

Article

# Joint Sparse and Low-Rank Multitask Learning with Laplacian-Like Regularization for Hyperspectral Classification

Zhi He \* , Yiwen Wang and Jie Hu

Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China; yizhigiser@126.com (Y.W.); verohu@163.com (J.H.)

\* Correspondence: hezh8@mail.sysu.edu.cn; Tel.: +86-20-8411-3057

Received: 8 November 2017; Accepted: 17 February 2018; Published: 21 February 2018

**Abstract:** Multitask learning (MTL) has recently provided significant performance improvements in supervised classification of hyperspectral images (HSIs) by incorporating shared information across multiple tasks. However, the original MTL cannot effectively exploit both local and global structures of the HSI and the class label information is not fully used. Moreover, although the mathematical morphology (MM) has attracted considerable interest in feature extraction of HSI, it remains a challenging issue to sufficiently utilize multiple morphological profiles obtained by various structuring elements (SEs). In this paper, we propose a joint sparse and low-rank MTL method with Laplacian-like regularization (termed as sllMTL) for hyperspectral classification by utilizing the three-dimensional morphological profiles (3D-MPs) features. The main steps of the proposed method are twofold. First, the 3D-MPs are extracted by the 3D-opening and 3D-closing operators. Different SEs are adopted to result in multiple 3D-MPs. Second, sllMTL is proposed for hyperspectral classification by taking the 3D-MPs as features of different tasks. In the sllMTL, joint sparse and low-rank structures are exploited to capture the task specificity and relatedness, respectively. Laplacian-like regularization is also added to make full use of the label information of training samples. Experiments on three datasets demonstrate the OA of the proposed method is at least about 2% higher than other state-of-the-art methods with very limited training samples.

**Keywords:** hyperspectral image (HSI); classification; low-rank; sparse; regularization; 3D morphological profile

## 1. Introduction

By simultaneously acquiring hundreds of continuous narrow spectral wavelengths for each image pixel, hyperspectral remote sensors can generate three-dimensional (3D) hyperspectral cubes containing rich spectral and spatial information. Thanks to the high spectral and spatial resolution, hyperspectral images (HSIs) have opened up new avenues for the remote sensing applications, including earth observation [1], environmental science [2], agriculture [3], mineral identification [4], etc. Supervised classification [5–8], which aims at assigning each pixel in the HSI into an accurate class label with representative training samples, plays an important role in those applications. However, it still faces some challenges in the classification task. For example, a large number of labeled samples are crucial to producing good classification results due to the Hughes phenomenon, but in reality, it is extremely difficult even impossible to identify the labels of samples. Other difficulties, such as the spectral signature from identical material may vary while different materials may have similar signatures, will also give a negative effect on the classification performance.

Much work has been carried out in the literature to address the above-mentioned dilemmas. The first category is to design reliable classifiers, among which support vector machine (SVM) [9,10] is one of the state-of-the-art methods. The success of SVM stems from the fact that it is robust to the high spectral dimension by using kernel tricks. Variations of the SVM-based methods are proposed to improve the classification performance, such as incorporating spatial information by composite kernels (SVMCK) [11] or multiple kernels [12], combining the kernel matrices of a family of SVM classifiers by ensemble learning [13], and the semi-supervised learning extension [14,15] by exploiting both labeled and unlabeled samples, etc. However, the kernel tricks in the SVM require proper tuning of parameters. Sparse representation-based classification (SRC), which is pioneered by Wright et al. [16] for face recognition, has become another widely used hyperspectral classifier [17–25]. The SRC is suitable for classifying the HSIs because it relies on the assumption that the spectral signatures of hyperspectral pixels belonging to the same class approximately lie in the same low-dimensional subspace and the unknown test sample can be sparsely represented by a linear combination of a few atoms from the whole training dictionary. Haq et al. [26] propose a homotopy-based SRC to alleviate the problem introduced by the high-dimensional data with too few labeled samples. The contextual information of neighboring pixels can also be incorporated to solve the problem of lack of sufficient training samples [18,19,22,23]. Multitask learning (MTL) [27–31], which can learn shared information across multiple tasks, has also demonstrated to be very powerful. In [28], several complementary features (i.e., spectral, gradient, Gabor texture and shape) are combined in a MTL fusion to enhance the discrimination of hyperspectral pixels, while kernel joint collaborative representation with adaptive weighted MTL is proposed in [29] to obtain desirable classification results. A SVM-based MTL method is proposed in [31] to effectively utilize the rich information included in the extracted two-dimensional (2D) Gabor features. Although many contributions have been carried out to improve the classification performance of MTL, till now, existing MTL methods rarely exploit both the task specificity and relatedness, and the label information of the training samples is also not fully utilized.

The second category is to increase the class separability by feature extraction/selection. Apart from spectral profiles, the spatial context can be adopted to improve the classification results. Many spatial filters have been proposed over the past few years, such as the edge-preserving filtering [5,32] and the mathematical morphology (MM) processing [33,34], where the spatial-domain features are extracted by edge-preserving filter or opening/closing operators. In [32], the edge-preserving filtering is conducted on the probability maps (i.e., probability presentation of the classification maps) and the label of a test sample is determined by maximum probability. In [5], a principal component analysis (PCA)-based edge-preserving method is proposed to highlight the separability of pixels. This simple yet powerful filtering approach has announced impressive results for hyperspectral classification. The MM-based method proposed in [33,34] is termed as extended morphological profile (EMP), where the opening and closing operators with different sizes of structuring elements (SEs) are performed on the first several principal components. It is worth stressing that a HSI data is naturally formed as a 3D cube consisting of a spectral dimension and two spatial dimensions. However, the above-mentioned methods ignore the 3D nature of the HSI.

In this paper, we propose a joint sparse and low-rank MTL method with Laplacian-like regularization (abbreviated as sllMTL) for classification of HSI by utilizing the 3D morphological profiles (3D-MPs). The visual illustration of the proposed method is shown in Figure 1, which is consisted of two major steps. In the feature extraction step, the 3D-MPs [35] of the HSI are obtained by the 3D-opening and 3D-closing operators with various SEs. Each SE can produce two corresponding 3D-MPs (one is obtained by the 3D-opening and another is generated by the 3D-closing). The significant difference between the existing 3D-MPs-based method [35] and our proposed method is that: other than simply stacking multiple 3D-MPs together, we make full use of the 3D-MPs cubes simultaneously in a MTL manner. In the classification step, the sllMTL is proposed to exploit the joint sparse and low-rank structures by treating each 3D-MP as the feature of a specific task. Both sparse and low-rank constraints are added to the representation coefficients so as to capture the task specificity and relatedness.

A Laplacian-like regularization is also adopted to take advantage of the class label information. Finally, the class label of the test sample can be determined by the minimal total residual.

The advantages of the proposed method lie in the following two aspects:

- 3D-MPs are adopted for feature extraction of the HSI cube. Compared to the vector/image-based methods, we respect the 3D nature of the HSI and take the HSI cube as a whole entity to simultaneously extract the spectral-spatial features. Different from the existing 3D-MPs-based method, we do not simply stack multiple 3D-MPs but fully exploit the spectral-spatial feature of each 3DMP in a novel MTL framework.
- sllMTL is proposed to simultaneously classify the 3D-MPs by taking each 3D-MP as the feature of a specific task. Compared to the existing MTL, the proposed sllMTL can capture both specificities and shared factors of the tasks by utilizing the sparse and low-rank constraints. Moreover, the Laplacian-like regularization can improve the classification performance further by making full use of the class label information.

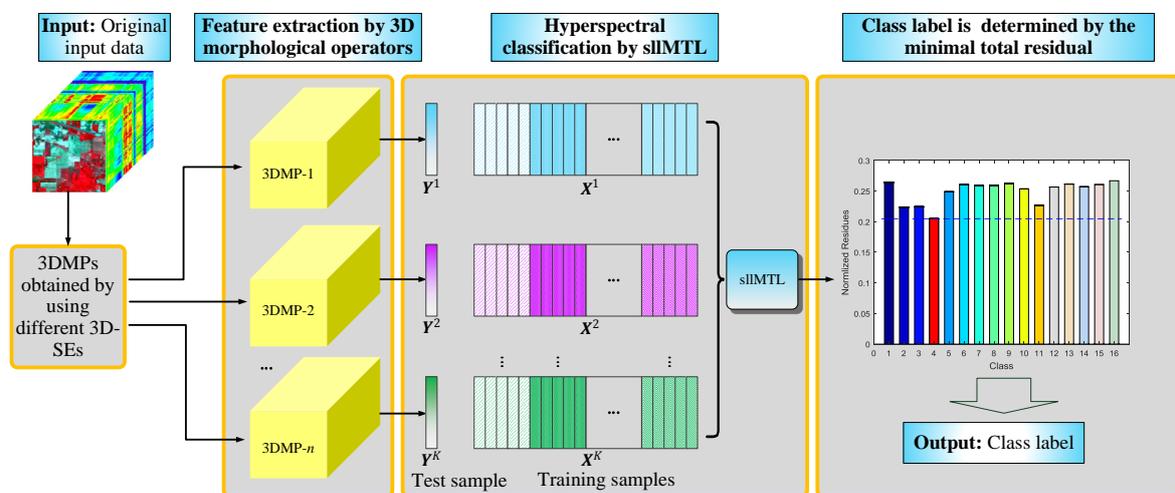


Figure 1. Schematic illustration of the proposed HSI classification method.

The remainder of this paper is organized as follows. Section 2 introduces two types of 3D mathematical morphology methods (i.e., 3D-opening and 3D-closing operators) to obtain multiple 3D-MPs. Section 3 describes the proposed sllMTL for hyperspectral classification in detail. Section 4 shows the experimental results, and discussions and conclusions are drawn in Sections 5 and 6.

## 2. Multiple Morphological Profiles Extracted by 3D Mathematical Morphology

Although the EMP has been successfully employed in feature extraction of the HSI, it ignores the 3D nature of the hyperspectral cube and thus cannot fully investigate the 3D spectral-spatial dependence among pixels. In this section, 3D mathematical operators (i.e., 3D-opening and 3D-closing) are applied for multiple morphological profiles extraction of HSI.

MM is initially proposed as a rigorous theoretic framework for analyzing the spatial relationship of different pixels by using SEs with given shapes and sizes as input. The relationship and structural character of the image can be captured by keeping the SEs sliding over the image. In analogy to the 2D MM operators, the 3D extension ones learn the structural characters through the moving of 3D-SEs in the image. The major advantage of the 3D-SEs lies in that they can effectively explore the spectral-spatial structure of the 3D images (e.g., HSI). It is noteworthy that erosion and dilation are two basic morphological operators. Suppose the original HSI cube is represented as  $I \in \mathbb{R}^{m \times n \times b}$ , where  $m$ ,  $n$  and  $b$  refer to the number of rows, columns and spectral bands, respectively, the value in the cube  $I$  for a given coordinate  $(x_0, y_0, z_0)$  is  $I(x_0, y_0, z_0)$  with  $I(x_0, y_0, z_0) \in \mathbb{R}$ ,  $B$  is a 3D-SE and  $B(x_0, y_0, z_0)$

indicates a set of values centered at the coordinate  $(x_0, y_0, z_0)$  of  $I$ , the 3D-erosion  $\epsilon_B^{3D}(I(x_0, y_0, z_0))$  of the cube  $I$  at the coordinate  $(x_0, y_0, z_0)$  with the 3D-SE  $B$  is determined by the minimum value of pixels inside  $B(x_0, y_0, z_0)$

$$\epsilon_B^{3D}(I(x_0, y_0, z_0)) = I \ominus B = \min_{(x,y,z)} (I(x, y, z) \in B(x_0, y_0, z_0)) \tag{1}$$

where “ $\ominus$ ” indicates the 3D-erosion operator.

The 3D-dilation  $\delta_B^{3D}(I(x_0, y_0, z_0))$  can be regarded as the dual operator, which switches the minimum operator to the maximum one

$$\delta_B^{3D}(I(x_0, y_0, z_0)) = I \oplus B = \max_{(x,y,z)} (I(x, y, z) \in B(x_0, y_0, z_0)) \tag{2}$$

where “ $\oplus$ ” indicates the 3D-dilation operator.

The 3D-erosion can expand objects of the HSI cube that are darker than their neighborhoods, while the 3D-dilation shrinks them, and vice versa for the objects that are brighter than their neighborhoods. The 3D-opening and 3D-closing operators can be derived from the combination of 3D-erosion and 3D-dilation. In greater detail, the 3D-opening  $\gamma_B^{3D}(I)$  is calculated by the 3D-erosion of the HSI cube  $I$  with  $B$  followed by the 3D-dilation with  $B$

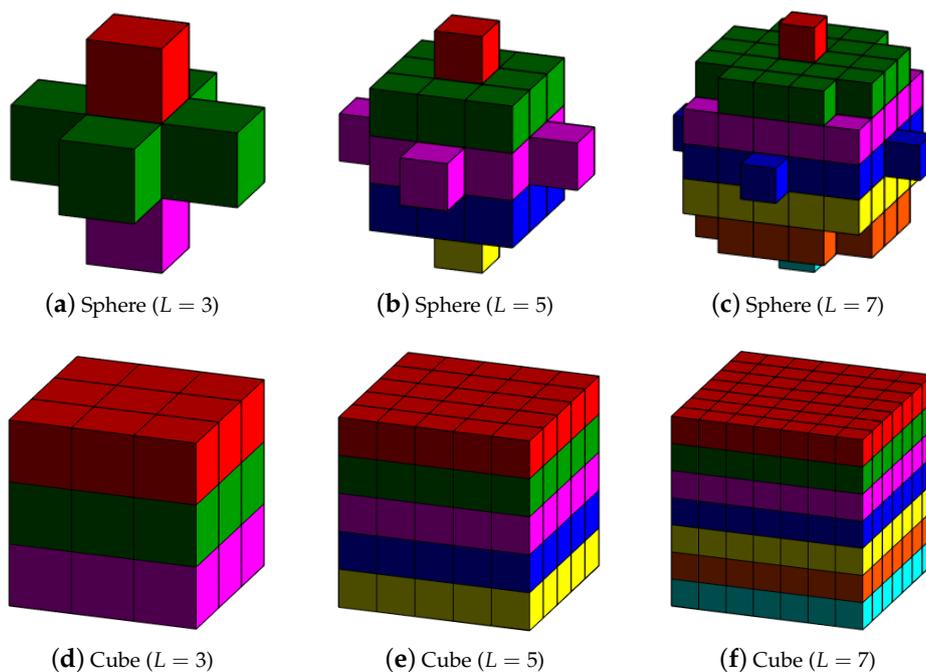
$$\gamma_B^{3D}(I) = I \circ B = (I \ominus B) \oplus B = \delta_B^{3D}(\epsilon_B^{3D}(I)) \tag{3}$$

where “ $\circ$ ” indicates the 3D-opening operator.

Similarly, the 3D-closing  $\phi_B^{3D}(I)$  can be obtained by the 3D-dilation of the HSI cube  $I$  with  $B$  followed by the 3D-erosion with  $B$

$$\phi_B^{3D}(I) = I \bullet B = (I \oplus B) \ominus B = \epsilon_B^{3D}(\delta_B^{3D}(I)) \tag{4}$$

where “ $\bullet$ ” indicates the 3D-closing operator.



**Figure 2.** 3D-SEs examples with various shapes (e.g., cube, sphere) and sizes (e.g., side length  $L$ ) is set to 3, 5 and 7). (a–c) are spheres with different sizes, and (d–f) are cubes with different sizes.

The multiple 3D-MPs of  $I$  in our proposed method are generated by performing a series of 3D-opening and 3D-closing operators with different 3D-SEs. As an example, Figure 2 depicts a set of 3D-SEs with various shapes (e.g., cube, sphere) and sizes (e.g., side length ( $L$ ) is set to 3, 5 and 7). Each 3D-SE can produce two corresponding 3D-MPs in merit of the 3D-opening and 3D-closing, respectively.

### 3. Joint Sparse and Low-Rank Multitask Learning with Laplacian-Like Regularization

In this section, we propose the sllMTL method to exploit both the local and global structures of the HSI by taking the 3D-MPs obtained from the 3D-opening and 3D-closing morphological operators as features of multiple tasks. Both sparse and low-rank constraints are considered in the sllMTL to capture the task specificity and relatedness. Moreover, Laplacian-like regularization is also added to improve the classification performance by taking full advantage of the label information of training samples. Suppose  $K$  3D-MPs are generated by the 3D morphological operators, there are  $K$  tasks in the sllMTL. For each task index  $k = 1, 2, \dots, K$ , represent  $\mathbf{X}^k = [\mathbf{X}_1^k, \mathbf{X}_2^k, \dots, \mathbf{X}_J^k]$  as the training feature matrix, in which  $\mathbf{X}_j^k \in \mathbb{R}^{m_k \times p_j}$ ,  $j = 1, 2, \dots, J$  is related to the  $j$ th class,  $J$  indicates the total number of classes contained in the HSI,  $m_k$  refers to the dimensionality of the  $k$ th task (note that no feature reduction is performed,  $m_k$  equals to the number of spectral bands  $b$ ),  $p_j$  denotes the number of training samples belong to class  $j$ , and  $\sum_{j=1}^J p_j = P$  is the total number of training samples. Given the unlabeled test sample  $\mathbf{Y} = [\mathbf{Y}^1, \mathbf{Y}^2, \dots, \mathbf{Y}^K]$ , the sllMTL method is formulated as the solution to the following problem

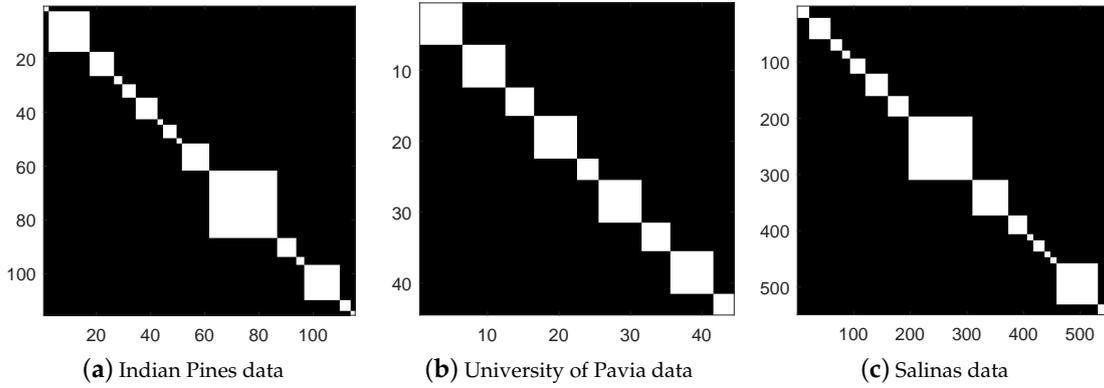
$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{Z}\|_* + \lambda \|\mathbf{Z}\|_{1,2} + \beta \text{tr}(\mathbf{Z}^T \mathbf{L} \mathbf{Z}) \\ \text{s.t.} \quad & \mathbf{Y}^k = \mathbf{X}^k \mathbf{Z}^k, k = 1, 2, \dots, K \end{aligned} \tag{5}$$

where  $\mathbf{Z}$  denotes the coefficient matrix to be calculated, joint sparse and low-rank constraints are added to  $\mathbf{Z}$ ,  $\|\mathbf{Z}\|_*$  refers to the trace norm (also known as the nuclear norm, i.e., the sum of singular values of a matrix), which can be utilized to identify the shared characters among multiple tasks,  $\|\mathbf{Z}\|_{1,2}$  indicates the row-sparse constraint to identify the task specificities,  $\mathbf{L}$  is the Laplacian-like matrix,  $\text{tr}(\cdot)$  is the trace of a matrix, i.e., the sum of the elements on the main diagonal. The matrix  $\mathbf{L}$  is generated by the natural assumption that the coefficients  $\mathbf{Z}_p$  and  $\mathbf{Z}_{\tilde{p}}$ ,  $p, \tilde{p} = 1, 2, \dots, P$  from the same class are more closer or similar than those belong to different classes. This assumption can be realized in terms of similarity. In mathematics, a reasonable way is to minimize the following function

$$\begin{aligned} \frac{1}{2} \sum_{p, \tilde{p}=1}^P \|\mathbf{Z}_p - \mathbf{Z}_{\tilde{p}}\|_2^2 \mathbf{W}_{p\tilde{p}} &= \sum_{p=1}^P \mathbf{Z}_p \mathbf{Z}_p^T \mathbf{D}_{pp} - \sum_{p, \tilde{p}=1}^P \mathbf{Z}_p \mathbf{Z}_{\tilde{p}}^T \mathbf{W}_{p\tilde{p}} \\ &= \text{tr}(\mathbf{Z}^T \mathbf{D} \mathbf{Z}) - \text{tr}(\mathbf{Z}^T \mathbf{W} \mathbf{Z}) \\ &= \text{tr}(\mathbf{Z}^T \mathbf{L} \mathbf{Z}) \end{aligned} \tag{6}$$

where  $\mathbf{D}$  is a diagonal matrix in which  $\mathbf{D}_{pp} = \sum_{\tilde{p}=1}^P \mathbf{W}_{p\tilde{p}}$ ,  $\mathbf{L}$  is the Laplacian-like matrix obtained by  $\mathbf{L} = \mathbf{D} - \mathbf{W}$ , and  $\mathbf{W}_{p\tilde{p}} = \begin{cases} 1 & \text{if the } p\text{th and } \tilde{p}\text{th training samples belong to the same class} \\ 0 & \text{otherwise} \end{cases}$ . To give an intuitive explanation, Figure 3 visually shows the  $\mathbf{W}$  used in the three HSI datasets. It should be mentioned that the label information of the training samples are adopted to obtain the matrix  $\mathbf{W}$ , which can then be utilized to generate  $\mathbf{L}$  ( $\mathbf{L} = \mathbf{D} - \mathbf{W}$ ) in Equation (5). Therefore, the training samples are utilized more fully in the proposed method. Moreover, although we add a Laplacian-like regularization in the joint sparse and low-rank multitask learning method, it cannot state that the Laplacian-like regularization is more effective than the sparse and low-rank regularizations. Actually, the Laplacian-like regularization is complementary to the sparse and low-rank regularizations. Note that the usage of the low-rank and sparsity priors have achieved great success in HSI classification, we

also use the joint sparse and low-rank regularizations (i.e.,  $\|Z\|_*$  and  $\|Z\|_{1,2}$ ) in the proposed method (see Equation (5)).



**Figure 3.** The  $W$  used in (a) Indian Pines data, (b) University of Pavia data and (c) Salinas data.

To solve Equation (5), we first introduce two auxiliary matrices  $P$  and  $Q$  to make the optimization problem in Equation (5) separable and simplify the solution procedure, in this regard, the problem in Equation (5) can be rewritten as

$$\begin{aligned} \min_{Z, P, Q} \quad & \|P\|_* + \lambda \|Q\|_{1,2} + \beta \text{tr}(P^T L P) \\ \text{s.t.} \quad & Y^k = X^k Z^k, Z^k = Q^k, Z^k = P^k, k = 1, 2, \dots, K \end{aligned} \quad (7)$$

Linearized alternating direction method with Adaptive Penalty (LADMAP) [36] can be adopted to solve the above-mentioned problem, whose augmented Lagrangian function yields

$$\begin{aligned} \mathcal{L}(Z, P, Q) = & \|P\|_* + \lambda \|Q\|_{1,2} + \beta \text{tr}(P^T L P) \\ & + \sum_{k=1}^K \left\{ \langle M_1^k, Y^k - X^k Z^k \rangle + \langle M_2^k, Z^k - Q^k \rangle + \langle M_3^k, Z^k - P^k \rangle \right. \\ & \left. + \frac{\mu}{2} \left( \|Y^k - X^k Z^k\|_F^2 + \|Z^k - Q^k\|_F^2 + \|Z^k - P^k\|_F^2 \right) \right\} \end{aligned} \quad (8)$$

where  $\mu > 0$  is a penalty parameter,  $M_1^k, M_2^k, k = 1, 2, \dots, K$  are Lagrange multipliers,  $Z^k, P^k, Q^k$  are the  $k$ th column of  $Z, P, Q$ , respectively, and  $\|\cdot\|_F$  is the Frobenius norm obtained by the square root of the sum of the absolute squares of matrix elements, e.g.,  $\|A\|_F = \sqrt{\sum_i \sum_j |a_{ij}|^2}$ .

The variables  $Z, P$  and  $Q$  can be updated alternately by minimizing the augmented Lagrangian function (see Equation (8)) with other variables fixed. In case the augmented Lagrangian function is difficult to minimize with a specific variable, we will linearize the corresponding smooth component. Specifically, the updating process is as follows.

Fix other variables, and the variable  $P$  in the  $(t + 1)$ th iteration can be updated by minimizing  $\mathcal{L}(Z^{(t)}, P, Q^{(t)})$

$$\begin{aligned}
 \mathbf{P}^{(t+1)} &= \arg \min_{\mathbf{P}} \mathcal{L}(\mathbf{Z}^{(t)}, \mathbf{P}, \mathbf{Q}^{(t)}) \\
 &= \arg \min_{\mathbf{P}} \|\mathbf{P}\|_* + \beta \text{tr}(\mathbf{P}^T \mathbf{L} \mathbf{P}) + \frac{\mu}{2} \sum_{k=1}^K \left\{ \frac{2}{\mu} \langle \mathbf{M}_3^{k,(t)}, \mathbf{Z}^{k,(t)} - \mathbf{P}^k \rangle + \|\mathbf{Z}^{k,(t)} - \mathbf{P}^k\|_{\mathbb{F}}^2 \right\} \\
 &= \arg \min_{\mathbf{P}} \|\mathbf{P}\|_* + \beta \text{tr}(\mathbf{P}^T \mathbf{L} \mathbf{P}) + \frac{\mu}{2} \sum_{k=1}^K \left\{ \left\| \mathbf{Z}^{k,(t)} - \mathbf{P}^k + \frac{\mathbf{M}_3^{k,(t)}}{\mu} \right\|_{\mathbb{F}}^2 \right\} \\
 &= \arg \min_{\mathbf{P}} \|\mathbf{P}\|_* + \beta \text{tr}(\mathbf{P}^T \mathbf{L} \mathbf{P}) + \frac{\mu}{2} \left\| \mathbf{Z}^{(t)} - \mathbf{P} + \frac{\mathbf{M}_3^{(t)}}{\mu} \right\|_{\mathbb{F}}^2 \\
 &= \arg \min_{\mathbf{P}} \|\mathbf{P}^T\|_* + \beta \text{tr}(\mathbf{P}^T \mathbf{L} \mathbf{P}) + \frac{\mu}{2} \left\| \left( \mathbf{Z}^{(t)} - \mathbf{P} + \frac{\mathbf{M}_3^{(t)}}{\mu} \right)^T \right\|_{\mathbb{F}}^2 \\
 &= \arg \min_{\tilde{\mathbf{P}}} \|\tilde{\mathbf{P}}\|_* + \beta \text{tr}(\tilde{\mathbf{P}} \mathbf{L} \tilde{\mathbf{P}}^T) + \frac{\mu}{2} \left\| \mathbf{Z}^{(t)T} - \tilde{\mathbf{P}} + \frac{\mathbf{M}_3^{(t)T}}{\mu} \right\|_{\mathbb{F}}^2
 \end{aligned} \tag{9}$$

where  $\tilde{\mathbf{P}} = \mathbf{P}^T$ .

Note that Equation (9) does not have a closed-form solution, we express the smooth component of  $\mathcal{L}(\mathbf{Z}^{(t)}, \mathbf{P}, \mathbf{Q}^{(t)})$  as

$$q(\tilde{\mathbf{P}}) = \beta \text{tr}(\tilde{\mathbf{P}} \mathbf{L} \tilde{\mathbf{P}}^T) + \frac{\mu}{2} \left\| \mathbf{Z}^{(t)T} - \tilde{\mathbf{P}} + \frac{\mathbf{M}_3^{(t)T}}{\mu} \right\|_{\mathbb{F}}^2 \tag{10}$$

Inspired by the LADMAP method, minimizing  $\mathcal{L}(\mathbf{Z}^{(t)}, \mathbf{P}, \mathbf{Q}^{(t)})$  can be approximately substituted by solving the following problem

$$\begin{aligned}
 \tilde{\mathbf{P}}^{(t+1)} &\approx \arg \min_{\mathbf{P}=\tilde{\mathbf{P}}^T} \|\tilde{\mathbf{P}}\|_* + \langle \nabla_{\tilde{\mathbf{P}}} q(\tilde{\mathbf{P}}^{(t)}), \tilde{\mathbf{P}} - \tilde{\mathbf{P}}^{(t)} \rangle + \frac{\eta}{2} \|\tilde{\mathbf{P}} - \tilde{\mathbf{P}}^{(t)}\|_{\mathbb{F}}^2 \\
 &= \arg \min_{\mathbf{P}=\tilde{\mathbf{P}}^T} \frac{1}{\eta} \|\tilde{\mathbf{P}}\|_* + \frac{1}{\eta} \langle \nabla_{\tilde{\mathbf{P}}} q(\tilde{\mathbf{P}}^{(t)}), \tilde{\mathbf{P}} - \tilde{\mathbf{P}}^{(t)} \rangle + \frac{1}{2} \|\tilde{\mathbf{P}} - \tilde{\mathbf{P}}^{(t)}\|_{\mathbb{F}}^2 \\
 &= \arg \min_{\mathbf{P}=\tilde{\mathbf{P}}^T} \frac{1}{\eta} \|\tilde{\mathbf{P}}\|_* + \frac{1}{2} \left\{ \frac{2}{\eta} \langle \nabla_{\tilde{\mathbf{P}}} q(\tilde{\mathbf{P}}^{(t)}), \tilde{\mathbf{P}} - \tilde{\mathbf{P}}^{(t)} \rangle + \|\tilde{\mathbf{P}} - \tilde{\mathbf{P}}^{(t)}\|_{\mathbb{F}}^2 \right\} \\
 &= \arg \min_{\mathbf{P}=\tilde{\mathbf{P}}^T} \frac{1}{\eta} \|\tilde{\mathbf{P}}\|_* + \frac{1}{2} \left( \left\| \tilde{\mathbf{P}} - \tilde{\mathbf{P}}^{(t)} + \frac{1}{\eta} \nabla_{\tilde{\mathbf{P}}} q(\tilde{\mathbf{P}}^{(t)}) \right\|_{\mathbb{F}}^2 \right)
 \end{aligned} \tag{11}$$

where  $q(\tilde{\mathbf{P}})$  is approximated by its linearization  $\langle \nabla_{\tilde{\mathbf{P}}} q(\tilde{\mathbf{P}}^{(t)}), \tilde{\mathbf{P}} - \tilde{\mathbf{P}}^{(t)} \rangle$  at  $\tilde{\mathbf{P}}^{(t)}$  plus a proximal term  $\frac{\eta}{2} \|\tilde{\mathbf{P}} - \tilde{\mathbf{P}}^{(t)}\|_{\mathbb{F}}^2$ , and  $\nabla_{\tilde{\mathbf{P}}} q(\tilde{\mathbf{P}}^{(t)})$  denotes the gradient of  $q$  w.r.t.  $\tilde{\mathbf{P}}$ . As long as  $\eta > 2\beta \|\mathbf{L}\|_2 + \mu$ , the above substitution is valid. Therefore, the Equation (11) has a closed-form solution formulated as

$$\begin{aligned}
 \tilde{\mathbf{P}}^{(t+1)} &= \mathcal{D}_{\mu^{-1}} \left( \tilde{\mathbf{P}}^{(t)} - \frac{1}{\eta} \nabla_{\tilde{\mathbf{P}}} q(\tilde{\mathbf{P}}^{(t)}) \right) \\
 &= \mathbf{U} \mathcal{D}_{\mu^{-1}}(\boldsymbol{\Sigma}) \mathbf{V}^T \\
 &= \mathbf{U} \text{diag} \left( \left\{ \left( \pi_i - \frac{1}{\mu} \right)_+ \right\}_{1 \leq i \leq \tilde{r}} \right) \mathbf{V}^T
 \end{aligned} \tag{12}$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are the matrices with orthogonal columns,  $(\cdot)_+ = \max(\cdot, 0)$ ,  $\tilde{r}$  and  $\pi_i$  refer to the rank and singular values of  $\left( \tilde{\mathbf{P}}^{(t)} - \frac{1}{\eta} \nabla_{\tilde{\mathbf{P}}} q(\tilde{\mathbf{P}}^{(t)}) \right)$ , respectively, and  $\mathbf{P}^{(t+1)} = \tilde{\mathbf{P}}^{(t+1)T}$ .

Fix the others, and the variable  $\mathbf{Q}$  in the  $(t + 1)$ th iteration can be updated by

$$\begin{aligned}
 \mathbf{Q}^{(t+1)} &= \arg \min_{\mathbf{Q}} \mathcal{L}(\mathbf{Z}^{(t)}, \mathbf{P}^{(t+1)}, \mathbf{Q}) \\
 &= \arg \min_{\mathbf{Q}} \frac{\lambda}{\mu} \|\mathbf{Q}\|_{1,2} + \frac{1}{2} \left\| \mathbf{Q} - \left( \mathbf{Z}^{(t)} + \frac{\mathbf{M}_2^{(t)}}{\mu} \right) \right\|_{\mathbb{F}}^2
 \end{aligned} \tag{13}$$

which has the closed-form solution represented by the row-shrinkage thresholding operator.

Finally, fix the others and the variable  $Z$  in the  $(t + 1)$ th iteration can be updated by

$$\begin{aligned} Z^{(t+1)} &= \arg \min_Z \mathcal{L}(Z, P^{(t+1)}, Q^{(t+1)}) \\ &= \arg \min_Z \sum_{k=1}^K \left\{ Z^{k,(t+1)T} \left( X^{k,(t+1)T} X^{k,(t+1)} + 2I \right) Z^{k,(t+1)} - 2Z^{k,(t+1)T} \left( X^{k,(t+1)T} Y^{k,(t+1)} \right. \right. \\ &\quad \left. \left. - \frac{X^{k,(t+1)T} M_1^{k,(t)}}{\mu} + Q^{k,(t+1)} - \frac{M_2^{k,(t)}}{\mu} + P^{k,(t+1)} - \frac{M_3^{k,(t)}}{\mu} \right) \right\} \end{aligned} \tag{14}$$

According to Equation (14), the  $k$ th column of  $Z^{(t+1)}$  yields

$$Z^{k,(t+1)} = \left( X^{k,(t+1)T} X^{k,(t+1)} + 2I \right)^{-1} H^{k,(t+1)} \tag{15}$$

where  $H^{k,(t+1)} = X^{k,(t+1)T} Y^{k,(t+1)} - \frac{X^{k,(t+1)T} M_1^{k,(t)}}{\mu} + Q^{k,(t+1)} - \frac{M_2^{k,(t)}}{\mu} + P^{k,(t+1)} - \frac{M_3^{k,(t)}}{\mu}$  and  $I$  is an identity matrix.

The above-mentioned equations are the major steps for updating the variables  $Z, P, Q$ . The detailed procedure for solving the optimization problem (see Equation (5)) in the proposed sllMTL method is summarized in Algorithm 1, in which  $\mu$  is adjusted using an adaptive updating strategy [36] for faster convergence. As long as the coefficient matrix  $Z$  is obtained, the label of the unlabeled test sample  $Y$  can be determined by the minimal total residual

$$\text{Class } Y = \arg \min_j r_j = \arg \min_j \sum_k \left\| Y^k - X_j^k Z_j^k \right\|_2^2 \tag{16}$$

where  $X_j^k$  and  $Z_j^k$  are associated with the  $j$ th ( $j = 1, 2, \dots, J$ ) class.

---

**Algorithm 1:** LADMALP for Solving the Optimization Problem (see Equation (5)) in the Proposed sllMTL Method

---

- Require:** The test feature matrix  $Y$ , the training feature matrix  $X^k, k = 1, 2, \dots, K$ , and the parameters  $\lambda$  and  $\beta$
- Ensure:** The solution  $Z$  and class label of the unlabeled test sample  $Y$
- 1: Initialize:  $Z = P = Q = M_2 = M_3 = 0, M_1 = 0, \rho = 1.1, \mu = 10^{-6}, \mu_{\max} = 10^6$  and  $t = 0$
  - 2: **while** Convergence is not attained **do**
  - 3: Fix the others and update the variable  $P$  by Equation (12)
  - 4: Fix the others and update the variable  $Q$  by Equation (13)
  - 5: Fix the others and update the variable  $Z$  by Equation (15)
  - 6: Update the Lagrange multipliers  $M_1, M_2$  and  $M_3$  by
 
$$M_1^{k,(t+1)} = M_1^{k,(t)} + \mu(Y^{k,(t+1)} - X^{k,(t+1)} Z^{k,(t+1)})$$

$$M_2^{k,(t+1)} = M_2^{k,(t)} + \mu(Z^{k,(t+1)} - Q^{k,(t+1)})$$

$$M_3^{k,(t+1)} = M_3^{k,(t)} + \mu(Z^{k,(t+1)} - P^{k,(t+1)})$$
  - 7: Update the parameter  $\mu$  by  $\mu = \min(\mu_{\max}, \rho\mu)$
  - 8: Update  $t$  by  $t = t + 1$
  - 9: **end while**
  - 10: Determine the class label of  $Y$  by Equation (16)
-

## 4. Experiments

### 4.1. Dataset Description

In the experiments, we verify the performance of the proposed hyperspectral classification method on three benchmark datasets, i.e., the Indian Pines data, the University of Pavia data and the Salinas data. All the three hyperspectral datasets used in the experiments have been preprocessed by distortion correction. Details of each dataset are given below.

1. **Indian Pines data:** the first data was gathered by the National Aeronautics and Space Administration's (NASA) Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor on 12 June 1992 over the Indian Pines test site in Northwest Indiana, and it consists of  $145 \times 145$  pixels and 220 spectral reflectance bands cover the wavelength range of 0.4–2.5  $\mu\text{m}$ . The number of bands is reduced to 200 by removing the noisy and water-vapor absorption bands (bands 104–108, 150–163, and 220). The spatial resolution is about 20 m per pixel. Figure 4 depicts the three-band false color composite image together with its corresponding ground truth. The data contains 16 classes of interest land-covers and 10366 labeled pixels ranging unbalanced from 20 to 2468, which poses a big challenge to the classification problem. Table 1 displays the detailed number of samples of each class, and the background color represents different classes of land-covers.
2. **University of Pavia data:** the second dataset was captured by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor on 8 July 2002 over an urban area surrounding the University of Pavia, Italy. It contains  $610 \times 340$  pixels with a spatial resolution of 1.3 m per pixel. The original dataset has 115 spectral channels with a coverage range from 0.43 to 0.86  $\mu\text{m}$ . 12 most noisy bands are removed before experiments, remaining 103 bands for experiments. 9 classes of interest are considered in this dataset. The color composite image, the ground truth data as well as the available training samples are depicted in Figure 5. As shown in Table 2, there are more than 900 pixels in each class, but the available training samples of each class are less than 600. In analogy to the Indian Pines data, the background color in Table 2 also agrees with that in Figure 5b,c.
3. **Salinas data:** the third dataset was acquired by the AVIRIS sensor on 8 October 1998 over the Salinas Valley, Southern California, USA. There are 224 bands in the original dataset, and 24 bands (bands 108–112, 154–167, and 224) are removed for the water absorption. The size of each band is  $512 \times 217$  pixels with a spatial resolution of 3.7 m per pixel. The color composite image and the ground truth are plotted in Figure 6. This dataset contains 16 classes of ground truth, and the detailed number of samples in each class is listed in Table 3, whose background color corresponds to different classes of land-covers.

**Table 1.** Number of samples (NoS) used in the Indian Pines data.

Class	Name	NoS	Class	Name	NoS
1	 alfalfa	54	9	 oats	20
2	 corn-no till	1434	10	 soybean-no till	968
3	 corn-min till	834	11	 soybean-min till	2468
4	 corn	234	12	 soybean-clean till	614
5	 grass/pasture	497	13	 wheat	212
6	 grass/trees	747	14	 woods	1294
7	 grass/pasture-mowed	26	15	 bldg-grass-tree-drives	380
8	 hay-windrowed	489	16	 stone-steel towers	95
Total		10,366			

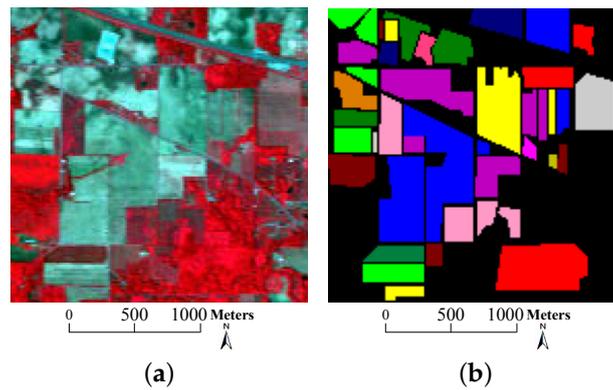


Figure 4. Indian Pines data. (a) Three-band false color composite and (b) ground truth data with 16 classes.

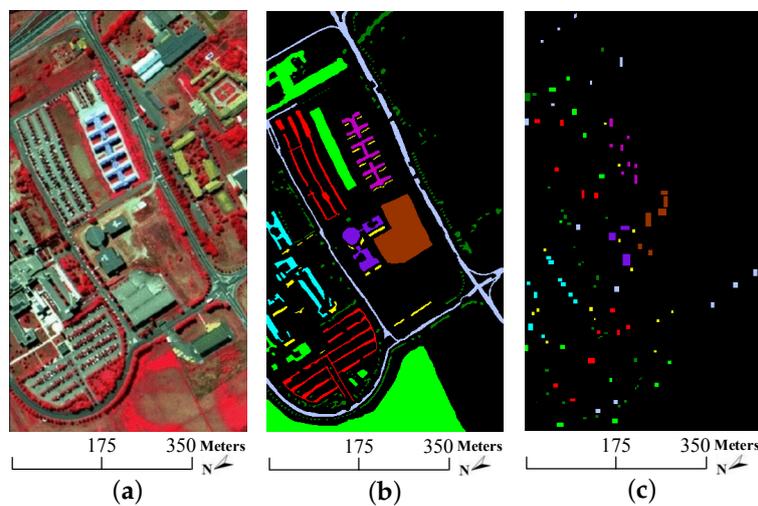


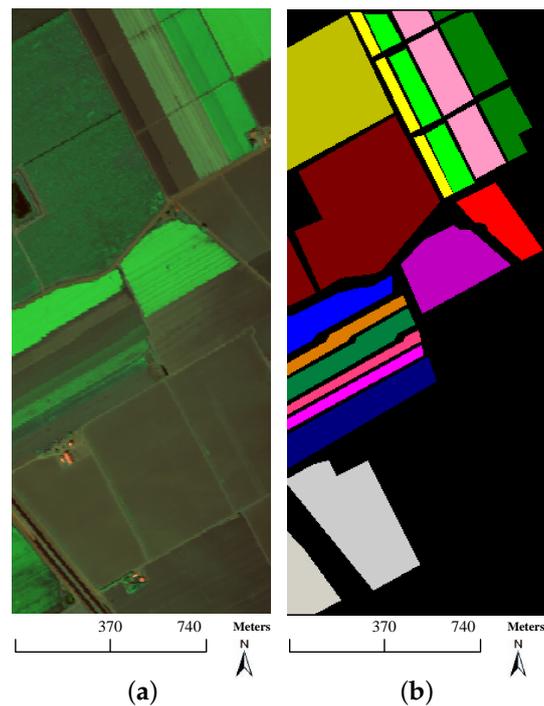
Figure 5. University of Pavia data. (a) Three-band false color composite, (b) ground truth data with 9 classes and (c) available training samples.

Table 2. NoS and available training samples (NoATS) used in the University of Pavia data.

Class	Name	NoS	NoATS	Class	Name	NoS	NoATS
1	asphalt	6631	548	6	bare soil	5029	532
2	meadows	18,649	540	7	bitumen	1330	375
3	gravel	2099	392	8	bricks	3682	514
4	trees	3064	524	9	shadows	947	231
5	metal sheets	1345	265	Total		42,776	3921

Table 3. NoS used in the Salinas data.

Class	Name	NoS	Class	Name	NoS
1	brocoli-green-weeds-1	2009	9	soil-vinyard-develop	6203
2	brocoli-green-weeds-2	3726	10	corn-senesced-green-weeds	3278
3	fallow	1976	11	lettuce-romaine-4wk	1068
4	fallow-rough-plow	1394	12	lettuce-romaine-5wk	1927
5	fallow-smooth	2678	13	lettuce-romaine-6wk	916
6	stubble	3959	14	lettuce-romaine-7wk	1070
7	celery	3579	15	vinyard-untrained	7268
8	grapes-untrained	11,271	16	vinyard-vertical-trellis	1807
Total				54,129	



**Figure 6.** Salinas data. (a) Three-band false color composite and (b) ground truth data with 16 classes.

#### 4.2. Experimental Design

To evaluate the effectiveness of the proposed method for hyperspectral classification, we compare our method with several baseline approaches, including the EMP [34], 3D discrete wavelet transform (3D-DWT) [37], SVM [9], SVMCK [11] and MTL [27]. The original spectral features (abbreviated as “Spec”) are also taken into consideration in the experiments. It is worthwhile to note that the “Spec”, EMP, 3D-DWT and 3D-MPs are feature extraction methods, whereas the SVM, SVMCK, MTL and sllMTL are classification methods. To investigate the performance of the proposed method in both feature extraction and classification, we compare the 3D-MPs with the “Spec”, EMP and 3D-DWT by fixing the classifier as sllMTL. Moreover, the sllMTL is compared with the SVM, SVMCK and MTL by utilizing the 3D-MPs-based features.

In greater detail, the 3D-MPs-based feature extraction method is validated by comparing with other widely-used methods, i.e., the original “Spec”, the multiple features obtained by the EMP, as well as the multiple features generated by the 3D-DWT. As to the above-mentioned feature extraction methods, (1) the original “Spec” features can be viewed as a special case of the sllMTL classifier with  $K = 1$ . (2) The EMP does not use the PCA for dimensional reduction but obtains the morphological profiles by performing the opening/closing operators with various SEs on all of the bands. (3) The 3D-DWT-based multiple cubes consisted of the subcubes with different scales, frequencies and orientations are also used for comparison. (4) The 3D-MPs-based multiple cubes are obtained by directly performing the 3D-opening and 3D-closing on the HSI cubes using different 3D-SEs. The 3D-SEs adopted in 3D-MPs are the cube and sphere shapes. Note that there are large areas of homogeneous regions in the Indian Pines data and Salinas data, the side lengths of the 3D-SEs are set to  $L = 5, 9, 13, 17$  and 21. Since many narrow regions are present in the University of Pavia data, smaller side lengths of 3D-SEs are chosen, i.e.,  $L = 3, 5, 9, 13$  and 17.

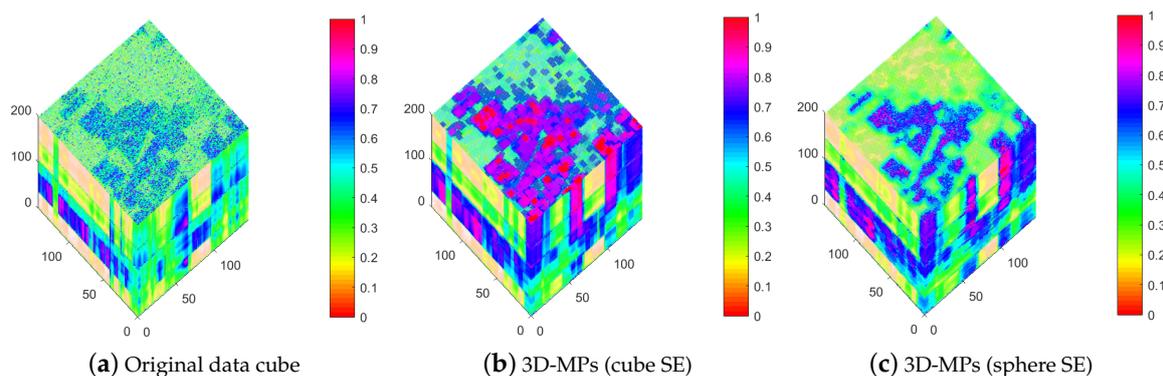
On the other hand, we test the performance of the sllMTL by fixing the input features as 3D-MPs and comparing the sllMTL with the SVM, SVMCK and MTL. Note that no nonlinear kernel is used in the proposed sllMTL, we adopt linear kernel in the SVM, SVMCK and MTL for consistency. For the SVM, which cannot handle the multiple-feature-based scenario, the multiple 3D-MPs are stacked together to form a single cube. For the SVMCK, the composite weight is set to  $\frac{1}{K}$  for all

of the 3D-MPs cubes. The penalty term in both SVM and SVMCK is set to 60. For the MTL, the regularization parameter  $\lambda$  is selected from  $[0.01, 0.05, 0.1, 0.2, \dots, 1]$ . For the sllMTL, three situations are taken into account: (1)  $\lambda = \beta = 0$  (abbreviated as sllMTL<sub>1</sub>), i.e., only low-rank constraint is considered; (2)  $\lambda \neq 0, \beta = 0$  (abbreviated as sllMTL<sub>2</sub>), i.e., joint sparse and low-rank constraints are considered; (3)  $\lambda \neq 0, \beta \neq 0$  (abbreviated as sllMTL<sub>3</sub>), i.e., all the sparse, low-rank and Laplacian-like constraints are considered. In the sllMTL<sub>2</sub> and sllMTL<sub>3</sub>, the nonzero parameters ( $\lambda$  or  $\beta$ ) are chosen from  $[0.01, 0.05, 0.1, 0.2, \dots, 1]$ .

Moreover, all the classification experiments are conducted on the aforementioned three hyperspectral datasets (see Section 4.1). Since the class labels are difficult to identify in reality, very limited number of labeled samples are utilized in the experiments. That means, 1% samples per class are randomly selected as training samples and the rest are for testing in the Indian Pines data and Salinas data. In case 1% samples of some classes are less than 2, we randomly choose 2 samples as training samples. Regarding to the University of Pavia data, whose available training samples (see Figure 5c) are separate from the whole ground truth data, 1% samples per class from the available training samples are randomly chosen for training. It should be highlighted that the experimental results in the tables are the average results of 10 independent trials to avoid possible bias. Three popular indexes (overall accuracy (OA), average accuracy (AA) and kappa coefficient ( $\kappa$ )) are adopted to compare different methods quantitatively.

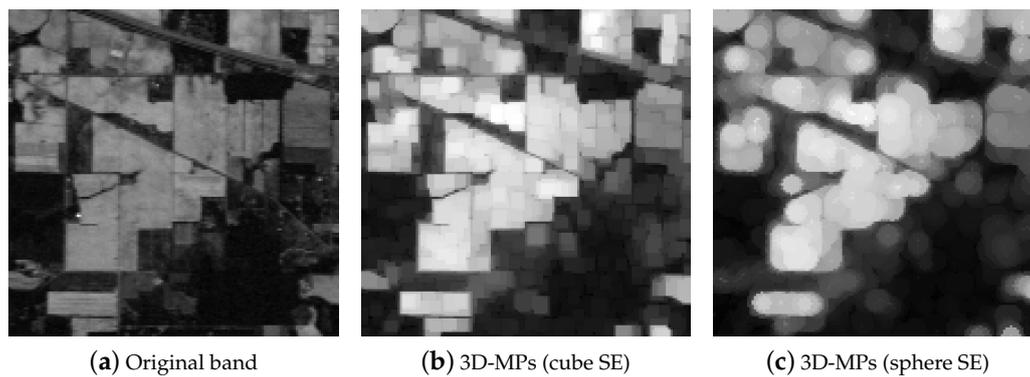
### 4.3. Experimental Results

In this section, the experimental results are provided to assess the performance of the 3D-MPs for spectral-spatial feature extraction and the sllMTL for hyperspectral classification. To this end, the 3D-MPs are compared with the “Spec”, EMP and 3D-DWT. The “Spec” utilizes the original spectral features, the EMP can incorporate spatial features, while 3D-DWT and 3D-MPs respect the 3D nature of the HSI and exploit spectral-spatial features. As an example, Figure 7 depicts the 3D-MPs obtained by 3D-opening with both cube and sphere 3D-SEs of side length  $L = 5$ , while Figure 8 plots the 3D-MPs of the 150th band. It can be observed from Figures 7 and 8 that the 3D-MPs highlight the structural characteristics of the HSI cube and provide smoother features in both spectral and spatial domains.

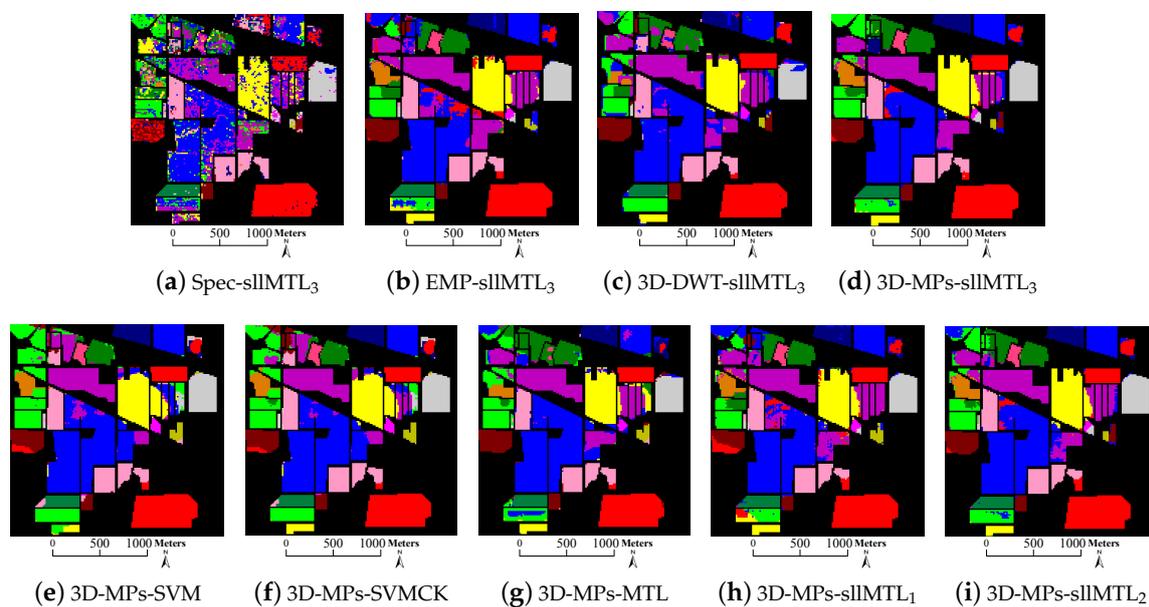


**Figure 7.** 3D-MPs obtained by 3D-opening with both cube and sphere SEs of side length  $L = 5$ . (a) is the original Indian Pines data, (b) is the 3D-MPs obtained by the cube SE, and (c) is the 3D-MPs obtained by the sphere SE.

The thematic maps of those methods in a trial are visually reported in Figures 9–11, while the detailed average classification accuracies are listed and compared in Tables 4–6. A few observations can be made based on the above-mentioned experimental results.



**Figure 8.** Spatial scenes of the 150th band. (a) is chosen from the original Indian Pines data, and (b,c) are chosen from the 3D-MPs.

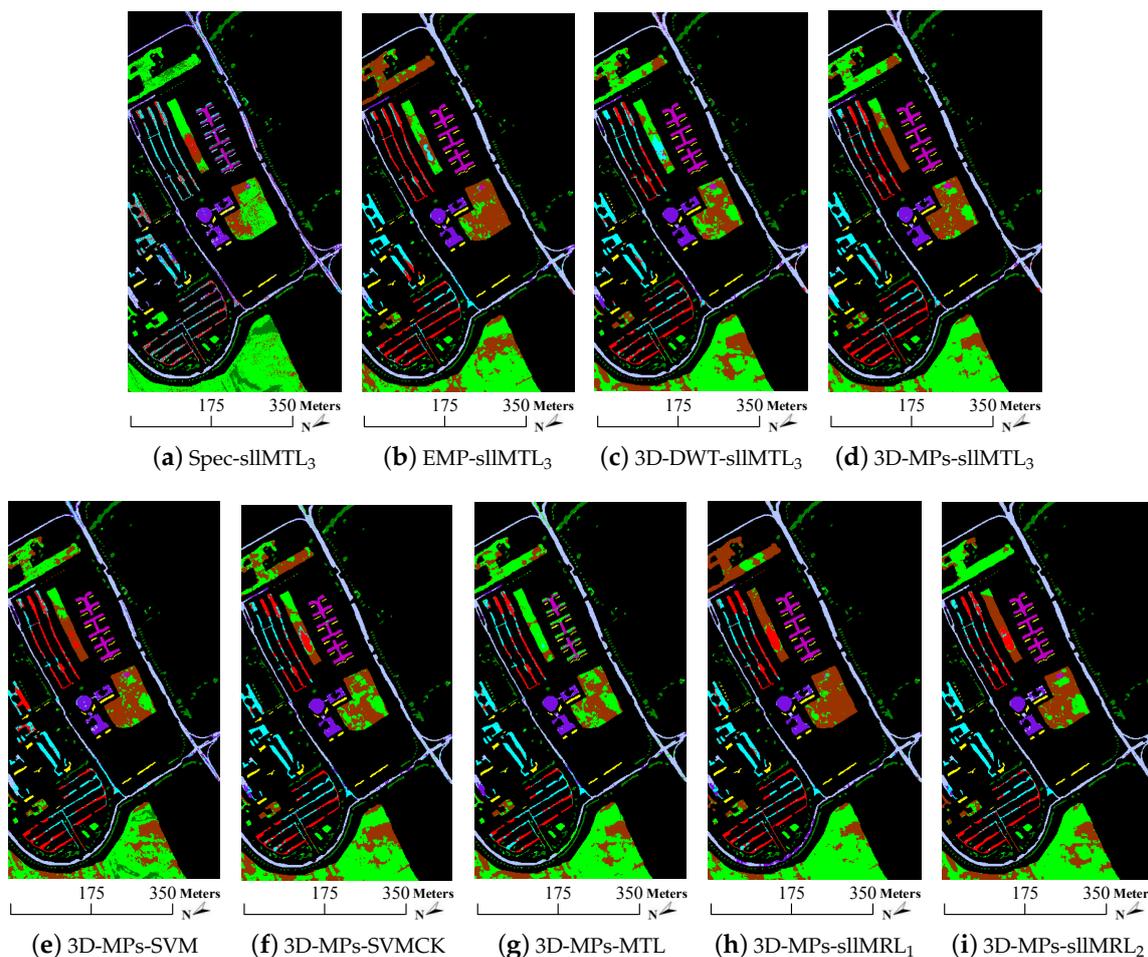


**Figure 9.** Classification maps of the Indian Pines data obtained by different methods.

It can be first seen that, the original “Spec”, which is based only on spectral characteristics, provides poor classification results compared with other methods. For instance, the OA of SVM in the Indian Pines data (see Table 4) is 16.86%, 17.30% and 20.01% lower than those of the EMP, 3D-DWT and 3D-MPs, respectively. As shown in Figure 9, the Spec-sllMTL<sub>3</sub> produces much more “salt-and-pepper” phenomenon than its counterparts. Similar properties can also be found in the University of Pavia data and the Salinas data. Note that the “Spec” utilizes the original spectral features without a spatial prior, it is not hard to infer that the spatial information can be utilized to alleviate the spectral variations and stabilize the pixel-based features. Therefore, the above-mentioned phenomena stress yet again the importance of spatial information for HSI classification.

**Table 4.** Classification accuracy (%) of different methods for the Indian Pines data, bold values indicate the best result for a row.

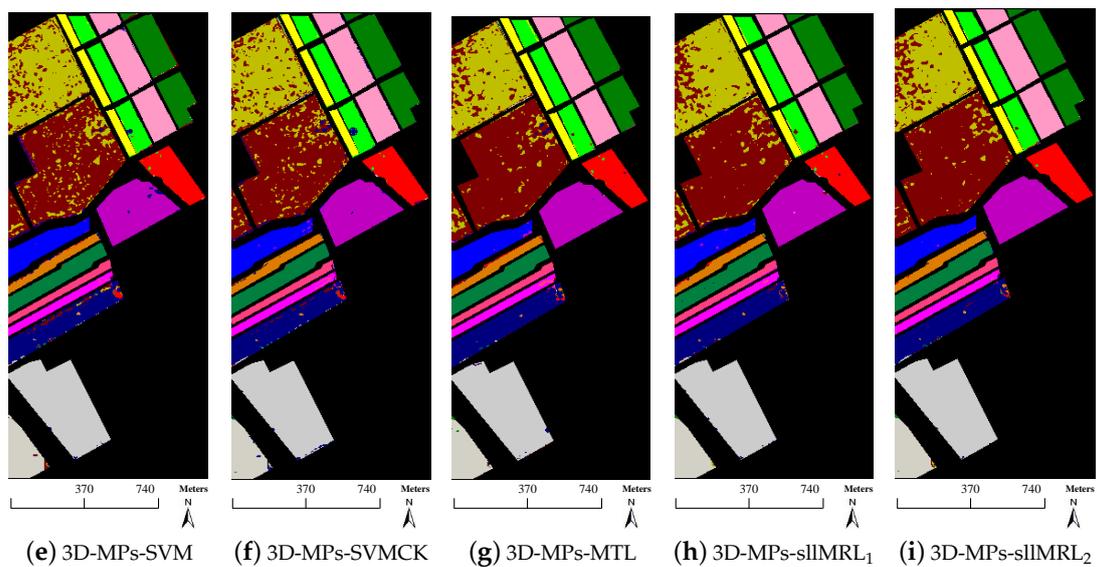
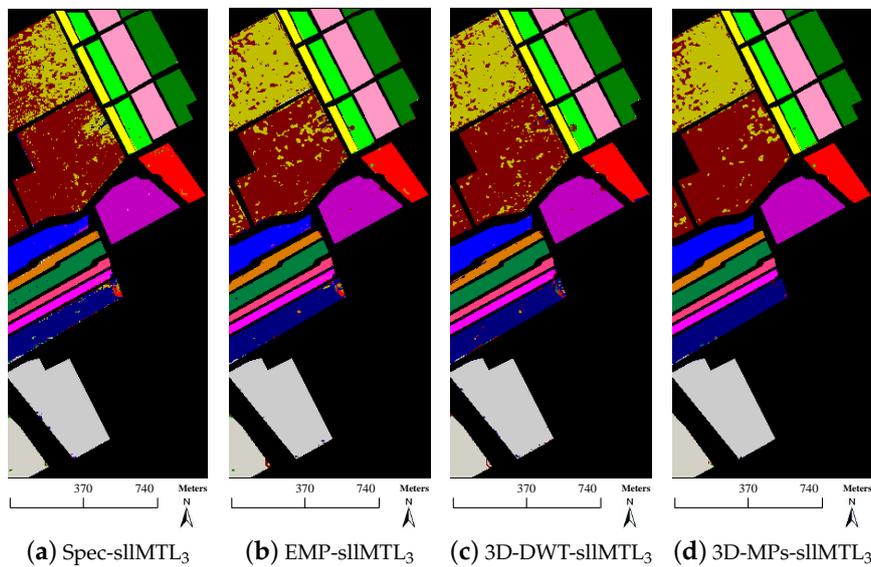
Class	sllMTL <sub>3</sub>				3D-MPs					
	Spec	EMP	3D-DWT	3D-MPs	SVM	SVMCK	MTL	sllMTL <sub>1</sub>	sllMTL <sub>2</sub>	sllMTL <sub>3</sub>
1	25.38	59.13	63.46	88.46	<b>93.27</b>	<b>93.27</b>	70.77	70.43	73.72	88.46
2	63.90	80.36	<b>82.52</b>	70.68	57.49	61.79	78.27	77.03	74.44	70.68
3	48.85	71.70	70.29	78.42	<b>87.79</b>	85.83	75.61	66.39	73.09	78.42
4	26.54	73.86	74.68	<b>86.15</b>	69.96	76.45	53.46	72.35	73.30	<b>86.15</b>
5	73.70	72.23	73.81	74.59	65.73	67.30	<b>75.94</b>	63.08	72.49	74.59
6	88.44	91.90	92.35	95.81	89.53	90.37	93.25	90.12	92.11	<b>95.81</b>
7	53.75	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	95.42	<b>100.00</b>	<b>100.00</b>	99.48	<b>100.00</b>	<b>100.00</b>
8	92.98	99.20	<b>99.20</b>	99.17	98.62	98.60	97.19	99.07	99.17	99.17
9	46.67	<b>95.14</b>	93.75	88.89	47.22	80.00	70.56	84.03	87.04	88.89
10	64.06	75.89	76.62	<b>89.98</b>	70.52	75.72	75.82	78.05	82.64	<b>89.98</b>
11	73.24	91.17	91.21	94.35	90.23	90.69	93.78	91.84	<b>96.64</b>	94.35
12	36.39	74.63	74.79	<b>86.33</b>	60.02	63.82	65.85	74.16	74.30	86.33
13	94.16	99.46	99.46	<b>99.52</b>	97.94	98.42	93.21	98.50	98.41	<b>99.52</b>
14	92.58	97.45	97.38	<b>99.53</b>	98.17	98.24	97.66	97.10	99.17	<b>99.53</b>
15	26.81	82.18	83.18	<b>85.11</b>	80.72	83.38	70.19	78.32	84.13	<b>85.11</b>
16	50.00	99.73	99.56	<b>100.00</b>	91.72	95.70	89.57	99.33	98.57	<b>100.00</b>
OA	68.54	85.40	85.84	<b>88.55</b>	81.35	83.11	84.62	84.07	86.93	<b>88.55</b>
AA	59.84	85.25	85.77	<b>89.81</b>	80.90	84.97	81.32	83.71	86.20	<b>89.81</b>
$\kappa$	63.92	83.28	83.79	<b>86.90</b>	78.73	80.76	82.35	81.73	85.00	<b>86.90</b>



**Figure 10.** Classification maps of the University of Pavia data obtained by different methods.

**Table 5.** Classification accuracy (%) of different methods for the University of Pavia data, bold values indicate the best result for a row.

Class	sIIIMTL <sub>3</sub>				3D-MPs					
	Spec	EMP	3D-DWT	3D-MPs	SVM	SVMCK	MTL	sIIIMTL <sub>1</sub>	sIIIMTL <sub>2</sub>	sIIIMTL <sub>3</sub>
1	68.90	87.91	88.13	<b>92.04</b>	86.62	89.64	87.20	86.07	91.25	<b>92.04</b>
2	<b>77.85</b>	67.59	70.63	69.31	69.31	68.13	73.70	65.15	67.74	69.31
3	51.60	65.60	73.70	<b>82.99</b>	72.46	74.85	73.56	66.22	71.80	<b>82.99</b>
4	86.85	63.94	82.21	<b>97.81</b>	72.06	83.58	80.29	79.96	88.19	<b>97.81</b>
5	72.19	<b>99.93</b>	<b>99.93</b>	79.33	59.70	76.58	60.59	99.85	98.74	79.33
6	38.38	72.94	61.13	<b>76.72</b>	55.90	63.59	58.62	75.98	69.85	76.72
7	89.77	93.08	<b>96.84</b>	93.01	96.69	96.54	91.20	95.34	94.74	93.01
8	41.01	78.79	61.68	74.99	63.25	63.82	69.07	<b>83.41</b>	74.96	74.99
9	59.13	97.99	95.67	99.26	96.09	98.63	96.20	<b>99.47</b>	99.37	99.26
OA	67.79	74.45	74.73	<b>78.62</b>	71.39	73.82	74.71	75.14	76.43	<b>78.62</b>
AA	65.08	80.86	81.10	<b>85.05</b>	74.68	79.48	76.71	83.49	84.07	<b>85.05</b>
$\kappa$	58.03	67.42	67.54	<b>73.02</b>	62.68	66.36	66.74	68.67	70.02	<b>73.02</b>



**Figure 11.** Classification maps of the Salinas data obtained by different methods.

**Table 6.** Classification accuracy (%) of different methods for the Salinas data, bold values indicate the best result for a row.

Class	sllMTL <sub>3</sub>				3D-MPs					
	Spec	EMP	3D-DWT	3D-MPs	SVM	SVMCK	MTL	sllMTL <sub>1</sub>	sllMTL <sub>2</sub>	sllMTL <sub>3</sub>
1	97.48	98.54	99.35	<b>99.90</b>	99.45	98.09	98.29	98.64	99.30	<b>99.90</b>
2	99.78	99.16	99.02	<b>100.00</b>	97.97	99.13	99.70	99.54	99.78	<b>100.00</b>
3	98.01	98.11	97.14	99.54	98.77	97.65	98.62	98.21	<b>99.64</b>	99.54
4	98.12	98.99	99.35	<b>99.71</b>	98.70	97.61	99.28	98.70	98.99	<b>99.71</b>
5	98.53	97.81	96.57	<b>98.83</b>	96.57	96.34	97.51	96.83	98.30	<b>98.83</b>
6	99.95	99.95	99.13	<b>99.97</b>	99.16	98.72	99.72	99.92	99.92	<b>99.97</b>
7	99.72	98.50	99.41	<b>99.86</b>	95.96	99.27	99.63	99.69	99.77	<b>99.86</b>
8	89.56	88.19	91.49	<b>93.85</b>	83.82	87.42	91.76	90.10	93.37	<b>93.85</b>
9	99.45	99.67	99.25	<b>99.90</b>	99.59	99.09	99.35	99.54	99.79	<b>99.90</b>
10	89.65	95.10	93.93	<b>98.34</b>	87.70	92.70	96.21	96.09	95.93	<b>98.34</b>
11	91.49	98.01	94.99	<b>99.62</b>	90.26	95.65	88.55	92.90	96.78	<b>99.62</b>
12	99.58	99.42	99.79	<b>100.00</b>	99.63	99.58	99.84	99.84	99.79	<b>100.00</b>
13	98.45	98.34	99.12	<b>99.45</b>	97.90	98.12	98.23	94.81	<b>99.45</b>	<b>99.45</b>
14	94.81	95.85	97.26	<b>99.34</b>	96.79	93.67	93.86	97.92	97.54	<b>99.34</b>
15	57.05	85.24	83.16	<b>86.46</b>	82.34	81.26	80.93	81.14	81.42	<b>86.46</b>
16	98.04	97.26	98.49	<b>99.22</b>	96.98	96.48	98.27	98.64	97.54	<b>99.22</b>
OA	90.67	94.54	94.80	<b>96.64</b>	92.33	93.34	94.66	94.43	95.46	<b>96.64</b>
AA	94.35	96.76	96.71	<b>98.37</b>	95.10	95.67	96.24	96.39	97.33	<b>98.37</b>
$\kappa$	89.57	93.93	94.20	<b>96.25</b>	91.45	92.59	94.05	93.79	94.94	<b>96.25</b>

Second, the 3D-based feature extraction methods (i.e., 3D-DWT and 3D-MPs) achieve comparable or better performance than the 2D-based one (i.e., EMP). For instance, as observed from Table 5, the OA, AA and  $\kappa$  of 3D-DWT are slightly better than the EMP, while the OA, AA and  $\kappa$  of 3D-MPs are 4.17%, 4.19%, 5.60% higher than the EMP, respectively. The classification results of the 3D-DWT and 3D-MPs are also comparable or better than the EMP in Tables 4 and 6. The reason for good results of 3D-DWT and 3D-MPs-based feature extraction methods is that they can respect the 3D nature of the HSI cube by taking the HSI cube as a whole entity to extract the spectral-spatial features.

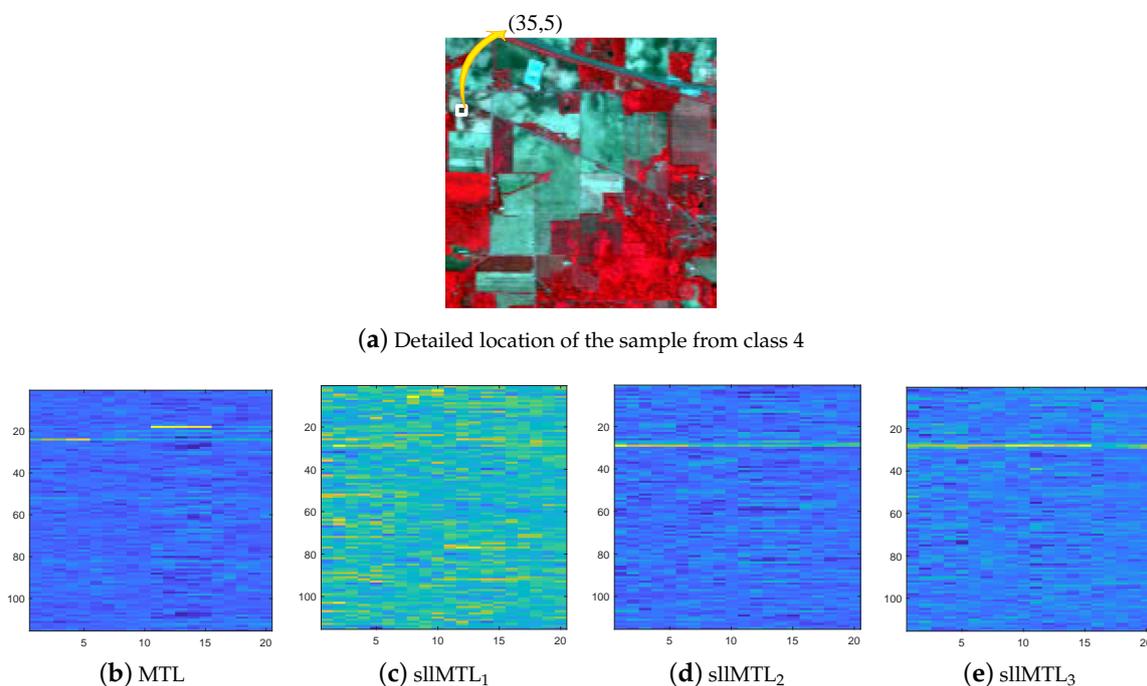
Third, the 3D-MPs yields the best classification performance in all the feature extraction methods by fixing the sllMTL<sub>3</sub> as classifier. For instance, in the Indian Pines data (see Table 4), the OA of 3D-MPs is 20.01%, 3.15% and 2.71% higher than those of the “Spec”, EMP and 3D-DWT, respectively. In the University of Pavia data (see Table 5), the OA of 3D-MPs is also 10.83%, 4.17% and 3.89% higher than those of the “Spec”, EMP and 3D-DWT, respectively. In the Salinas data (see Table 6), the OA of 3D-MPs is 5.97%, 2.10% and 1.84% higher than other feature extraction methods. Similarly, the AA and  $\kappa$  of the 3D-MPs are also much higher than its counterparts on the three HSI datasets. It demonstrates that multiple features obtained by varying the shapes and sizes of SEs can offer additional complementary information, and thus, effectively improve the performance of hyperspectral classification.

Fourth, the classification results of the SVM are inferior to other ones. For instance, it can be observed from Table 4 that the OA of SVM is about 2% to 7% lower than other classifiers (i.e., SVMCK, MTL, sllMTL<sub>1</sub>, sllMTL<sub>2</sub> and sllMTL<sub>3</sub>), the AA (or  $\kappa$ ) of SVM is also about 0.4% to 9% (or 2% to 8%) lower than its competitors. It is notable that the classification results of the University of Pavia data (see Table 5) and the Salinas data (see Table 6) also yield similar properties. This is due to the fact that the SVM cannot able to cope with the multiple-feature-based scenario but simple stacks the multiple 3D-MPs together to generate a single cube, while the other methods can make full use of the multiple 3D-MPs features by taking each 3D-MPs cube as the feature of a task.

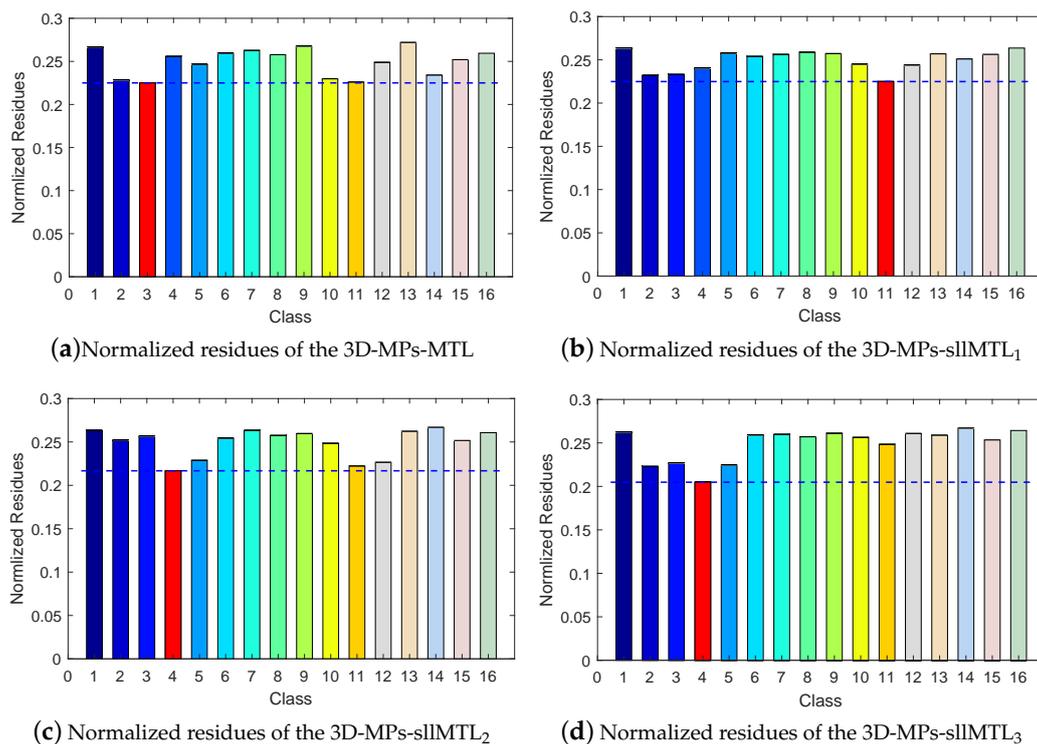
Fifth, we compare the multitask-based classifiers (i.e., MTL, sllMTL<sub>1</sub>, sllMTL<sub>2</sub> and sllMTL<sub>3</sub>) by utilizing the 3D-MPs-based features. To provide intuitive insight, the coefficient matrices of a test sample from class 4 (i.e., corn) located at (35,5) in the Indian Pines data are compared in Figure 12. The x-axis labels the task number and the y-axis labels the representation number. We can see from

Figure 12 that, for an instance in a certain class of data, a distinctive training feature is active in the  $sllMTL_2$  and  $sllMTL_3$ . The joint sparse and low-rank constraints in the  $sllMTL_2$  and  $sllMTL_3$  can effectively capture the task specificity and relatedness, and thus alleviate the misclassification problem. The normalized residues of the same sample from class 4 (i.e., corn) located at (35,5) are plotted in Figure 13, from which one can observe that, by utilizing the 3D-MPs features, the  $sllMTL_2$  and  $sllMTL_3$  yield more accurate classification results than the MTL and  $sllMTL_1$ . Moreover, it is shown in Tables 4–6 that the  $sllMTL_2$  and  $sllMTL_3$  outperform the MTL and  $sllMTL_1$ , and the  $sllMTL_3$  performs much better than the  $sllMTL_2$  since a Laplacian-like regularization is added in the  $sllMTL_3$  to take full advantage of the label information of training samples. In addition, as displayed in Figures 9–11, the classification maps of  $sllMTL_3$  are more close to the ground truth (see Figures 4a, 5b and 6b) than other methods.

Finally, it is worthwhile to note that the proposed method is much better than those in previous work (e.g., [17,34]). Although we use the same hyperspectral datasets as [17,34], the number of training and test samples are completely different. For instance, in the Indian Pines data, around 10% of the labeled samples are chosen for training in [17], while only 1% samples per class are selected as training samples in this paper. Therefore, it is not very surprising to see that the OA of our proposed method is much lower than the results shown in Table 2 of [17]. As can be seen in Figure 5 of [17], the OA is lower than 75% when only 1% labeled samples in each class are chosen as training samples, while the OA of 3D-MPs- $sllMTL_3$  achieves to 88.55% with the same number of training samples. Similarly, in the University of Pavia data, the number of training samples is 3921 in [34], while very limited training samples (1% samples from the available 3921 training samples are chosen for training, i.e., about 39 samples in total) are chosen in this paper. It is observed from Table 5 that the classification performance of 3D-MPs is better than the EMP with the same classifier.



**Figure 12.** The coefficient matrices of a test sample from class 4 (i.e., corn) located at (35,5) in the Indian Pines data for (b) MTL, (c)  $sllMTL_1$ , (d)  $sllMTL_2$ , and (e)  $sllMTL_3$ . The x-axis labels the task number and the y-axis labels the representation number.



**Figure 13.** Normalized residues of a sample from class 4 (i.e., corn) located at (35,5). (a–d) are the normalized residues of the 3D-MPs-MTL, 3D-MPs-sllMTL<sub>1</sub>, 3D-MPs-sllMTL<sub>2</sub>, and 3D-MPs-sllMTL<sub>3</sub>, respectively.

## 5. Discussion

### 5.1. Statistical Significance Analysis of the Results

To discuss whether the proposed method significantly outperforms its counterparts, we apply the McNemar’s test to evaluate the statistical significance in accuracy improvement. The McNemar’s test is a non-parametric test based upon the standardized normal test statistic

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \tag{17}$$

where  $f_{ij}$  indicates the number of samples classified correctly by the classifier  $i$  but incorrectly by the classifier  $j$ . This test allows two methods to be compared. In this paper, we compare the 3D-MPs-sllMTL<sub>3</sub> with the other methods one by one. If the absolute of the McNemar’s value  $Z$  is greater than 1.96 (i.e.,  $|Z| > 1.96$ ), the difference in accuracy between the  $i$ th and  $j$ th classifiers is considered statistically significant at the 5% level of significance. The average McNemar’s values  $Z$  of the proposed method and other methods by utilizing the afore-mentioned three experimental datasets are depicted in Figure 14. It is clearly shown in Figure 14 that the  $Z$  is larger than 1.96 in all cases, and therefore, the 3D-MPs-sllMTL<sub>3</sub> performs significantly better than the other methods according to the McNemar’s test.

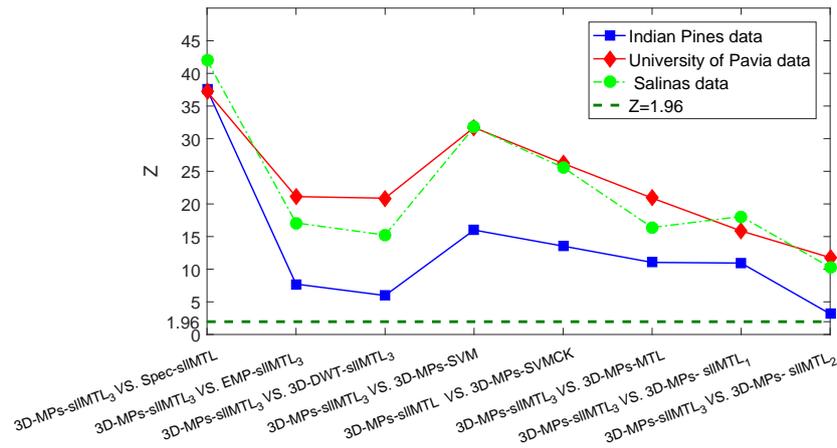


Figure 14. McNemar’s test between 3D-MPs-sllMTL<sub>3</sub> and other methods.

5.2. Sensitivity Analysis of the Parameters

Another important issue is to discuss the sensitivity of the parameters in the proposed 3D-MPs-sllMTL<sub>3</sub>. There are two key parameters in the 3D-MPs-sllMTL<sub>3</sub>:  $\lambda$  and  $\beta$ . The impacts of the parameters  $\lambda$  and  $\beta$  are shown in Figure 15, whose x-axis and y-axis are the value ranges of the corresponding parameters and the z-axis denotes the OA (%) of various hyperspectral datasets. It can be seen from Figure 15 that the classification accuracies for the three datasets generally improve slowly as the parameters  $\lambda$  and  $\beta$  increase, and then, begin to decrease after obtaining the maximum values. When the parameters  $\lambda$  and  $\beta$  are very small, the low-rank constraint term will dominate the optimization process and the sparse and Laplacian-like regularizations will be weakened. The influence of the sparse and Laplacian-like regularizations will become strong with the increase of  $\lambda$  and  $\beta$ . Based on the above analysis, it is better to make a tradeoff among the three constraints so as to achieve satisfactory classification performance.

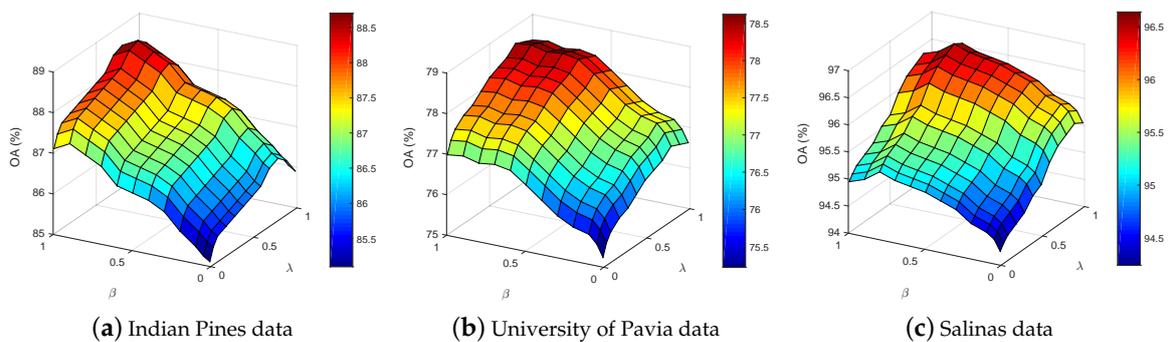


Figure 15. The impact of parameters  $\lambda$  and  $\beta$  on the OA in (a) Indian Pines data, (b) University of Pavia data and (c) Salinas data.

5.3. Influence Analysis of the 3D-SEs

Note that one can extract different structure features of the HSI by applying a series of 3D-SEs, it is interesting to analyze the influence of the 3D-SEs with different shapes and sizes. As an example, we discuss the influence of the 3D-SEs in the Indian Pines data. Both cube and sphere shapes with side length  $L = 5, 9, 13, 17$  and  $21$  are adopted in the experiments. We set  $K = 1$  in the sllMTL<sub>3</sub> and analyze the influence of the 3D-SEs by classifying each 3D-MP feature separately. The classification results of different 3D-MPs is shown in Figure 16, from which we can observe that the sphere(closing) with  $L = 5$  provides the lowest classification accuracy, while cube(opening) with  $L = 17$  achieves the highest

classification accuracy among all the 3D-MPs. However, it is shown in Table 4 that the combination of multiple 3D-MPs gives much better performance than using each 3D-MP feature separately. The reason is that a single 3D-MP can only reflect the characteristic of the HSI in one aspect, while multiple features can represent the characteristic of the HSI more effectively from different perspectives.

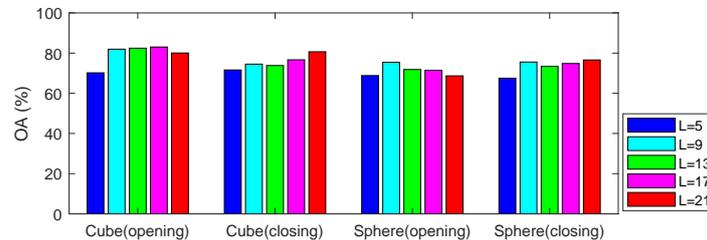


Figure 16. Classification results the sllMTL<sub>3</sub> by using different 3D-MPs.

## 6. Conclusions

In this paper, we have proposed a 3D-MPs-based MTL framework, namely, 3D-MPs-sllMTL, for hyperspectral imagery classification. On the one hand, the 3D-opening and 3D-closing operators are applied for feature extraction of the HSI cubes. Multiple SEs with different shapes and sizes are used to extract the spectral-spatial features from different aspects. Comparing against the vector/image-based methods in the literature, the proposed 3D-MPs-based method respects the 3D nature of the HSI by taking the HSI cube as a whole entity. On the other hand, the sllMTL method is proposed in this paper to exploit the joint sparse and low-rank structures by taking each 3D-MP as features of a specific task. Comparing against the existing MTL methods, the proposed sllMTL can capture both specificities and shared factors of the tasks by adopting the joint sparse and low-rank constraints. The Laplacian-like regularization can also improve the classification performance further by making full use of the class label information. We have validated the superiority of the 3D-MPs-sllMTL approach on three benchmark hyperspectral datasets and compared our method with a number of alternatives, i.e., the original “Spec”, EMP, 3D-DWT, SVM, SVMCK and MTL. The experimental results consistently demonstrate the effectiveness of the proposed method. Quantitatively, by utilizing the 3D-MPs features, the OA of sllMTL<sub>3</sub> is at least about 4%, 3% and 2% higher than the SVM, SVMCK and MTL, respectively. By utilizing the sllMTL<sub>3</sub>, the OA of 3D-MPs is at least about 6%, 2% and 2% higher than the “Spec”, EMP and 3D-DWT, respectively. Note that the sllMTL considers an equal contribution for any 3D-MPs-based features in the classification process, a future research direction is to investigate the importance of different features by assigning different weights to the features. Extending the sllMTL to its kernel versions is also a promising research topic for further improving the classification performance. Moreover, as to other kinds of images (e.g., the sequence of the 2D images or medical images), we can also take them as 3D cubes, whose 3D-MPs features can be extracted by the 3D MM. However, the sllMTL cannot be directly used for classification of those images because a pixel in those images usually comes from different classes. One should extract and arrange the features from the same class together before applying the sllMTL. Therefore, applying the proposed method to classify other images (e.g., the sequence of the 2D images or medical images) is also an interesting area of future work.

**Acknowledgments:** This work was supported by the National Natural Science Foundation of China under Grant 41501368, and the Fundamental Research Funds for the Central Universities under Grant 16lgpy04. The authors would like to thank Prof. D. Landgrebe from Purdue University for providing the AVIRIS image of Indian Pines and the Prof. Gamba from University of Pavia for providing the ROSIS data set. Last but not least, we would like to take this opportunity to thank the Editors and the Anonymous Reviewers for their detailed comments and suggestions, which greatly helped us to improve the clarity and presentation of our manuscript.

**Author Contributions:** All coauthors made significant contributions to the manuscript. Zhi He designed the research framework, analyzed the results and wrote the manuscript. Yiwen Wang and Jie Hu assisted in

the prepared work and validation work. Moreover, all coauthors contributed to the editing and review of the manuscript.

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

## References

1. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C.; et al. The EnMAP spaceborne imaging spectroscopy mission for Earth Observation. *Remote Sens.* **2015**, *7*, 8830–8857.
2. Angelliaume, S.; Ceamanos, X.; Viallefont-Robinet, F.; Baqué, R.; Déliot, P.; Miegbielle, V. Hyperspectral and Radar airborne imagery over controlled release of oil at sea. *Sensors* **2017**, *17*, 1772, doi:10.3390/s17081772.
3. Fu, Y.; Zhao, C.; Wang, J.; Jia, X.; Yang, G.; Song, X.; Feng, H. An improved combination of spectral and spatial features for vegetation classification in hyperspectral images. *Remote Sens.* **2017**, *9*, 261.
4. Schneider, S.; Murphy, R.J.; Melkumyan, A. Evaluating the performance of a new classifier—the GP-OAD: A comparison with existing methods for classifying rock type and mineralogy from hyperspectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2014**, *98*, 145–156.
5. Sun, B.; Kang, X.; Li, S.; Benediktsson, J.A. Random-walker-based collaborative learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 212–222.
6. Kang, X.; Xiang, X.; Li, S.; Benediktsson, J.A. PCA-based edge-preserving features for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 7140–7151.
7. Gao, L.; Zhao, B.; Jia, X.; Liao, W.; Zhang, B. Optimized kernel minimum noise fraction transformation for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 548.
8. Zhong, Y.; Jia, T.; Zhao, J.; Wang, X.; Jin, S. Spatial-spectral-emissivity land-cover classification fusing visible and thermal infrared hyperspectral imagery. *Remote Sens.* **2017**, *9*, 910.
9. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: New York, NY, USA, 2013.
10. Kuo, B.C.; Ho, H.H.; Li, C.H.; Hung, C.C.; Taur, J.S. A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 317–326.
11. Camps-Valls, G.; Gomez-Chova, L.; Muñoz-Marí, J.; Vila-Francés, J.; Calpe-Maravilla, J. Composite kernels for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 93–97.
12. Gu, Y.; Wang, C.; You, D.; Zhang, Y.; Wang, S.; Zhang, Y. Representative multiple kernel learning for classification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2852–2865.
13. Gurram, P.; Kwon, H. Sparse kernel-based ensemble learning with fully optimized kernel parameters for hyperspectral classification problems. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 787–802.
14. Maulik, U.; Chakraborty, D. Learning with transductive SVM for semisupervised pixel classification of remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2013**, *77*, 66–78.
15. Wang, L.; Hao, S.; Wang, Q.; Wang, Y. Semi-supervised classification for hyperspectral imagery based on spatial-spectral label propagation. *ISPRS J. Photogramm. Remote Sens.* **2014**, *97*, 123–137.
16. Wright, J.; Yang, A.; Ganesh, A.; Sastry, S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227.
17. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3973–3985.
18. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral image classification via kernel sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 217–231.
19. Zhang, H.; Li, J.; Huang, Y.; Zhang, L. A nonlocal weighted joint sparse representation classification method for hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2056–2065.
20. Srinivas, U.; Suo, Y.; Dao, M.; Monga, V.; Tran, T.D. Structured sparse priors for image classification. *IEEE Trans. Image Process.* **2015**, *24*, 1763–1776.
21. Dao, M.; Nguyen, N.H.; Nasrabadi, N.M.; Tran, T.D. Collaborative multi-sensor classification via sparsity-based representation. *IEEE Trans. Signal Process.* **2016**, *64*, 2400–2415.
22. Bian, X.; Chen, C.; Xu, Y.; Du, Q. Robust hyperspectral image classification by multi-layer spatial-spectral sparse representations. *Remote Sens.* **2016**, *8*, 985.

23. Zhang, S.; Li, S.; Fu, W.; Fang, L. Multiscale superpixel-based sparse representation for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 139.
24. Xue, Z.; Du, P.; Su, H.; Zhou, S. Discriminative sparse representation for hyperspectral image classification: a semi-supervised perspective. *Remote Sens.* **2017**, *9*, 386.
25. Dao, M.; Kwan, C.; Koperski, K.; Marchisio, G. A joint sparsity approach to tunnel activity monitoring using high resolution satellite images. In Proceedings of the IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, NY, USA, 19–21 October 2017; pp. 322–328.
26. Ul Haq, Q.S.; Tao, L.; Sun, F.; Yang, S. A fast and robust sparse approach for hyperspectral data classification using a few labeled samples. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2287–2302.
27. He, Z.; Wang, Q.; Shen, Y.; Sun, M. Kernel sparse multitask learning for hyperspectral image classification with empirical mode decomposition and morphological wavelet-based features. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5150–5163.
28. Li, J.; Zhang, H.; Zhang, L.; Huang, X.; Zhang, L. Joint collaborative representation with multitask learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5923–5936.
29. Li, J.; Zhang, H.; Zhang, L. A nonlinear multiple feature learning classifier for hyperspectral images with limited training samples. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2015**, *8*, 2728–2738.
30. Zhang, Y.; Wu, K.; Du, B.; Zhang, L.; Hu, X. Hyperspectral target detection via adaptive joint sparse representation and multi-task learning with locality information. *Remote Sens.* **2017**, *9*, 482.
31. Jia, S.; Deng, B.; Zhu, J.; Jia, X.; Li, Q. Superpixel-based multitask learning framework for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2575–2588.
32. Kang, X.; Li, S.; Benediktsson, J.A. Spectral-spatial hyperspectral image classification with edge-preserving filtering. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2666–2677.
33. Benediktsson, J.; Palmason, J.; Sveinsson, J. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 480–491.
34. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.A.; Chanussot, J.; Tilton, J.C. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE* **2013**, *101*, 652–675.
35. Hou, B.; Huang, T.; Jiao, L. Spectral-spatial classification of hyperspectral data using 3-D morphological profile. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2364–2368.
36. Lin, Z.; Liu, R.; Su, Z. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*; Curran Associates, Inc.: Granada, Spain, 2011; pp. 612–620.
37. Qian, Y.; Ye, M.; Zhou, J. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2276–2291.



© 2018 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/>).