

Article

Multiscale Superpixel-Based Sparse Representation for Hyperspectral Image Classification[†]

Shuzhen Zhang ^{1,2}, Shutao Li ^{1,*}, Wei Fu ¹ and Leiyuan Fang ¹

¹ College of Electrical and Information Engineering, Hunan University, Changsha 418002, China; shuzhen_zhang@hnu.edu.cn (S.Z.); fuweihunandaxue@gmail.com (W.F.); fangleyuan@gmail.com (L.F.)

² School of Information Science and Engineering, Jishou University, Jishou 416000, China

* Correspondence: shutao_li@hnu.edu.cn; Tel.: +86-731-8882-2866

[†] This paper is an extended version of our paper published in 2016 IEEE International Geoscience and Remote Sensing Symposium.

Academic Editors: Qi Wang, Nicolas H. Younan, Carlos López-Martínez, Lenio Soares Galvao and Prasad S. Thenkabail

Received: 30 November 2016; Accepted: 25 January 2017; Published: 7 February 2017

Abstract: Recently, superpixel segmentation has been proven to be a powerful tool for hyperspectral image (HSI) classification. Nonetheless, the selection of the optimal superpixel size is a nontrivial task. In addition, compared with single-scale superpixel segmentation, the same image segmented on a different scale can obtain different structure information. To overcome such a drawback also utilizing the structural information, a multiscale superpixel-based sparse representation (MSSR) algorithm for the HSI classification is proposed. Specifically, a modified segmentation strategy of multiscale superpixels is firstly applied on the HSI. Once the superpixels on different scales are obtained, the joint sparse representation classification is used to classify the multiscale superpixels. Furthermore, majority voting is utilized to fuse the labels of different scale superpixels and to obtain the final classification result. Two merits are realized by the MSSR. First, multiscale information fusion can more effectively explore the spatial information of HSI. Second, in the multiscale superpixel segmentation, except for the first scale, the superpixel number on a different scale for different HSI datasets can be adaptively changed based on the spatial complexity of the corresponding HSI. Experiments on four real HSI datasets demonstrate the qualitative and quantitative superiority of the proposed MSSR algorithm over several well-known classifiers.

Keywords: multiscale superpixels; sparse representation; hyperspectral image; spectral-spatial image classification

1. Introduction

A hyperspectral sensor can capture hundreds of narrow contiguous spectral bands from the visible to infrared spectrum for each image pixel. Therefore, hyperspectral images consist of rich spectral-spatial information, which have attracted great attention in different application domains, such as national defense [1], urban planning [2], precision agriculture [3,4] and environment monitoring [5–7].

In the last few decades, HSI classification has been an important issue in remote sensing. In the earlier research, different pixel-wise approaches have been developed [8–11]. However, without considering the spatial information, the obtained classification results by these approaches usually contain much noise. To further improve the classification performance, methods incorporating the spatial information of the HSI have been proposed recently. In these methods, pixels in a small region are assumed to belong to the same material and have similar spectral properties. Various texture feature extraction methods used in the traditional two-dimensional images have been

extend to the HSI for improving the classification performance, such as the Gabor filter [12], the local binary pattern (LBP) filter [13], the edge-preserving filter (EPF) [14], the two-dimensional Gaussian derivative (GD) filter [15] and the extended morphological profiles (EMPs) filter [16,17]. In addition, to exploit the nonlinearity information, kernel technique has been widely used in HSI. For instance, the generalized composite kernel [18], graphic kernel [19], spatial-spectral derivative-aided kernel [20] and probabilistic kernel [21] have been introduced to the HSI classification. Furthermore, HSI can be regarded as “cube” data. In this case, the tensor-based classification methods have been used for the HSI [22–24]. Moreover, with the rise of deep learning, spectral-spatial information-based deep learning algorithms [25–27] are also applied in the HSI classification, which can extract potential and invariant features of the HSI.

In light of the operating mechanism of human vision, the key information of a nature image can be captured by learning a sparse coding. The sparse representation technique has been developed according to the above-mentioned theory. In the last few years, this technique has been extremely employed in the computer vision domain, such as face recognition [28], feature matching [29,30] and image fusion [31]. In these different signal processing tasks, state-of-the-art performance is usually obtained. Recently, sparse representation classification (SRC) has also attracted much attention for the classification of the HSI [27,32–35]. The SRC assumes that a test pixel can be approximately represented by a linear combination of all training samples. The class label of the test pixel is determined by which class leads to the minimum reconstruction error. For the pixel-wise SRC, the HSI classification result usually appears very noisy. To gain better classification accuracies, Chen et al. [36] proposed a joint sparse representation classification (JSRC) for the HSI. The JSRC assumes that pixels in a fixed window belong to the same class, and thus, these pixels can be simultaneously represented by a set of common atoms of a training dictionary. In the past few years, several modified versions of the JSRC have been proposed [20,37–41]. Although these approaches obtain improved performance, the neighborhood of the test pixel is a fixed square window. That is, for each test pixel in these methods, if pixels locate at the image edges or in a detailed region, the neighborhood may contain pixels from different classes, and the classification results are usually unsatisfactory. Therefore, to solve the aforementioned problem, the shape of the regions should be adaptively changed according to the different spatial information of the HSI.

In the image processing field, various superpixel segmentation methods have been widely used [42–44]. The superpixel has also been introduced for the HSI classification in recent years [45–49]. Each superpixel of the image is an adaptive segmentation region according to the spatial structure. Therefore, it can effectively exploit the spatial information compared with the fixed window centered at the test pixel. Meanwhile, developing classification methods in a superpixel-by-superpixel manner has lower computational complexity than the pixel-wise approaches. However, for single-scale superpixel based algorithms, the accuracy of the superpixel segmentation will directly affect the final results [45,46,50]. Therefore, the choice of the superpixel size is important. However, it is not a trivial work to choose the optimal superpixel size. The reason is that the small size may be short of enough information and, the large size may result in the error segmentation. In fact, for single-scale superpixel segmentation, some mixed superpixels that consist of pixels from different classes will still exist in the segmentation image. In addition, for the same region of an image, different structural information can be explored by segmenting superpixels on different scales. In view of this reason, multiscale superpixel-based methods are used for feature representation, target detection and recognition in some very recent works [50–52]. For different applications, the superpixel information of different scales is usually integrated via different strategies, such as adopting the similarity between a pixel and the average of pixels within the superpixel [50], converting to a sparse constraint problem [51] and utilizing the convolutional neural network (CNN) [52]. These methods can effectively integrate multiscale information to obtain the optimal result.

In this paper, a modified segmentation strategy of multiscale superpixels is proposed. In the strategy, the number of each scale superpixel is related with the complexity of the first principal component of the HSI. Adopting the segmentation strategy, a multiscale superpixel-based sparse representation (MSSR) algorithm is proposed.

The rest of this paper is organized as follows. In Section 2, the JSRC algorithm for the HSI classification is briefly introduced. The proposed MSSR for the HSI classification is detailed in Section 3. In Section 4, the experimental results and discussions are given. Finally, in Section 5, the paper is summarized, and the future works are suggested.

2. JSRC Algorithm for HSI Classification

For HSI, pixels in a fixed window are assumed to come from the same ground materials and share the same spectral characteristics. According to the sparse representation theory, the correlations among the pixels within the window can be represented by the joint sparse regularization. Specifically, we denote one pixel in a HSI with B bands as $\mathbf{y}_c \in R^B$ and pixels in the $\sqrt{p} \times \sqrt{p}$ window as $\mathbf{Y}_c = [\mathbf{y}_{c_1}, \mathbf{y}_{c_2}, \dots, \mathbf{y}_{c_p}] \in R^{B \times p}$. Let $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_M] \in R^{B \times T}$ represent the structure dictionary with T training samples from M distinct classes, where $\{\mathbf{D}_j\}_{j=1,2,\dots,M}$ are sub-dictionaries. Let t_j be the number of training samples from the j -th class, and $\sum_{j=1}^M t_j = T$. Then, pixels \mathbf{Y}_c in the window can be represented as:

$$\mathbf{Y}_c = \mathbf{D}\mathbf{A} + \mathbf{N} \quad (1)$$

where \mathbf{N} is the possible noise and $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M]$ is the sparse coefficients matrix of \mathbf{Y}_c . The vector $\{\mathbf{A}_j\}_{j=1,2,\dots,M}$ is the corresponding component of sub-dictionary $\{\mathbf{D}_j\}_{j=1,2,\dots,M}$ in \mathbf{A} .

According to the JSRC algorithm, the sparse regularization places a $l_{row,0}$ -norm on the sparse matrix \mathbf{A} , which means to select a number of the most representative nonzero rows in \mathbf{A} . The joint sparse matrix \mathbf{A} can be obtained by solving the following optimization problem:

$$\hat{\mathbf{A}} = \arg \min \|\mathbf{Y}_c - \mathbf{D}\mathbf{A}\|_2 \quad s.t. \|\mathbf{A}\|_{row,0} \leq K \quad (2)$$

where K represents the sparsity level. The simultaneous orthogonal matching pursuit (SOMP) algorithm [53] can efficiently solve (2). After the sparse coefficients matrix $\hat{\mathbf{A}}$ is obtained, the class label of \mathbf{Y}_c is determined by the minimum residual error:

$$Class(\mathbf{Y}_c) = \arg \min_{j=1,2,\dots,M} E_j(\mathbf{Y}_c) \quad (3)$$

where $E_j(\mathbf{Y}_c) = \|\mathbf{Y}_c - \mathbf{D}_j \cdot \hat{\mathbf{A}}_j\|_2$, $j = 1, 2, \dots, M$ is the corresponding reconstruction residual error of the j -th class.

3. Proposed MSSR for HSI Classification

Compared with the fixed-shape neighborhood in the JSRC method, a superpixel is an adaptively spatial region, which is beneficial to obtain a better classification performance [45–47]. However, as previously mentioned, it is difficult to determine the optimal superpixel size. Meanwhile, the land covers in HSI have very complex structures with different sizes. Therefore, the multiscale superpixel-based approach is applied in the MSSR algorithm, which can more effectively exploit the spatial information of the HSI. The proposed MSSR algorithm for HSI classification consists of three parts: (1) the generation of multiscale superpixels in HSI; (2) the sparse representation for HSI with multiscale superpixels; (3) the fusion of multiscale classification results. The algorithmic schematic is demonstrated in Figure 1, and the detailed description is given below.

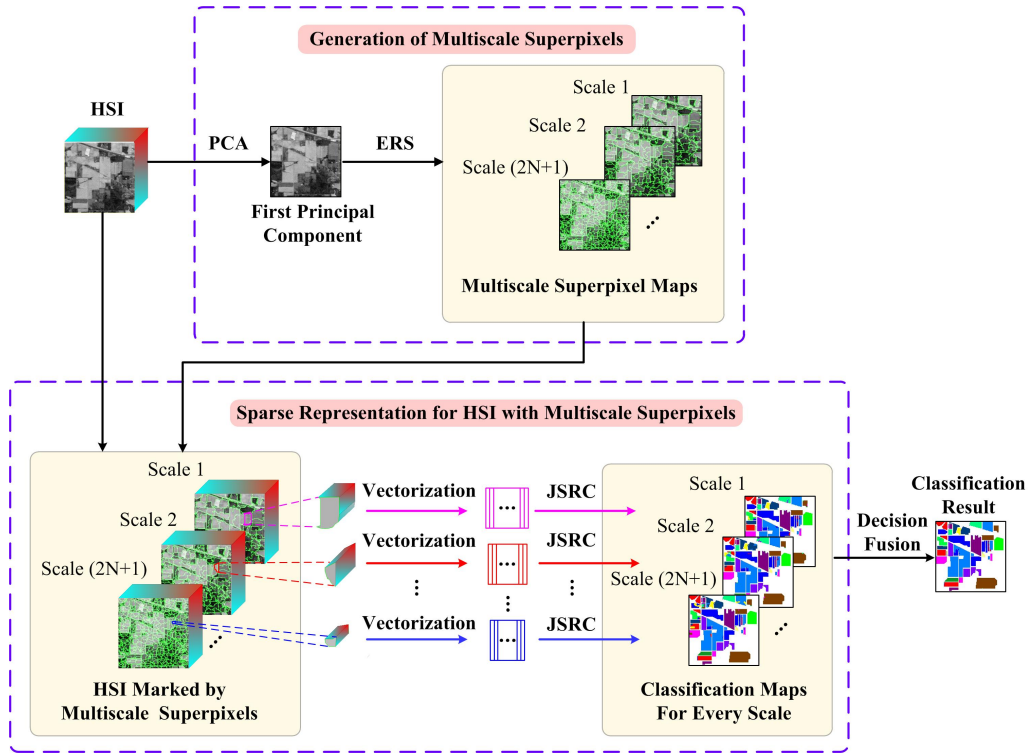


Figure 1. Schematic illustration of the multiscale superpixel-based sparse representation (MSSR) algorithm for HSI classification.

3.1. Generation of Multiscale Superpixels in HSI

As shown in Figure 1, to reduce the computational cost, the principal component analysis (PCA) algorithm [54] is firstly applied on the original HSI. Since the first principal component contains the major information of the HSI, we denote it as a fundamental image. Then, multiscale superpixel segmentation is applied on the fundamental image. For the multiscale superpixel segmentation, let \mathbf{F} represent the fundamental image. Let S_n denote the number of superpixels in the n -th scale. Let \mathbf{Y}_k^n represent the k -th superpixel in the n -th scale. Then, the fundamental image \mathbf{F} can be described as:

$$\mathbf{F} = \bigcup_{k=0}^{S_n} \mathbf{Y}_k^n, (n = 0, \pm 1, \pm 2, \dots, \pm N) \text{ and } \mathbf{Y}_k^n \cap \mathbf{Y}_g^n = \phi, (k \neq g) \quad (4)$$

In terms of Equation (4), the total number of superpixel scales is $(2N + 1)$.

In general, the more complicated structure of the fundamental image is, the greater the number of segmented superpixels should be. Therefore, in the MSSR algorithm, we connect the number of superpixels in the n -th scale S_n with the complexity of the fundamental image. To be specific, the Canny operator [55] is applied for the fundamental image \mathbf{F} to gain the corresponding edge image. The edge ratio C [56], which is the proportion of nonzero pixels accounting for the total pixels in the edge image, reflects the complexity of the fundamental image. Then, S_n is defined as:

$$S_n = 2^{n/2} \times S_f \times C \quad (n = 0, \pm 1, \pm 2, \dots, \pm N) \quad (5)$$

where S_f is the fundamental number of superpixels, which is empirically selected. In general, the more complicated the fundamental image is, the larger the value of S_f should be. In addition, in terms of Equation (5), when the fundamental image is more complicated, the number of superpixels in the same scale is also larger. Therefore, the step length variation among multiscale superpixels is related with the complexity of the fundamental image. At the same time, the variation range of multiscale superpixels is also connected with the image complexity. It should be noted that other

advanced methods depicting the image complexity might be applied to enhance the performance, but that will increase the computation amount [57,58].

According to the number of superpixels in the n -th scale S_n , a graph-based segmentation algorithm is used to generate the n -th scale superpixel segmentation result. Graph-based image segmentation algorithms are widely used in superpixel segmentation [59–61]. Among these, the entropy rate superpixel (ERS) [61] segmentation method has been demonstrated to be very efficient. Specifically, the fundamental image \mathbf{F} is firstly mapped to a graph $G = (V, E)$, where V is the vertex set denoting pixels of the fundamental image and E is the edge set representing the pairwise similarities given in the form of a similarity matrix. In the ERS, for the n -th scale of superpixel segmentation, the graph is partitioned into connected S_n subgraphs by choosing a subset of edges $A_n \subseteq E$. To obtain the compact and homogeneous superpixels, an entropy rate term $H_n(A_n)$ is adopted. Meanwhile, a balancing term $B_n(A_n)$ is utilized to enable superpixels with similar sizes. Therefore, the objective function of the ERS method is given by:

$$\max_{A_n} \{H_n(A_n) + \omega_n B_n(A_n)\} \quad s.t. A_n \subseteq E \quad (6)$$

where $\omega_n \geq 0$ is the weight of the balancing term. As described in [62], a greedy algorithm effectively solves the optimization problem in (6). After multiscale superpixel segmentation, for each test pixel, there are $(2N + 1)$ corresponding superpixels, which incorporate the test pixel. Then, the spatial information of each superpixel will be combined with the spectral information of pixels within the superpixel for HSI classification. Therefore, for each test pixel, there will be $(2N + 1)$ classification results.

3.2. Sparse Representation for HSI with Multiscale Superpixels

Multiscale superpixel segmentation results combine the original HSI to acquire a group of HSI marked with multiscale superpixels. Therefore, there are $(2N + 1)$ different marked regions corresponding to each test pixel in the HSI. For pixels within each region, they are supposed to have similar spectral characteristics. Hence, these pixels are simultaneously represented by a few common atoms from a structure dictionary. Assume the superpixel \mathbf{Y}_k^n contains p spectral pixels, i.e., $\mathbf{Y}_k^n = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p] \in R^{B \times p}$. Let $\mathbf{A}_k^n = [\mathbf{A}_{k_1}^n, \mathbf{A}_{k_2}^n, \dots, \mathbf{A}_{k_M}^n]$ be the sparse coefficients matrix of \mathbf{Y}_k^n and $\{\mathbf{A}_{k_j}^n\}_{j=1,2,\dots,M}$ is the corresponding component of sub-dictionary $\{\mathbf{D}_j\}_{j=1,2,\dots,M}$ in \mathbf{A}_k^n . The joint sparse matrix \mathbf{A}_k^n can be obtained by applying (2):

$$\hat{\mathbf{A}}_k^n = \arg \min \|\mathbf{Y}_k^n - \mathbf{D}\mathbf{A}_k^n\|_2 \quad s.t. \|\mathbf{A}_k^n\|_{row,0} \leq K \quad (7)$$

The reconstruction residual error of each class can be described as:

$$E_j(\mathbf{Y}_k^n) = \|\mathbf{Y}_k^n - \mathbf{D}_j \cdot \hat{\mathbf{A}}_{k_j}^n\|_2, \quad j = 1, 2, \dots, M \quad (8)$$

The class label of \mathbf{Y}_k^n is represented as:

$$Class(\mathbf{Y}_k^n) = \arg \min_{j=1,2,\dots,M} E_j(\mathbf{Y}_k^n) \quad (9)$$

3.3. Fusion of Multiscale Classification Results

For each test pixel, the class labels of corresponding multiscale superpixels may be different. That is, there are $(2N + 1)$ different classification results for an HSI. For these multiscale classification results, a quick and effective decision fusion strategy (i.e., the majority voting) is utilized to obtain the final classification result. Specifically, assume the class labels of a test pixel under different scales

respectively are $l_1, l_2, \dots, l_{2N+1}$. We count the number of each class occurrence, and denote them as L_1, L_2, \dots, L_M , where $(2N + 1) = \sum_{j=1}^M L_j$. The class label of the test pixel can be obtained by:

$$L_y = \arg \max(L_1, L_2, \dots, L_M) \quad (10)$$

4. Experimental Results and Discussion

In this section, the effectiveness of the proposed MSSR algorithm is tested in the classification of four hyperspectral datasets, i.e., the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines image, the Reflective Optics System Imaging Spectrometer (ROSIS-03) University of Pavia image, the AVIRIS Salinas image and the Hyperspectral Digital Image Collection Experiment (HYDICE) Washington DC image. The performance of the proposed MSSR algorithm is compared with those of seven competing classification algorithms, i.e., SVM [8], EMP [16], SRC [36], JSRC [36], multiscale adaptive sparse representation (MASR) [63], superpixel-based classification via multiple kernels (SCMK) [46] and the superpixel-based discriminative sparse model (SBDSM) [64]. The EMP, JSRC, MASR, SCMK, SBDSM and MSSR algorithms take advantage of the spectral-spatial information for HSI classification, while the SVM and SRC algorithms only exploit the spectral information. It should be noted that, for the SBDSM algorithm, the sparse dictionary is built by directly extracting pixels from HSI. Therefore, compared with the MSSR algorithm, the SBDSM algorithm is based on the single-scale superpixel and sparse representation.

4.1. Datasets Description

The Indian Pines image was acquired by the AVIRIS sensor over the agricultural Indian Pines site in northwestern Indiana. The size of this image is $145 \times 145 \times 220$, where 20 water absorption bands are discarded. The spatial resolution of the image is 20 m per pixel and the spectral coverage ranges from 0.2 to $2.4 \mu\text{m}$. The reference of this image contains sixteen classes, most of which are different kinds of crops. Figure 2 demonstrates the false-color composite of the Indian Pines image and the corresponding reference data.

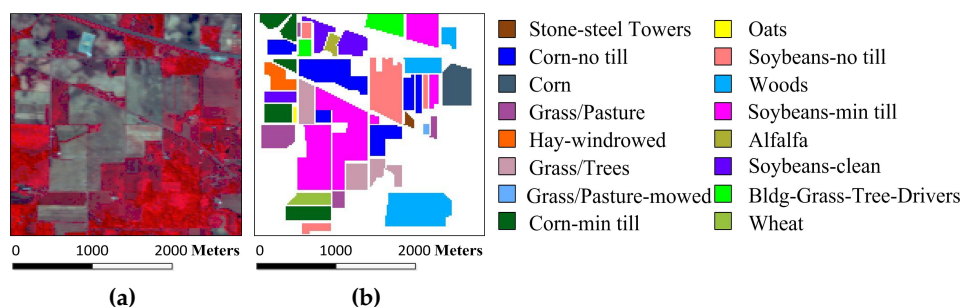


Figure 2. Indian Pines image: (a) false-color image; and (b) reference image.

The University of Pavia image was captured by the ROSIS-03 sensor over an urban area surrounding the University of Pavia, Italy. The ROSIS-03 sensor generated an image with a geometric resolution of 1.3 m per pixel and the spectral coverage ranging from 0.43 to $0.86 \mu\text{m}$. This image is of a size of $610 \times 340 \times 120$, where 12 spectral bands are removed due to high noise. The reference of this image contains nine ground-truth classes. Figure 3 shows the false-color composite of the University of Pavia image and the corresponding reference data.

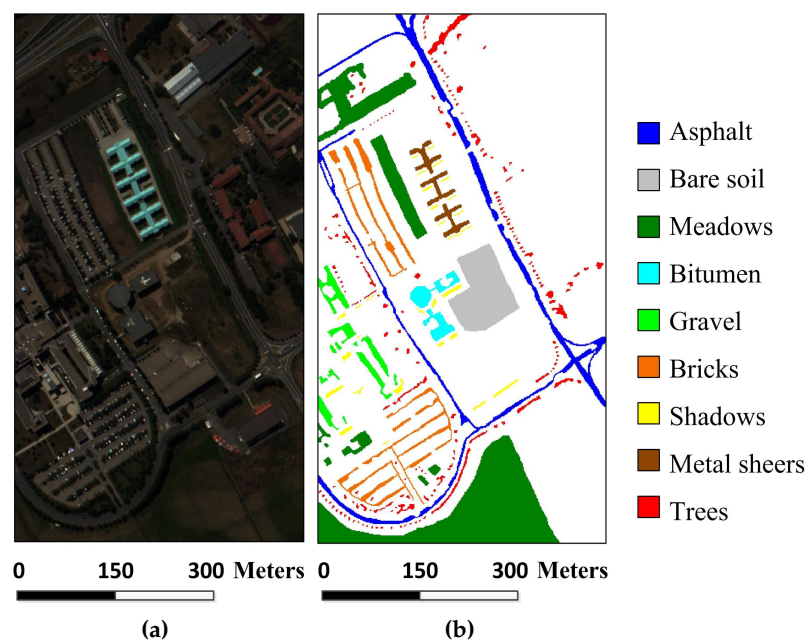


Figure 3. University of Pavia image: (a) false-color image; and (b) reference image.

The Salinas image was captured by the AVIRIS sensor over Salinas Valley, California. The image consists of 512×217 pixels and 224 spectral bands, where 20 water absorption spectral bands were removed. The geometric resolution of this image is 3.7 m. The reference of this image contains sixteen ground-truth classes. Figure 4 shows the false-color composite of the Salinas image and the corresponding reference data.

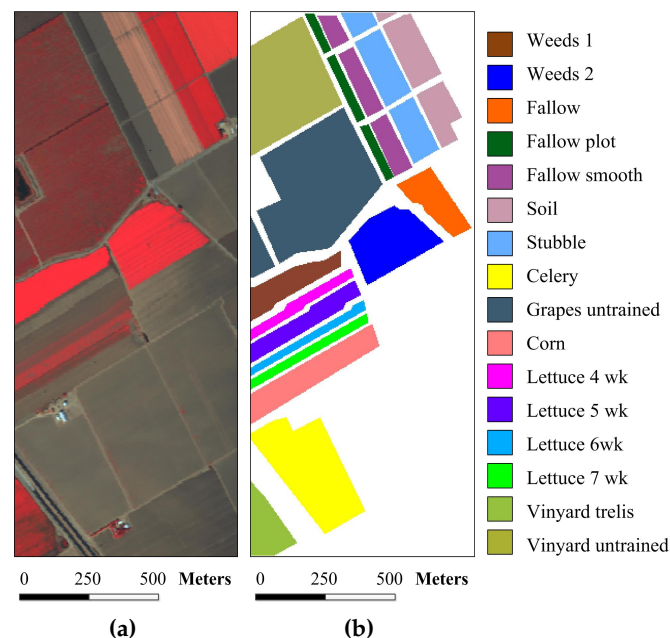


Figure 4. Salinas image: (a) false-color image; and (b) reference image.

The Washington DC image was recorded by the HYDICE sensor over the Washington DC Mall. The image consists of 280×307 pixels, each pixel including 210 spectral bands. The spectral coverage

ranges from 0.4 to 2.5 μm and the spatial resolution of the image is 3 m per pixel. In the experiments, bands ranging from 0.9 to 1.4 μm , where the atmosphere of these bands is opaque, are discarded from the dataset, leaving 191 bands. Figure 5 demonstrates the false-color composite of the Washington DC image and the corresponding reference data, which considers six classes of interest.

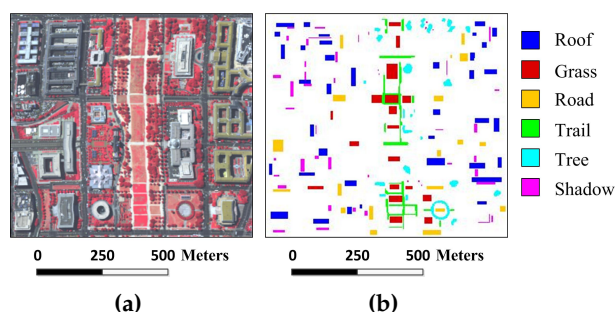


Figure 5. Washington DC image: (a) false-color image; and (b) reference image.

4.2. Comparison of Results

In the experiments, the SVM algorithm adopting a spectral Gaussian kernel is implemented with the LIBSVM [65] package, which is accelerated with Visual C++ software (6.0 Version). The parameters C and σ of the SVM are obtained by ten-fold cross validation. For the EMP algorithm, the parameters of feature extraction are set to the default in [16]. Once these morphological features are acquired, an SVM classifier is applied for the HSI classification. For the MASR and SCMK algorithms, the parameters are set to the same in [46,63], respectively. The parameters for the SRC and JSRC algorithms are tuned to reach the best results in these experiments. For the MSSR algorithm, the fundamental number of superpixel S_f is set to 3200, 3200, 1600 and 12,800 for the Indian Pine, University of Pavia, Salinas and Washington DC images, respectively. The number of multiscales for the four images is respectively set to 7, 7, 5 and 11. For the SBDSM algorithm, the number of superpixels is obtained by applying Equation (5), in which the fundamental numbers of superpixels for the four images are the same values in the MSSR algorithm, and the power exponent n in Equation (5) is set to zero. In the following subsection, the parameters of the proposed MSSR algorithm and the parameters of the SBDSM algorithm will be further analyzed. In addition, different algorithms are compared based on the overall accuracy (OA), average accuracy (AA) and kappa coefficient. These quantitative values of each algorithm were averaged over ten runs to diminish the possible bias.

In the experiment on the Indian Pines image, 10% of the labeled samples for each class are randomly selected as the training set and the remainder as the test set (see Table 1). Figures 6 and 7, respectively, show the superpixel segmentation maps and classification maps under different single scales. In the two figures, the number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 3200 and the power exponent n is an integer changing from -3 to 3 . Table 2 lists the quantitative values under different scales. Obviously, different scales yield different performances for different classes. For example, when the power exponent is set to -3 , the value of OA is minimum. However, for this scale, the sixth, eighth, ninth and thirteenth classes, the optimal classification performances are obtained. Conversely, although the optimal OA is acquired when the power exponent n is zero, the classification accuracy of the third class at this scale is minimum. These results demonstrate that, for the HSI classification, the optimal single scale is not suitable for all of the spatial structure regions; multiscale information fusion may be a better approach. In addition, the classification maps from the compared classifiers are illustrated in Figure 8, and the quantitative results are tabulated in Table 3. As can be seen, the SVM and SRC algorithms, which only consider the spectral information, deliver the classification maps with much noise. Meanwhile,

the spectral-spatial-based classification algorithms (EMP, JSRC, MASR, SCMK, SBDSM and MSSR) significantly outperform the pixel-wise algorithms. Compared with the SBDSM algorithm, the MSSR algorithm achieves more accurate estimations in the detailed area. This result indicates that the multiscale strategy can overcome the problem of the non-uniformity of spatial structure. At the same time, the problem bringing from the existing mixed superpixels of the single-scale superpixel segmentation can be well solved. Moreover, in terms of OA, AA and the kappa coefficient, the proposed MSSR algorithm also outperforms the other compared algorithms.

Table 1. Number of training and test samples of sixteen classes in the Indian Pines image.

Class	Name	Train	Test
1	Alfalfa	5	41
2	Corn-no till	143	1285
3	Corn-min till	83	747
4	Corn	24	213
5	Grass/pasture	49	434
6	Grass/tree	73	657
7	Grass/pasture-mowed	3	25
8	Hay-windrowed	48	430
9	Oats	2	18
10	Soybean-no till	98	874
11	Soybean-min till	246	2209
12	Soybean-clean till	60	533
13	Wheat	21	184
14	Woods	127	1138
15	Bldg-grass-trees-drives	39	347
16	Stone-steel towers	10	83
Total		1031	9218

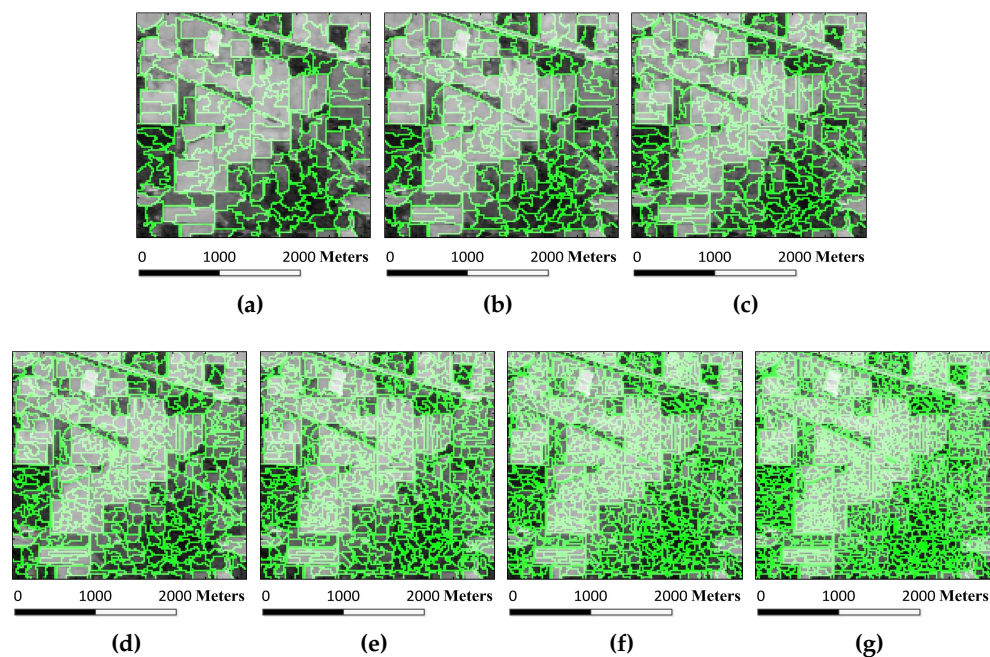


Figure 6. Superpixel segmentation results of the Indian Pines image under different scales. The number of single-scale superpixels is gained by using the Equation (5), in which the fundamental superpixel number is set to 3200 and the power exponent n is an integer changing from -3 to three: (a) $n = -3$; (b) $n = -2$; (c) $n = -1$; (d) $n = 0$; (e) $n = 1$; (f) $n = 2$; and (g) $n = 3$.

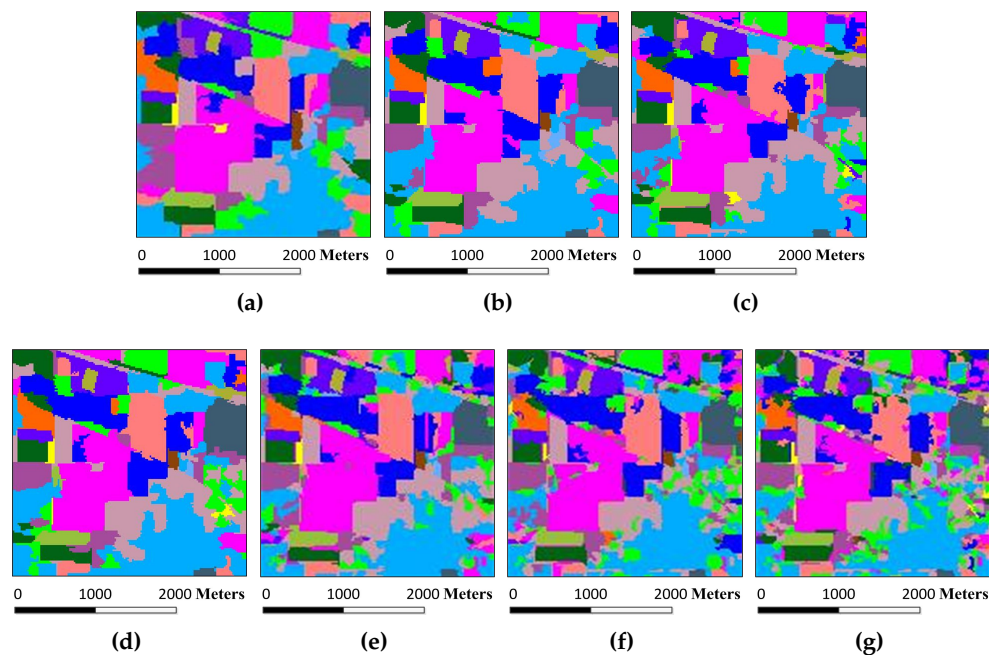


Figure 7. Classification results of the Indian Pines image under different scales. The number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 3200 and the power exponent n is an integer changing from -3 to three: (a) $n = -3$, OA = 93.14%; (b) $n = -2$, OA = 96.42%; (c) $n = -1$, OA = 96.62%; (d) $n = 0$, OA = 97.08%; (e) $n = 1$, OA = 95.64%; (f) $n = 2$, OA = 95.61%; and (g) $n = 3$, OA = 93.65%.

Table 2. Classification accuracy of the Indian Pines image under different scales. The number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 3200 and the power exponent n is an integer changing from -3 to three. Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface. AA, average accuracy.

Class	$n = -3$	$n = -2$	$n = -1$	$n = 0$	$n = 1$	$n = 2$	$n = 3$
1	99.01	98.05	99.02	97.56	98.05	97.56	90.73
2	87.21	90.27	95.04	96.34	90.41	92.54	90.33
3	92.05	95.53	97.11	89.69	95.61	92.34	92.88
4	95.31	99.06	99.04	99.06	88.64	91.55	87.98
5	91.71	94.10	94.79	93.55	94.47	94.01	92.30
6	99.88	99.76	99.76	99.85	99.30	98.48	97.53
7	96.00	96.00	97.60	96.43	96.00	96.80	96.00
8	100	100	100	100	100	98.79	98.28
9	100	100	100	100	100	94.44	86.67
10	90.48	96.84	92.20	94.74	93.91	94.74	93.46
11	93.86	97.32	96.51	96.92	97.71	96.43	95.15
12	82.66	92.35	95.20	97.75	90.73	91.22	84.80
13	100	99.46	99.57	99.46	99.46	99.13	99.46
14	99.74	99.93	99.74	99.82	99.46	98.54	98.98
15	93.60	98.04	91.87	93.08	95.33	93.03	83.80
16	96.63	96.39	97.59	96.39	97.11	97.56	91.57
OA (%)	93.37	96.44	96.56	96.92	95.77	95.32	93.62
AA (%)	94.88	95.07	95.19	95.19	95.01	95.45	92.50
Kappa	0.92	0.95	0.95	0.95	0.95	0.95	0.93

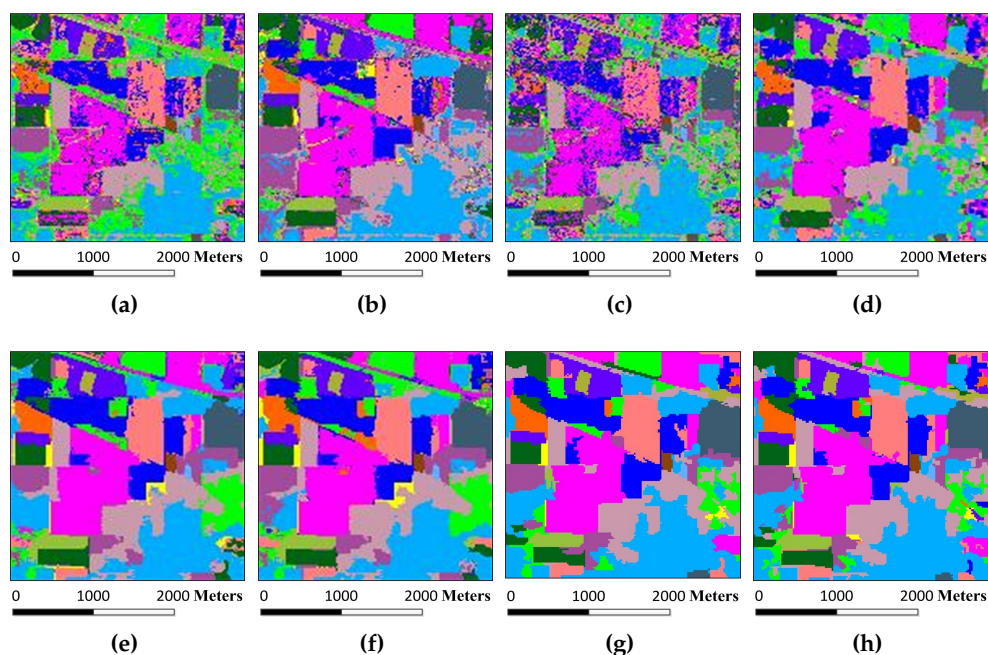


Figure 8. Classification maps for the Indian Pines image by different algorithms (OA values are reported in parentheses): (a) SVM (78.01%); (b) EMP (92.71%); (c) SRC (68.91%); (d) JSRC (94.42%); (e) MASR (98.27%); (f) SCMK (97.96%); (g) SBDSM (97.08%); and (h) MSSR (98.58%).

Table 3. Classification accuracy of the Indian Pines image by the classification algorithms used in this work for comparison. Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface. EMP, extended morphological profile; JSRC, joint sparse representation classification; MASR, multiscale adaptive sparse representation; SCMK, superpixel-based classification via multiple kernel; SBDSM, superpixel-based discriminative sparse model; MSSR, multiscale superpixel-based sparse representation.

Class	SVM	EMP	SRC	JSRC	MASR	SCMK	SBDSM	MSSR
1	80.24	98.22	58.64	92.68	94.68	99.99	97.56	97.45
2	72.99	85.84	52.35	94.68	97.25	97.24	96.34	97.83
3	66.76	90.86	53.65	93.44	97.34	97.20	89.69	99.60
4	84.10	93.75	36.78	91.93	97.41	96.89	99.06	98.59
5	90.68	93.40	82.35	94.05	97.20	96.31	93.55	98.16
6	94.09	98.47	89.98	95.58	99.62	99.59	99.85	99.82
7	83.96	92.75	88.56	83.20	96.41	99.68	96.43	96.73
8	96.15	99.90	90.23	99.86	99.89	100	100	100
9	92.00	100	71.35	36.67	76.42	100	100	92.23
10	77.11	87.45	68.32	91.21	97.99	92.35	94.74	98.28
11	69.84	91.91	75.32	95.98	98.64	98.61	96.92	99.28
12	73.72	87.85	42.56	88.89	97.72	96.78	97.75	97.19
13	98.28	97.97	91.21	83.04	99.01	99.13	99.46	99.46
14	90.67	99.22	88.52	99.56	99.99	99.64	99.82	100
15	70.67	97.85	36.25	93.26	98.62	97.21	93.08	92.04
16	96.12	98.27	88.58	90.12	95.87	97.02	96.39	96.39
OA (%)	78.37	92.49	68.34	94.56	98.26	98.01	96.92	98.56
AA (%)	83.58	94.61	65.20	89.01	96.66	97.95	95.19	97.98
Kappa	0.75	0.91	0.64	0.94	0.98	0.98	0.95	0.98

In the experiment on the Pavia University image, we randomly select 1% labeled pixels for each class as training samples and the rest of labeled pixels as the testing samples (see Table 4). Superpixel

segmentation maps and classification maps under different scales are shown in Figures 9 and 10. The quantitative results are illustrated in Table 5. As can be seen from Figure 9, Figure 10 and Table 5, with the increase of superpixel number, more details are presented in the segmentation maps and the classification maps. In this case, for the regions containing an abundance of details, the classification performances are improved. For example, the region size of the forth and the ninth classes is small and their classification accuracies are improved with the increase of superpixel number. Meanwhile, the classification accuracy of the seventh class is always 100%. The reason is that the region of this class is relatively smooth, and the superpixel-based classification method can obtain a good classification result. In addition, the classification maps and quantitative results from different classifiers on the University of Pavia image are shown in Figure 11 and Table 6. As shown from Figure 11 and Table 6, the proposed MSSR algorithm achieves competitive results, in terms of visual quality and quantitative metrics. Moreover, the MSSR algorithm obtains higher classification accuracies than the SBDSM algorithm for almost all classes. For some classes, the increase is obvious. For instance, in Table 6, the classification accuracy of pixels representing gravel climbs from 78.19% to 98.78%. The results demonstrate the superiority of multiple scales compared with the single scale, which can more accurately classify pixels in the detailed or near-edge regions.

Table 4. Number of training and test samples of nine classes in the University of Pavia image.

Class	Name	Train	Test
1	Asphalt	67	6631
2	Meadows	187	18,649
3	Gravel	21	2099
4	Trees	31	3064
5	Metal sheets	14	1345
6	Bare soil	51	5029
7	Bitumen	14	1330
8	Bricks	37	3682
9	Shadows	10	947
Total		432	42,776

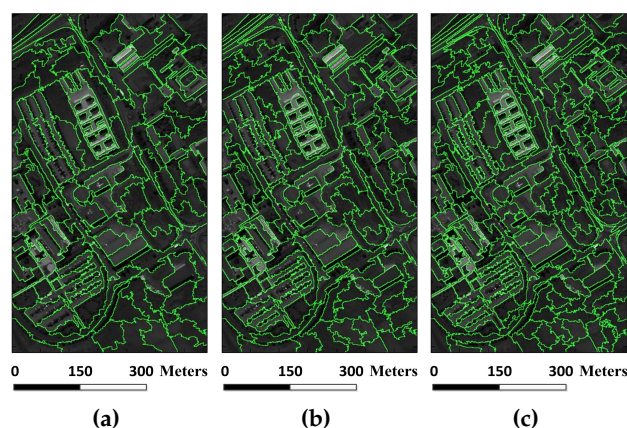


Figure 9. Cont.

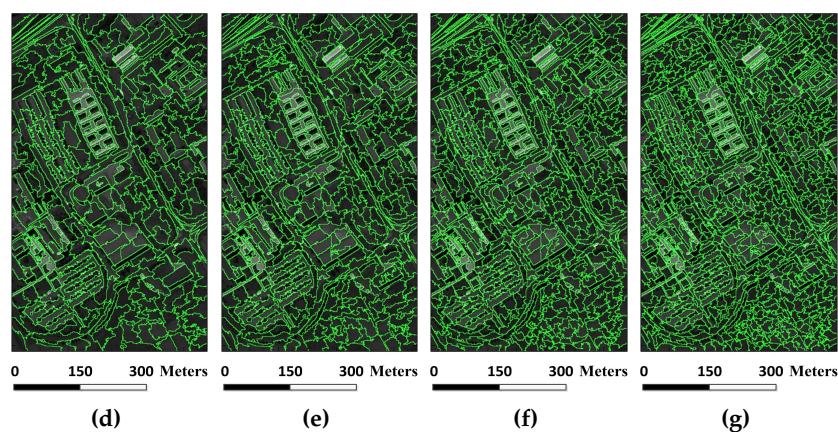


Figure 9. Superpixel segmentation results of the University of Pavia image under different scales. The number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 3200 and the power exponent n is an integer changing from -3 to three: (a) $n = -3$; (b) $n = -2$; (c) $n = -1$; (d) $n = 0$; (e) $n = 1$; (f) $n = 2$; and (g) $n = 3$.

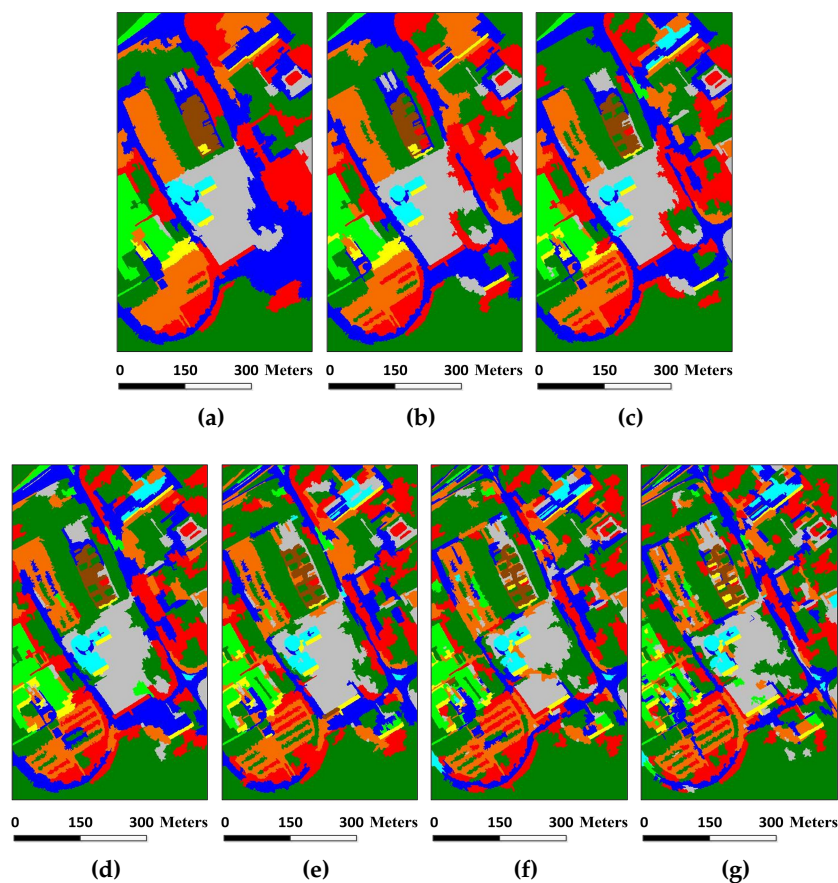


Figure 10. Classification results of the University of Pavia image under different scales. The number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 3200 and the power exponent n is an integer changing from -3 to three: (a) $n = -3$, OA = 91.74%; (b) $n = -2$, OA = 91.42%; (c) $n = -1$, OA = 92.39%; (d) $n = 0$, OA = 92.60%; (e) $n = 1$, OA = 92.12%; (f) $n = 2$, OA = 91.54%; and (g) $n = 3$, OA = 91.35%.

Table 5. Classification accuracy of the University of Pavia image under different scales. The number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 3200 and the power exponent n is an integer changing from -3 to three. Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface.

Class	$n = -3$	$n = -2$	$n = -1$	$n = 0$	$n = 1$	$n = 2$	$n = 3$
1	94.50	89.10	83.10	86.43	82.04	81.84	78.85
2	95.08	96.75	98.97	98.00	97.29	98.13	98.82
3	100	100	99.18	78.19	88.99	87.25	91.15
4	58.69	60.27	67.03	69.07	66.67	68.18	74.35
5	88.43	85.15	91.59	96.39	93.91	95.94	92.04
6	99.26	99.18	99.98	96.82	97.91	95.52	87.81
7	100	100	100	100	100	100	100
8	99.37	99.05	88.82	91.28	91.35	91.66	97.91
9	29.24	33.51	43.01	57.52	68.62	68.81	72.40
OA (%)	91.70	91.51	92.06	92.15	91.92	91.64	91.37
AA (%)	84.83	84.91	86.28	86.96	88.52	86.66	87.70
Kappa	0.89	0.89	0.90	0.90	0.89	0.89	0.88

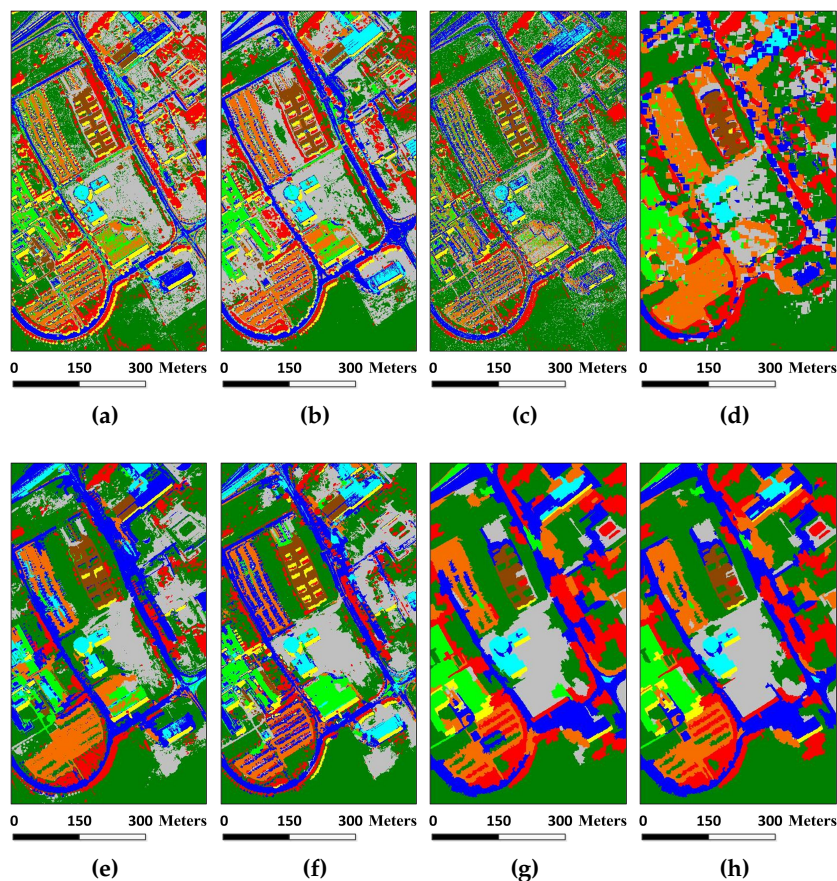


Figure 11. Classification maps for the Pavia University image by different algorithms (OA values are reported in parentheses): (a) SVM (86.52%); (b) EMP (91.80%); (c) SRC (77.90%); (d) JSRC (86.78%); (e) MASR (89.45%); (f) SCMK (94.96%); (g) SBDSM (92.60%); and (h) MSSR (95.47%).

Table 6. Classification accuracy of the University of Pavia image by the classification algorithms used in this work for comparison. Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface.

Class	SVM	EMP	SRC	JSRC	MASR	SCMK	SBDSM	MSSR
1	81.44	94.39	73.34	60.65	74.01	90.19	86.43	94.20
2	91.39	90.40	91.84	97.82	98.71	99.58	98.00	99.97
3	82.13	95.17	52.39	86.77	96.77	93.71	78.19	98.78
4	93.47	96.02	74.55	82.86	87.60	88.82	69.07	74.18
5	99.30	99.85	99.62	95.79	100	99.70	96.39	96.92
6	88.27	78.83	45.37	87.20	84.53	98.05	96.82	97.38
7	93.84	96.95	61.55	98.17	98.25	85.79	100	100
8	74.46	97.60	76.13	74.62	80.58	90.75	88.82	99.34
9	76.52	99.15	85.38	36.25	49.95	98.65	57.52	65.78
OA (%)	86.13	91.58	77.81	84.53	89.73	94.73	92.15	95.54
AA (%)	78.25	94.25	72.91	78.20	85.60	91.85	86.96	89.19
Kappa	0.82	0.89	0.70	0.81	0.86	0.93	0.90	0.94

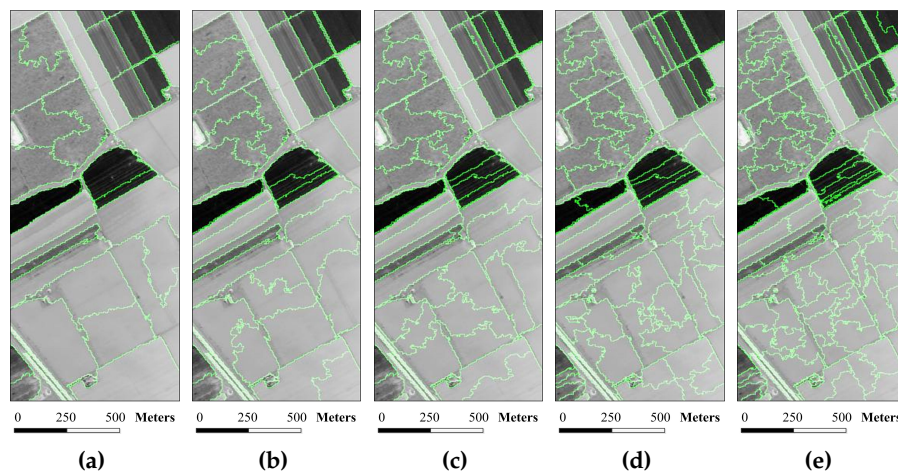
The third and fourth experiments are conducted on the Salinas and Washington DC images. For the Salinas image, 0.2% of the labeled data are randomly chosen for training, and the remaining 99.8% of data for testing (see Table 7). For the very small proportion of the training samples, the experiment is quite challenging. In the Washington DC image, 2% of the labeled pixels are selected as a training set and the remaining 98% as a testing set (see Table 8). For the Salinas image, Figures 12 and 13 respectively show the superpixel segmentation maps and classification maps under different scales. The corresponding quantitative values of classification results are tabulated in Table 9. For the Washington DC image, the superpixel segmentation maps and classification maps under different scales are illustrated in Figures 14 and 15. Table 10 shows the classification accuracies under different scales. As can be seen, for the Salinas image, although the proportion of training samples is very small, the high OA under all scales is acquired. Meanwhile, some classes, such as the first class, the third class and the ninth class, can obtain 100% classification accuracies. This is because the Salinas image has many large homogeneous regions, which make superpixel segmentation become easy. Obviously, error segmentation will appear when the number of superpixels is too small. In this case, the classification performance of some classes with small size regions will be deteriorated sharply, such as the eleventh class and the twelfth class. For the Washington DC image, it consists of many heterogeneous regions. Therefore, the optimal number of segmented superpixels is large, and the OA value is relatively steady near of the optimal scale. The qualitative and quantitative results from different algorithms on the Salinas and the Washington DC are shown in Figures 16 and 17 and Tables 11 and 12. We can observe that the proposed MSSR algorithm is usually superior to the other classifiers on the two datasets. Especially, for the Salinas image, compared with the other methods, the superpixel-based sparse representation algorithms greatly improve the classification accuracy under the condition of very limited training samples.

Table 7. Number of training and test samples of sixteen classes in the Salinas image.

Class	Name	Train	Test
1	Weeds_1	5	2004
2	Weeds_2	8	3718
3	Fallow	4	1973
4	Fallow plow	3	1391
5	Fallow smooth	6	2672
6	Stubble	8	3951
7	Celery	8	3571
8	Grapes	23	11,248
9	Soil	13	6190
10	Corn	7	3271
11	Lettuce 4 wk	3	1065
12	Lettuce 5 wk	4	1923
13	Lettuce 6 wk	2	914
14	Lettuce 7 wk	3	1068
15	Vineyard untrained	15	7253
16	Vineyard trellis	4	1803
Total		116	54,015

Table 8. Number of training and test samples of six classes in the Washington DC image.

Class	Name	Train	Test
1	Roof	63	3122
2	Road	36	1786
3	Trail	29	1399
4	Grass	26	1261
5	Shadow	24	1191
6	Tree	23	1117
Total		201	9698

**Figure 12.** Superpixel segmentation results of the Salinas image under different scales. The number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 1600 and the power exponent n is an integer changing from -2 to 2 : (a) $n = -2$; (b) $n = -1$; (c) $n = 0$; (d) $n = 1$; and (e) $n = 2$.

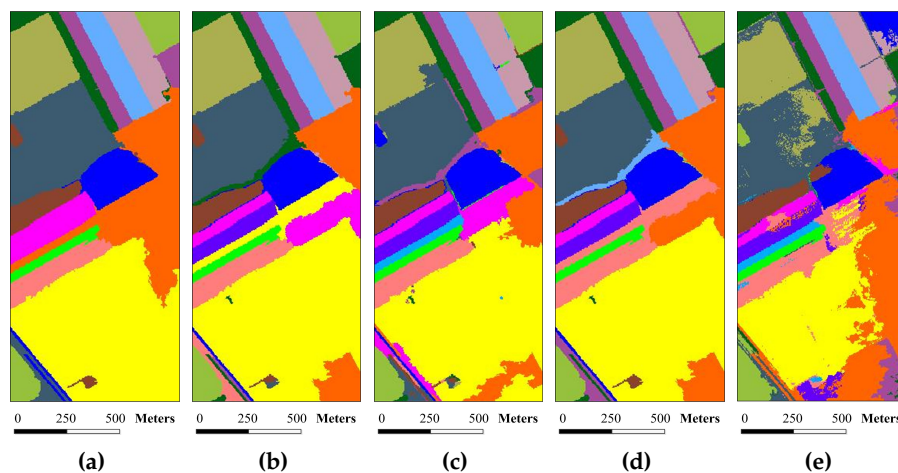


Figure 13. Classification results of the Salinas image under different scales. The number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 1600 and the power exponent n is an integer changing from -2 to two: (a) $n = -2$, OA = 95.21%; (b) $n = -1$, OA = 97.04%; (c) $n = 0$, OA = 98.38%; (d) $n = 1$, OA = 97.70%; and (e) $n = 2$, OA = 97.00%.

Table 9. Classification accuracy of the Salinas image under different scales. The number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 1600 and the power exponent n is an integer changing from -2 to two. Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface.

Class	$n = -2$	$n = -1$	$n = 0$	$n = 1$	$n = 2$
1	100	100	100	100	100
2	99.78	99.87	99.91	99.87	99.83
3	100	100	100	100	100
4	99.93	99.93	96.93	93.08	84.67
5	99.33	99.33	99.93	99.33	99.40
6	99.95	99.33	99.95	99.95	99.89
7	99.89	99.90	99.95	99.75	99.66
8	99.11	99.88	99.18	99.02	96.01
9	100	100	100	100	100
10	97.34	98.08	97.40	77.92	95.92
11	40.00	100	100	100	100
12	60.00	100	100	100	100
13	78.42	69.61	98.03	98.21	97.81
14	95.97	95.99	96.29	78.91	95.90
15	99.94	90.93	96.49	99.94	92.42
16	97.23	97.24	97.23	97.24	97.24
OA (%)	95.35	96.84	98.68	97.67	97.23
AA (%)	90.68	93.73	98.85	96.46	97.42
Kappa	0.96	0.96	0.98	0.97	0.97

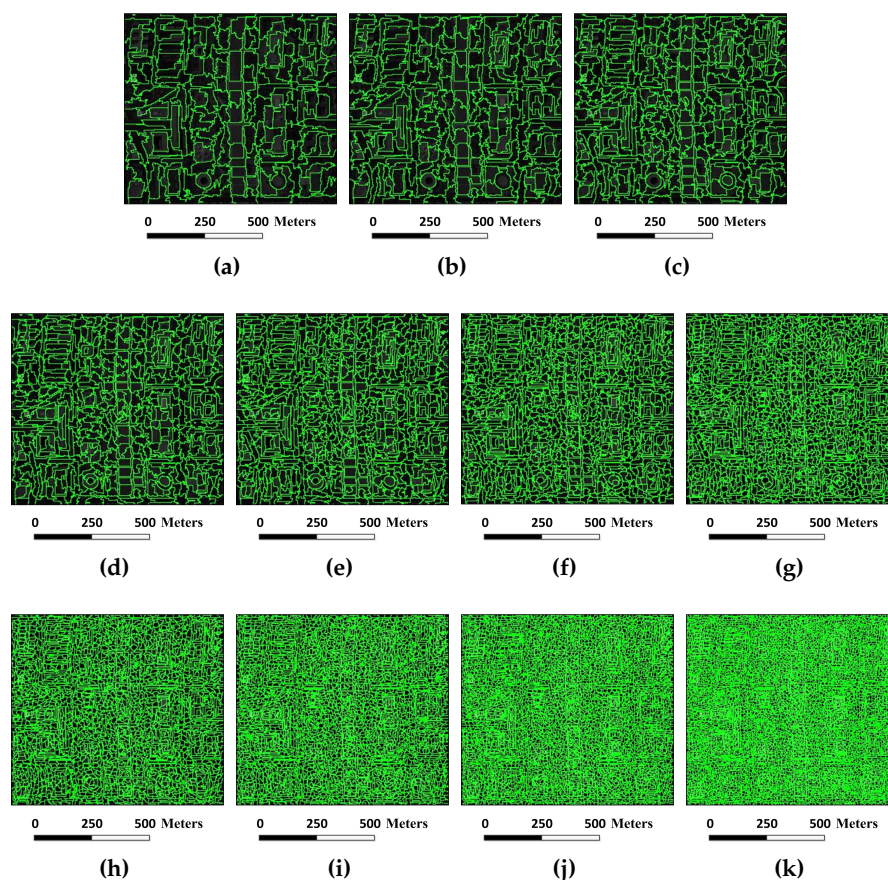


Figure 14. Superpixel segmentation results of the Washington DC image under different scales. The number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 12,800 and the power exponent n is an integer changing from -5 to five: (a) $n = -5$; (b) $n = -4$; (c) $n = -3$; (d) $n = -2$; (e) $n = -1$; (f) $n = 0$; (g) $n = 1$; (h) $n = 2$; (i) $n = 3$; (j) $n = 4$; and (k) $n = 5$.

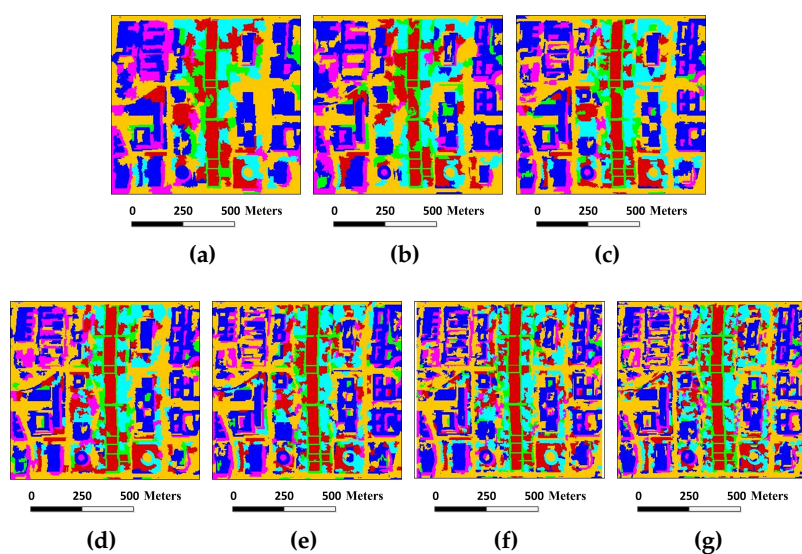


Figure 15. Cont.

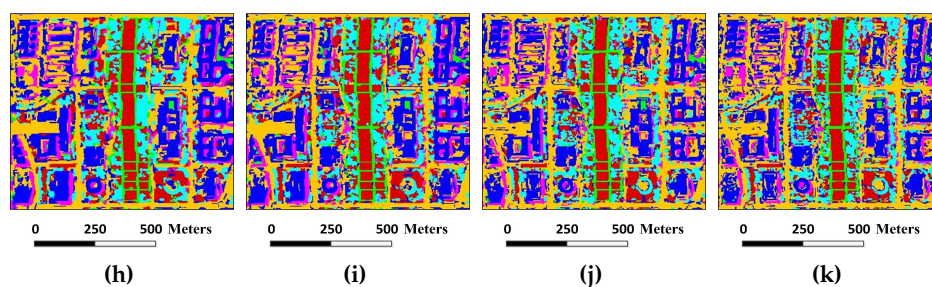


Figure 15. Classification results of the Washington DC image under different scales. The number of single-scale superpixel is gained by using Equation (5), in which the fundamental superpixel number is set to 12,800 and the power exponent n is an integer changing from -5 to five: (a) $n = -5$, OA = 86.48%; (b) $n = -4$, OA = 90.29%; (c) $n = -3$, OA = 92.56%; (d) $n = -2$, OA = 91.43%; (e) $n = -1$, OA = 91.84%; (f) $n = 0$, OA = 93.33%; (g) $n = 1$, OA = 92.25%; (h) $n = 2$, OA = 92.49%; (i) $n = 3$, OA = 93.12%; (j) $n = 4$, OA = 92.55%; and (k) $n = 5$, OA = 92.31%.

Table 10. Classification accuracy of the Washington DC image under different scales. The number of single-scale superpixels is gained by using Equation (5), in which the fundamental superpixel number is set to 12,800 and the power exponent n is an integer changing from -5 to five. Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface.

Class	$n = -5$	$n = -4$	$n = -3$	$n = -2$	$n = -1$	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
1	90.12	89.37	93.28	90.77	87.90	93.48	91.65	92.04	90.02	89.07	88.42
2	98.63	97.38	97.66	99.83	98.18	97.21	97.78	96.41	99.37	99.20	99.37
3	83.61	92.64	94.83	94.97	97.67	85.29	95.05	92.57	92.43	92.50	90.97
4	83.68	91.52	93.30	92.41	92.89	96.77	97.50	95.88	96.77	96.69	95.96
5	78.46	93.33	90.43	89.32	91.79	94.02	87.95	93.33	93.33	93.25	93.68
6	68.55	73.75	81.59	80.40	82.86	92.62	82.04	85.87	86.87	88.06	86.78
OA (%)	86.07	90.27	92.62	91.86	91.68	93.38	92.45	92.85	92.96	92.75	92.17
AA (%)	83.84	89.66	91.85	91.28	91.88	93.23	0.92	92.68	93.13	93.13	92.53
Kappa	0.83	0.88	0.91	0.90	0.90	0.92	0.91	0.91	0.91	0.91	0.90

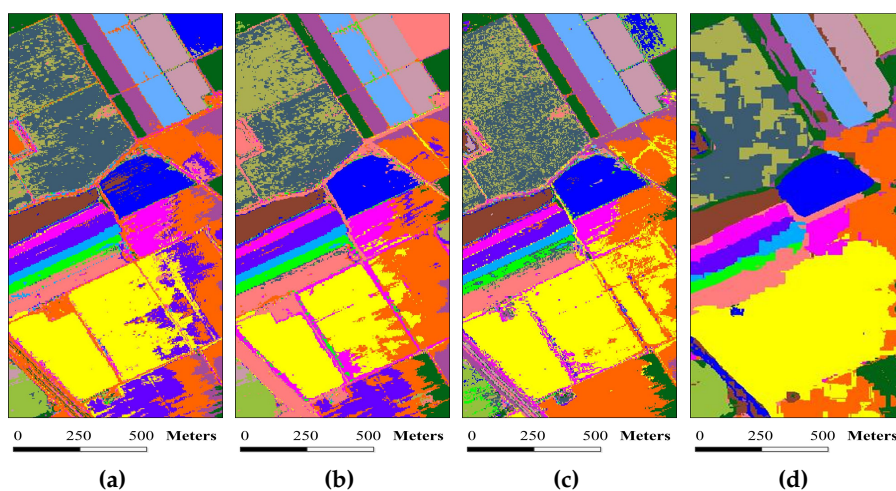


Figure 16. Cont.

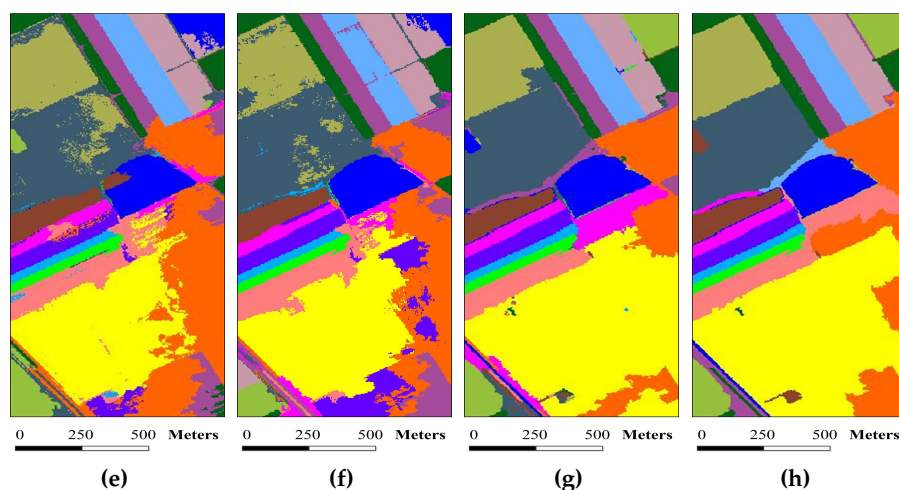


Figure 16. Classification maps for the Salinas image by different algorithms (OA values are reported in parentheses): (a) SVM (80.23%); (b) EMP (85.84%); (c) SRC (81.94%); (d) JSRC (84.79%); (e) MASR (92.21%); (f) SCMK (94.53%); (g) SBDSM (98.38%); and (h) MSSR (99.41%).

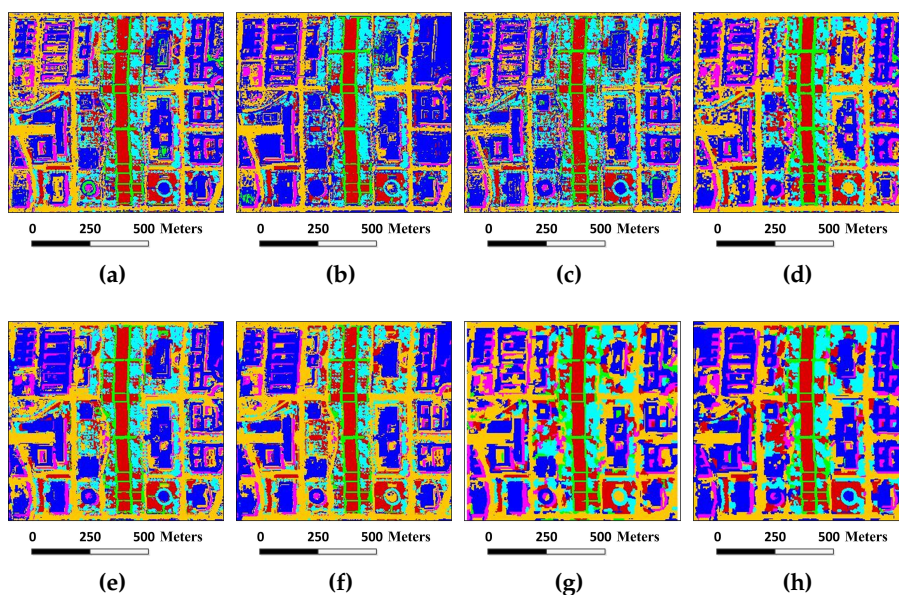


Figure 17. Classification maps for the Washington DC image by different algorithms (OA values are reported in parentheses): (a) SVM (90.98%); (b) EMP (90.28%); (c) SRC (91.95%); (d) JSRC (92.79%); (e) MASR (95.62%); (f) SCMK (94.55%); (g) SBDSM (93.33%); and (h) MSSR (96.60%).

The average running times over ten realizations of the proposed MSSR algorithm and the other algorithms are given in Table 13. We implemented these experiments using MATLAB on a computer with an Intel(R) Xeon(R) CPU E5-2603 v3 @1.60 GHz and 96 GB of RAM. As can be seen, in the case of a comparatively great proportion of training samples, the SBDSM algorithm consumes much less execution time compared with other algorithms. This demonstrates the high efficiency of the superpixel-based sparse classification strategy. However, on account of the multiscale information procedure, the MSSR algorithm consumes more computation time. Meanwhile, under the condition of less training samples, for the SBDSM algorithm, there is no advantage on the computation speed. The main reason is that the computational cost by the training processes for the SVM, EMP and SCMK algorithms is significantly reduced. In addition, for the multiscale-based sparse algorithms,

the time consumption of the MSSR algorithm decreased compared with the MASR method, which illustrates the effectiveness of the superpixel-based strategy. Moreover, the running time is expected to be further reduced by adopting a general-purpose graphics processing unit.

Table 11. Classification accuracy of the Salinas image by the classification algorithms used in this work for comparison. Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface.

Class	SVM	EMP	SRC	JSRC	MASR	SCMK	SBDSM	MSSR
1	98.35	99.23	92.56	99.99	99.90	99.94	100	100
2	98.76	98.90	95.39	99.67	99.35	97.21	99.91	99.78
3	92.54	93.53	74.67	68.25	97.88	87.54	100	100
4	96.28	98.92	98.68	62.47	88.95	99.93	96.93	99.93
5	95.75	94.25	89.88	82.55	92.86	99.35	99.93	99.93
6	95.90	95.61	98.76	99.27	99.97	99.95	99.95	99.90
7	98.47	99.24	99.02	95.35	100	99.53	99.95	99.97
8	63.26	53.35	75.30	86.63	84.98	97.85	99.18	99.28
9	97.55	98.75	93.79	100	99.76	99.64	100	100
10	82.38	93.41	76.02	85.27	88.10	87.94	97.40	97.32
11	88.97	96.51	88.20	89.17	98.97	97.18	100	100
12	85.57	95.36	95.48	67.54	98.85	100	100	98.03
13	98.68	98.57	96.81	51.82	99.69	92.58	98.03	95.97
14	86.82	93.84	89.75	95.25	94.16	95.97	96.29	95.97
15	59.66	79.00	88.20	67.40	80.84	88.48	96.49	99.94
16	35.88	38.52	71.44	83.42	92.68	94.93	97.23	97.25
OA (%)	80.65	85.75	81.72	85.33	92.33	94.14	98.68	99.41
AA (%)	87.01	93.40	85.21	82.63	94.75	94.92	98.85	99.17
Kappa	0.81	0.84	0.79	0.84	0.91	0.93	0.98	0.99

Table 12. Classification accuracy of the Washington DC image by the classification algorithms used in this work for comparison. Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface.

Class	SVM	EMP	SRC	JSRC	MASR	SCMK	SBDSM	MSSR
1	83.40	86.09	89.95	91.13	98.75	93.56	93.48	93.98
2	98.35	96.80	94.98	97.95	98.14	98.92	97.21	99.45
3	93.23	89.31	81.35	89.80	98.26	98.18	85.29	95.92
4	96.17	93.46	95.40	93.38	90.94	92.57	96.77	97.58
5	91.26	97.58	95.98	94.36	99.16	95.04	94.02	97.69
6	92.83	79.64	94.53	92.98	82.74	91.25	92.62	96.81
OA (%)	91.11	90.04	91.59	93.06	95.97	94.96	93.38	96.74
AA (%)	92.54	90.48	92.03	93.27	94.75	94.91	93.23	96.58
Kappa	0.89	0.88	0.90	0.91	0.95	0.94	0.92	0.95

Table 13. Average running time (seconds) over ten realizations for the classification of the Indian Pines, University of Pavia, Salinas and Washington DC images by the algorithms used in this work.

Images	SVM	EMP	SRC	JSRC	MASR	SCMK	SBDSM	MSSR
Indian Pines	242.3	63.5	17.4	118.6	1580.9	254.6	9.2	67.4
U.Pavia	30.4	24.6	50.4	425.3	3010.4	48.7	26.5	197.5
Salinas	13.4	10.5	76.2	800.9	4129.1	20.6	18.4	92.1
Washington DC	14.6	23.3	16.7	34.2	210.6	16.5	21.9	187.3

4.3. Effect of Superpixel Scale Selection

We first analyze the effect of the number of single-scale superpixels. In this analysis, the training and testing sets for the Indian Pines, University of Pavia, Salinas and Washington DC images are the same sets in the aforementioned comparison experiments. The average results over ten runs for this analysis are obtained. Figure 18 shows the OA values under different fundamental superpixel numbers. The number of single-scale superpixels is gained by applying Equation (5), in which the power exponent n is set to zero. In these experiments, the fundamental superpixel number S_f is varied from 400 to 51,200. Because the range of S_f is large, we adopt the log scale to represent. The log value of 400, 800, 1600, 3200, 6400, 12,800, 25,600 and 51,200, respectively, is 2.6, 2.9, 3.2, 3.5, 3.8, 4.1, 4.4 and 4.7. As shown in Figure 18, with the increase of the fundamental superpixel number, the OA values of the four images firstly increase and then decrease. This demonstrates that the classification accuracy will be deteriorated at a too small or a too large superpixel scale. Moreover, in Figure 18, the highest OA values of the four images are obtained when S_f reaches 3200, 3200, 1600 and 12,800 for the Indian Pines, University of Pavia, Salinas and Washington DC images, respectively. This result illustrates that the classification accuracy is closely relative with the complexity of the image. Compared with the other three images, the Washington DC image needs the most superpixels to realize the optimal classification performance, although its size is relatively small. The main reason is that the spatial structure of this image is comparatively complicated. In addition, the OA values keep high in a large dynamic range. For example, when the fundamental superpixel number is between 6400 and 51,200, the OA values of the Washington DC image are always over 90%. Therefore, multiscale superpixel information can be used for HSI classification.

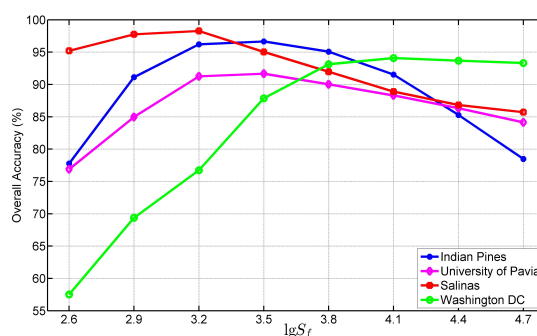


Figure 18. Classification accuracy OA versus different fundamental superpixel numbers S_f on the four test images.

Figure 19 illustrates the relationship among the OA value, the number of multiscale superpixels and the fundamental superpixel number S_f . In the same way, for S_f in Figure 19, we adopt the log value. The log value of 400, 800, 1600, 3200, 6400, 12,800, 25,600 and 51,200, respectively, is 2.6, 2.9, 3.2, 3.5, 3.8, 4.1, 4.4 and 4.7. The training sets for the four images are set to the same as before. The number of superpixel multiscales is an odd number rising from three to 15. In these experiments, five contiguous fundamental superpixel numbers in the previous experiment are selected, in which the third number corresponds to the optimal OA in the previous experiment. As can be seen, for the Indian Pines and University of Pavia images, the optimal segmentation accuracies are obtained when the fundamental superpixel number is set to 3200 and the scale number of multiscale superpixels is set to seven. For the Salinas image, when the fundamental superpixel number reaches to 1600 and the scale number of multiscale superpixels reaches to five, the optimal OA can be acquired. For the Washington DC image, when the fundamental superpixel number is 12,800 and the scale number of multiscale superpixels is 11, the best classification performance is obtained. Among these four hyperspectral images, the Salinas image contains many homogeneous regions, and the Washington DC has a large number of heterogeneous regions. The experimental results show that, to obtain

perfect classification performance, the relatively complex image requires more scale number and more superpixel number at each scale.

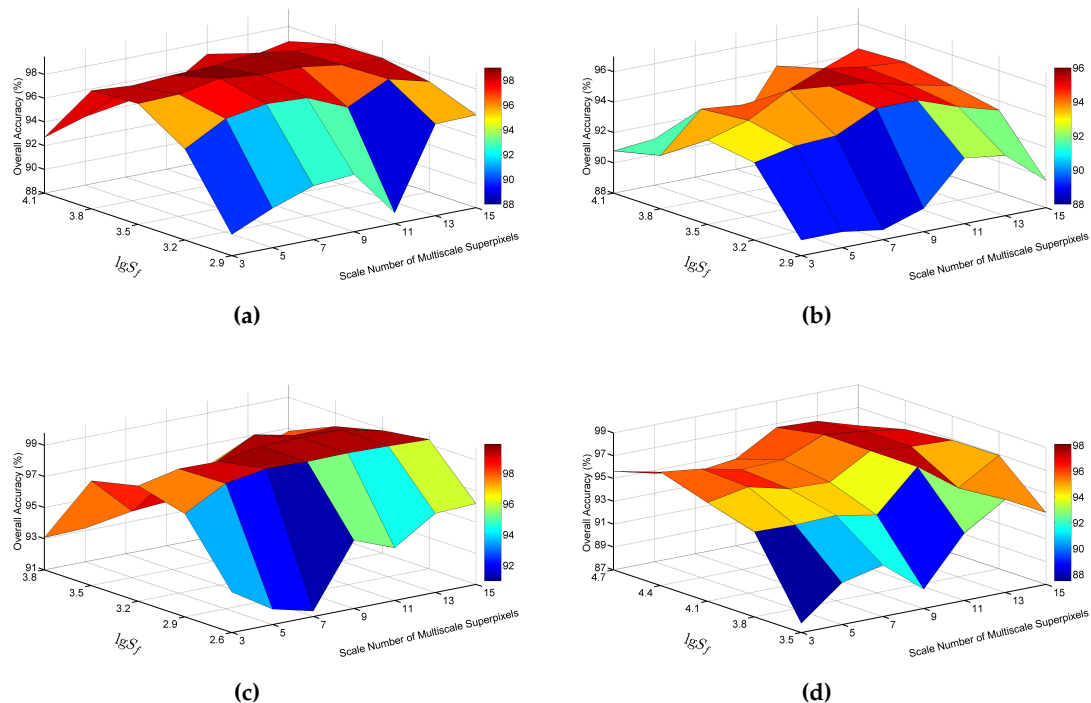


Figure 19. Relationship among the OA value, the number of multiscale superpixels and the fundamental superpixel number S_f : (a) Indian Pine image; (b) University of Pavia image; (c) Salinas image; and (d) Washington DC image.

4.4. Comparison of Different Superpixel Segmentation Methods

In this section, we compare the performance of the adopted ERS algorithm with the performances of two competing superpixel segmentation algorithms, i.e., the Felzenszwalb-Huttenlocher (FH) algorithm [59] and the simple linear iterative clustering (SLIC) [42] algorithm. In the comparison, the Indian Pine image is utilized. The training and testing sets are the same sets as before. In the ERS algorithm, the fundamental superpixel number is 3200 and the power exponent n in Equation (5) is an integer changing from -3 to three. In the FH algorithm, multiscale superpixels are generated with various scales and smoothing parameters, σ and k_S , where σ is the Gaussian smoothing parameter and k_S controls the region size. In the SLIC algorithm, the number of multiscale superpixels is obtained by presetting the number of superpixel segmentation. In the two comparison experiments, for each scale, the superpixel number is approximately the superpixel number generated by applying the ERS algorithm. Figures 6, 20 and 21 illustrate the superpixel segmentation results under different scales by adopting ERS, FH and SLIC algorithms, respectively. The qualitative and quantitative results under different scales are shown in Figures 7, 22 and 23, and Tables 2, 14 and 15. In addition, the three over-segmentation methods are utilized in the proposed MSSR algorithm, which are called the MSSR_ERS, MSSR_FH and MSSR_SLIC algorithms, respectively. The classification accuracies and maps by these algorithms are also shown in Table 16 and Figure 24. As can be seen from Figures 6, 20 and 21, for the FH algorithm, since there is no explicit constraint on length, the segmentation shapes are the most irregular. The SLIC algorithm yields similar size segmentation regions by setting uniform grid spacing. For the ERS algorithm, a balancing term is utilized to enable superpixels with similar sizes. From the point of classification performances on the single scale, with the increase of superpixel number, the OA values present an

increase firstly, then a descending tendency. When superpixel number is too small, some classes with few pixels are completely misclassified by applying the FH and SLIC algorithms, such as the seventh class. The error classification is induced by error segmentation. On the contrary, when the superpixel number is too large, small segmented regions will deteriorate classification performance, since it lacks sufficient spatial information for classification. Three superpixel segmentation methods are applied in the proposed MSSR algorithm. These methods almost consume the same computational time. The classification results show that ERS-based classification algorithm outperforms the other two over-segmentations-based classification algorithms.

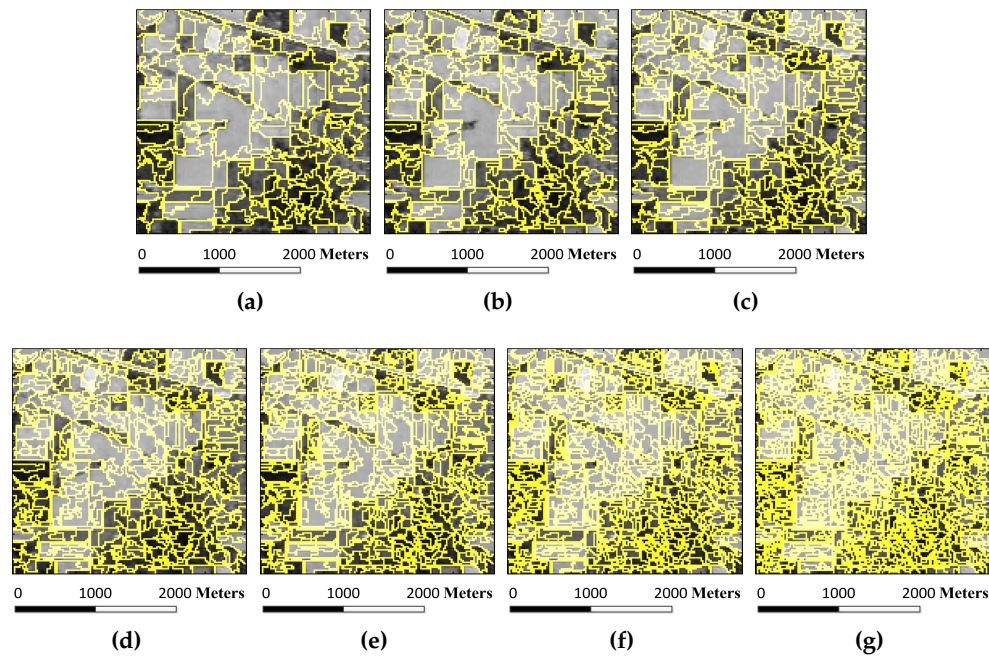


Figure 20. Felzenszwalb-Huttenlocher (FH) segmentation results of the Indian Pines image under different scales. Multiscale superpixels are generated with various scales and smoothing parameters, σ and k_S : (a) $\sigma = 0.2, k_S = 43$; (b) $\sigma = 0.2, k_S = 31$; (c) $\sigma = 0.2, k_S = 23$; (d) $\sigma = 0.3, k_S = 17$; (e) $\sigma = 0.4, k_S = 12$; (f) $\sigma = 0.4, k_S = 9$; and (g) $\sigma = 0.3, k_S = 7$.

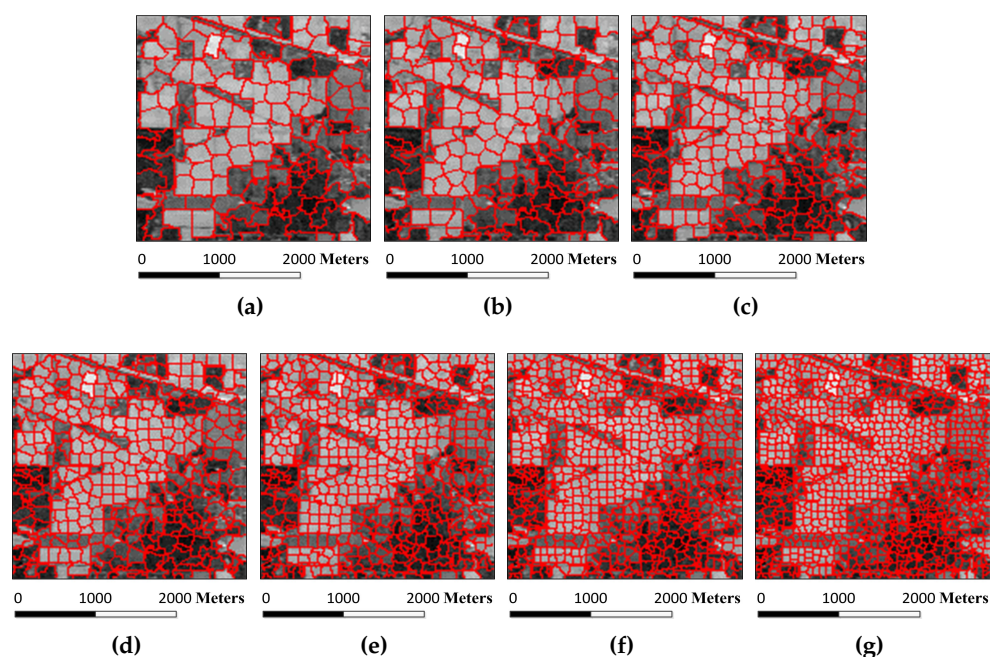


Figure 21. Simple linear iterative clustering (SLIC) segmentation results of the Indian Pines image under different scales. The number of multiscale superpixels is obtained by presetting the number of superpixel segmentations n_S : (a) $n_S = 147$; (b) $n_S = 206$; (c) $n_S = 288$; (d) $n_S = 415$; (e) $n_S = 562$; (f) $n_S = 780$; and (g) $n_S = 1055$.

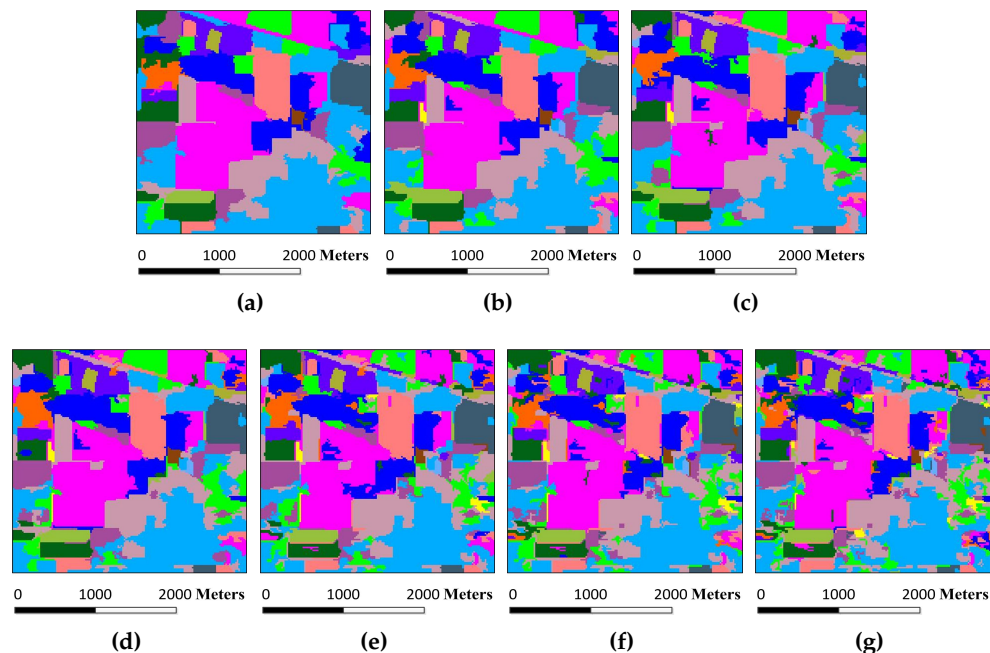


Figure 22. Classification results of the Indian Pines image under different scales. The FH segmentation method is applied in the SBDSM algorithm. Multiscale superpixels are generated with various scales and smoothing parameters, σ and k_S : (a) $\sigma = 0.2, k_S = 43$, OA = 83.56%; (b) $\sigma = 0.2, k_S = 31$, OA = 93.21%; (c) $\sigma = 0.2, k_S = 23$, OA = 93.52%; (d) $\sigma = 0.3, k_S = 17$, OA = 96.25%; (e) $\sigma = 0.4, k_S = 12$, OA = 94.52%; (f) $\sigma = 0.4, k_S = 9$, OA = 94.32%; and (g) $\sigma = 0.3, k_S = 7$, OA = 94.28%.

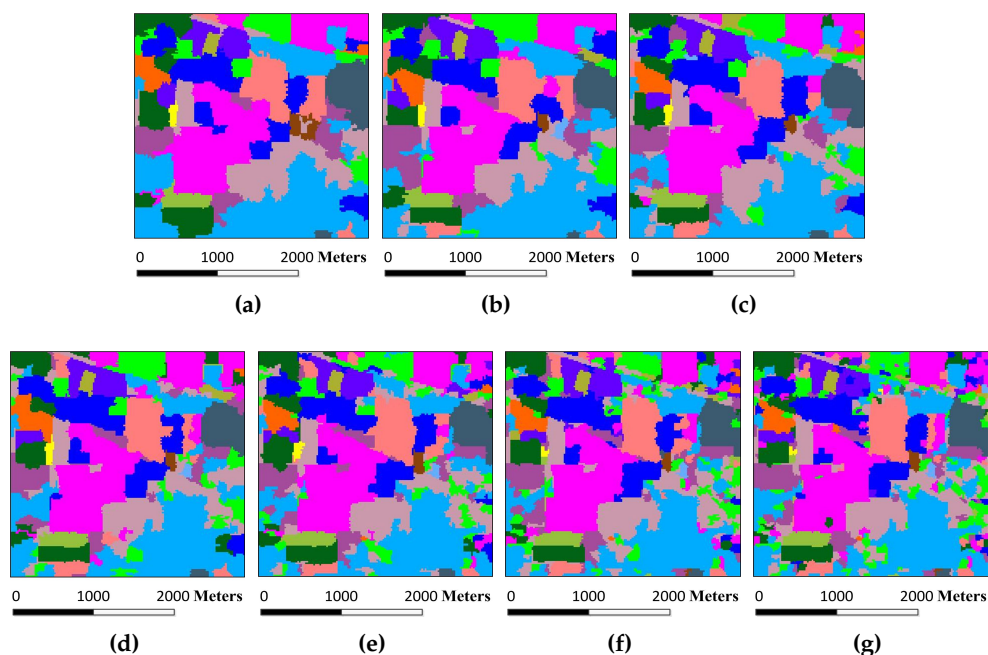


Figure 23. Classification results of the Indian Pines image under different scales. The SLIC segmentation method is applied in the SBDSM algorithm. The number of multiscale superpixels is obtained by presetting the number of superpixel segmentations n_S : (a) $n_S = 147$, OA = 89.10%; (b) $n_S = 206$, OA = 91.45%; (c) $n_S = 288$, OA = 93.16%; (d) $n_S = 415$, OA = 96.81%; (e) $n_S = 562$, OA = 96.52%; (f) $n_S = 780$, OA = 95.32%; and (g) $n_S = 1055$, OA = 95.24%.

Table 14. Classification accuracy of the Indian Pines image under different scales. The FH segmentation method is applied in the SBDSM algorithm. Multiscale superpixels are generated with various scales and smoothing parameters, σ and k_S . Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface.

Class	$\sigma = 0.2$ $k_S = 43$	$\sigma = 0.2$ $k_S = 31$	$\sigma = 0.2$ $k_S = 23$	$\sigma = 0.3$ $k_S = 17$	$\sigma = 0.4$ $k_S = 12$	$\sigma = 0.4$ $k_S = 9$	$\sigma = 0.3$ $k_S = 7$
1	92.68	92.68	92.68	92.68	92.68	92.68	92.68
2	86.93	81.48	79.92	96.74	93.62	90.50	93.93
3	99.46	86.85	98.66	92.90	94.51	93.31	91.16
4	42.72	94.01	87.79	90.74	75.59	90.61	82.63
5	90.78	99.54	99.08	99.08	95.62	91.94	94.93
6	99.71	96.00	99.85	99.54	99.70	99.24	96.35
7	0	100	96.00	96.00	96.00	96.00	96.00
8	100	100	100	100	98.60	100	100
9	0	100	100	100	100	100	83.33
10	91.19	90.93	94.51	96.91	97.94	96.57	90.96
11	62.61	93.75	96.20	92.03	91.58	95.52	94.34
12	95.87	91.93	79.36	81.43	91.56	87.24	86.12
13	99.46	99.46	99.46	99.46	99.46	99.46	98.37
14	83.04	99.82	99.82	99.82	99.74	99.74	99.74
15	97.98	97.98	98.27	98.85	98.85	98.27	97.41
16	97.59	97.59	97.59	97.59	97.59	97.59	97.59
OA (%)	83.62	93.62	93.97	96.04	94.96	94.56	94.26
AA (%)	77.50	94.99	94.95	95.86	95.19	95.92	93.47
Kappa	0.82	0.93	0.93	0.94	0.94	0.95	0.93

Table 15. Classification accuracy of the Indian Pines image under different scales. The SLIC segmentation method is applied in the SBDSM algorithm. The number of multiscale superpixels is obtained by presetting the number of superpixel segmentations n_S . Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface.

Class	$n_S = 147$	$n_S = 206$	$n_S = 288$	$n_S = 415$	$n_S = 562$	$n_S = 780$	$n_S = 1055$
1	97.56	97.56	75.61	97.56	97.56	73.17	75.61
2	83.66	82.33	91.91	94.16	91.36	92.22	93.70
3	86.75	84.75	85.68	93.84	98.66	97.99	95.45
4	81.67	66.20	96.71	92.96	92.96	97.65	89.67
5	87.33	94.01	95.16	88.71	87.79	94.24	91.94
6	98.17	98.93	96.19	98.33	98.33	99.09	98.63
7	0	100	100	96.00	96.00	92.00	96.00
8	99.07	99.07	100	100	99.30	99.77	99.97
9	98.45	99.58	100	100	88.89	66.77	55.56
10	71.40	89.82	84.67	97.71	94.05	95.31	93.48
11	92.39	93.57	93.62	96.29	95.02	97.01	97.15
12	80.86	88.74	90.81	90.99	87.62	86.49	91.37
13	99.43	86.41	99.46	99.46	99.46	98.91	98.91
14	98.42	99.89	99.56	99.65	98.51	99.30	99.82
15	99.14	98.85	97.41	97.98	91.64	89.34	87.32
16	98.80	97.59	97.59	98.80	97.59	97.59	96.39
OA (%)	89.28	91.78	93.06	97.08	96.69	95.62	95.57
AA (%)	85.91	92.49	87.77	96.40	94.42	92.55	92.68
Kappa	0.88	0.91	0.92	0.96	0.94	0.95	0.95

Table 16. Classification accuracy of the Indian Pines image by applying the MSSR_FH, MSSR_SLIC and MSSR_ERS algorithms. Class-specific accuracy values are in percentage. The best results are highlighted in bold typeface.

Class	MSSR_FH	MSSR_SLIC	MSSR_ERS
1	92.68	97.56	97.45
2	92.97	97.35	97.83
3	97.86	97.19	99.60
4	96.71	97.18	98.59
5	98.39	93.09	98.16
6	99.85	99.24	99.82
7	96.00	96.00	96.73
8	100	100	100
9	100	100	92.23
10	97.37	97.94	98.28
11	95.70	96.71	99.28
12	91.37	97.94	97.19
13	99.46	99.46	99.46
14	99.82	100	100
15	99.14	98.85	92.04
16	97.56	97.59	96.39
OA (%)	96.77	97.75	98.56
AA (%)	97.18	97.88	97.98
Kappa	0.96	0.97	0.98
time (s)	67.1	66.3	67.4

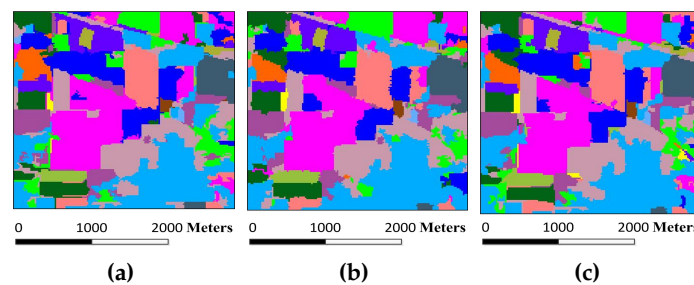


Figure 24. Classification maps by using MSSR_FH, MSSR_SLIC and MSSR_ERS algorithms: (a) MSSR_FH, OA = 96.54%; (b) MSSR_SLIC, OA = 97.38%; (c) MSSR_ERS, OA = 98.58%.

4.5. Effects of Training Sample Number

In this section, we analyze the effects of different training samples on the aforementioned classifiers on four HSI datasets. Except for the number of the training samples; the other parameters for all classifiers are the same as before. Different percentages of training samples are randomly selected for the Indian Pine image (from 1% to 20%), University of Pavia image (from 0.2% to 2%), Salinas (from 0.1% to 1%) and Washington DC (from 1% to 10%) images, and the rest of the samples are for testing. The OA values for different classifiers under the different number of training samples are indicated in Figure 25, which are averaged over ten runs. As shown in Figure 25, with the growth of training samples, the performance of classification generally improves. Moreover, the proposed MSSR algorithm generally outperforms all other algorithms. The reason is that, comparing with other algorithms, the superpixel-based segmentation method and the multiscale strategy can more effectively explore the spatial information for HSI classification. Particularly, for the Salinas image, under a limited number of training samples, the MSSR algorithm can obtain high classification accuracy. This is because this image has large homogenous regions.

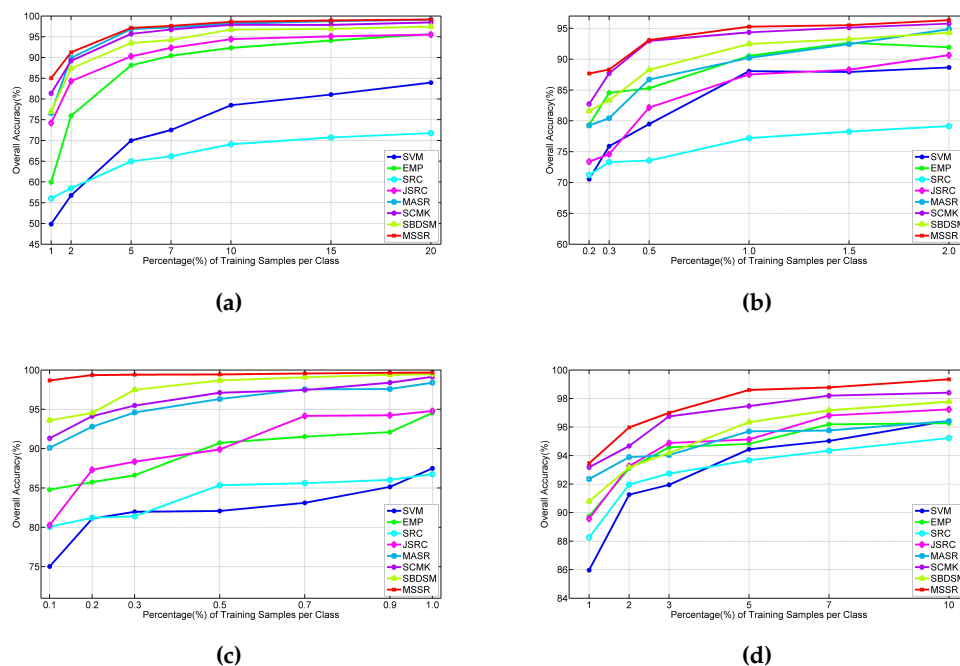


Figure 25. Effect of the number of training samples on SVM, EMP, SRC, JSRC, MASR, SCMK, SBDSM and MSSR for the: (a) Indian Pines image; (b) University of Pavia images; (c) Salinas image; and (d) Washington DC image.

5. Conclusions

In this paper, a novel MSSR algorithm is presented for spectral-spatial HSI classification. Instead of using a single-scale superpixel, the MSSR adopts multiscale superpixels to effectively explore the spatial information of the HSI. Then, the JSRC is used to classify the multiscale superpixels, and an effective decision fusion is applied to obtain the final classification result. Unlike the common multiscale superpixel segmentation, in the proposed MSSR, the step length variation among multiscale superpixels can be adaptively changed with the complexity of the fundamental image. Experiments on four well-known HSI datasets demonstrate that the proposed MSSR classifier outperforms other state-of-the-art classifiers, in terms of quantitative metrics and visual quality on the classification maps. Moreover, under limited training samples, for the HSI with many large homogenous regions, the proposed algorithm can obtain high classification accuracy.

In these experiments, for the multiscale strategy, the initial scale (i.e. the superpixel number in the first scale) is empirically selected. In future work, a more systematic way of adaptively selecting this parameter for different datasets will be studied. Moreover, multiple features fusion will be integrated into the MSSR method to further improve the classification performance.

Acknowledgments: This work was supported in part by the National Natural Science Fund of China for Distinguished Young Scholars under Grant 61325007, by the National Natural Science Fund of China for International Cooperation and Exchanges under Grant 61520126001 and by the 2017 Scientific Research Project of Education Department of Hunan Province under Grant 1800. The authors would like to thank David A. Landgrebe from Purdue University for providing the AVIRIS image of Indian Pines and Paolo Gamba from University of Pavia for providing the ROSIS data set. The authors would like to thank the National Aeronautics and Space Administration Jet Propulsion Laboratory for providing the AVIRIS image of Salinas and the Spectral Information Technology Application Center of Virginia for providing the HYDICE image of Washington DC. The authors would also like to thank the handling editor and anonymous reviewers for their valuable comments and suggestions, which significantly improved the quality of this paper.

Author Contributions: Shuzhen Zhang and Wei Fu designed the proposed model and implemented the experiments. Shuzhen Zhang drafted the manuscript. Leyuan Fang contributed to the improvement of the proposed model and edited the manuscript. Shutao Li provided overall guidance to the project, reviewed and edited the manuscript and obtained funding to support this research.

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

References

1. Yuan, Y.; Wang, Q.; Zhu, G. Fast hyperspectral anomaly detection via high-order 2-D crossing filter. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 620–630.
2. Heldens, W.; Heiden, U.; Esch, T.; Stein, E.; Muller, A. Can the future EnMAP mission contribute to urban applications? a literature survey. *Remote Sens.* **2011**, *3*, 1817–1846.
3. Lee, M.A.; Huang, Y.; Yao, H.; Thomson, S.J. Determining the effects of storage on cotton and soybean leaf samples for hyperspectral analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2562–2570.
4. Kanning, M.; Siegmann, B.; Jarmer, T. Regionalization of uncovered agricultural soils based on organic carbon and soil texture estimations. *Remote Sens.* **2016**, *8*, 927.
5. Clark, M.L.; Roberts, D.A. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sens.* **2012**, *4*, 1820–1855.
6. Ryan, J.P.; Davis, C.O.; Tufillaro, N.B.; Kudela, R.M.; Gao, B. Application of the hyperspectral imager for the coastal ocean to phytoplankton ecology studies in monterey bay, CA, USA. *Remote Sens.* **2014**, *6*, 1007–1025.
7. Brook, A.; Dor, E.B. Quantitative detection of settled dust over green canopy using sparse unmixing of airborne hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 884–897.
8. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790.
9. Ratle, F.; Camps-Valls, G.; Weston, J. Semisupervised neural networks for efficient hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2271–2282.
10. Zhong, Y.; Zhang, L. An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 894–909.

11. Jiao, H.; Zhong, Y.; Zhang, L. Artificial DNA computing-based spectral encoding and matching algorithm for hyperspectral remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4085–4104.
12. Jia, S.; Shen, L.; Li, Q. Gabor feature-based collaborative representation for hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1118–1129.
13. Li, W.; Chen, C.; Su, H.; Du, Q. Local binary patterns and extreme learning machine for hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3681–3693.
14. 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.
15. Mirzapour, F.; Ghassemian, H. Multiscale gaussian derivative functions for hyperspectral image feature extraction. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 525–529.
16. Benediktsson, J.A.; Palmason, J.A.; Sveinsson, J.R. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 480–491.
17. Quesada-Barriuso, P.; Arguello, F.; Heras, D.B. Spectral-spatial classification of hyperspectral images using wavelets and extended morphological profiles. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1177–1185.
18. Li, J.; Marpu, P.R.; Plaza, A.; Bioucas-Dias, J.M.; Benediktsson, J.A. Generalized composite kernel framework for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4816–4829.
19. Camps-Valls, G.; Shervashidze, N.; Borgwardt, K.M. Spatio-spectral remote sensing image classification with graph kernels. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 741–745.
20. Wang, J.; Jiao, L.; Liu, H.; Yang, S.; Liu, F. Hyperspectral image classification by spatial-spectral derivative-aided kernel joint sparse representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2485–2500.
21. Liu, J.; Wu, Z.; Li, J.; Plaza, A.; Yuan, Y. Probabilistic-kernel collaborative representation for spatial-spectral hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2371–2384.
22. Zhang, L.; Zhang, L.; Tao, D.; Huang, X. Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 242–256.
23. Guo, X.; Huang, X.; Zhang, L.; Zhang, L.; Plaza, A.; Benediktsson, J.A. Support tensor machines for classification of hyperspectral remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3248–3264.
24. He, Z.; Li, J.; Liu, L.; Zhang, L. Tensor block-sparsity based representation for spectral-spatial hyperspectral image classification. *Remote Sens.* **2016**, *8*, 636.
25. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep learning-based classification of hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107.
26. Chen, Y.; Zhao, X.; Jia, X. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2381–2391.
27. Liang, H.; Li, Q. Hyperspectral imagery classification using sparse representations of convolutional neural network features. *Remote Sens.* **2016**, *8*, 99.
28. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227.
29. Ma, J.; Zhou, H.; Zhao, J.; Gao, Y.; Jiang, J.; Tian, J. Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6469–6481.
30. Ma, J.; Zhao, J.; Yuille, A.L. Non-rigid point set registration by preserving global and local structures. *IEEE Trans. Image Process.* **2016**, *25*, 53–64.
31. Li, S.; Yin, H.; Fang, L. Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4779–2677.
32. Srinivas, U.; Chen, Y.; Monga, V.; Nasrabadi, N.M.; Tran, T.D. Exploiting sparsity in hyperspectral image classification via graphical models. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 505–509.
33. Qian, Y.; Ye, M.; Zhou, J. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture feature. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2276–2291.
34. Fu, W.; Li, S.; Fang, L.; Kang, X.; Benediktsson, J.A. Hyperspectral image classification via shape-adaptive joint sparse representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 556–567.
35. Li, C.; Ma, Y.; Mei, X.; Liu, C.; Ma, J. Hyperspectral image classification with robust sparse representation. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 641–645.

36. 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.
37. 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.* **2013**, *7*, 2056–2065.
38. Zhang, E.; Jiao, L.; Zhang, X.; Liu, H.; Wang, S. Class-level joint sparse representation for multifeature-based hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4160–4177.
39. Peng, X.; Zhang, L.; Yi, Z.; Tan, K.K. Learning locality-constrained collaborative representation for robust face recognition. *Pattern Recognit.* **2014**, *47*, 2794–2806.
40. Peng, X.; Yu, Z.; Yi, Z.; Tang, H. Constructing the L2-graph for robust subspace learning and subspace clustering. *IEEE Trans. Cybern.* **2016**, *PP*, 1–14.
41. Yuan, Y.; Lin, J.; Wang, Q. Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization. *IEEE Trans. Cybern.* **2016**, *46*, 2966–2977.
42. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282.
43. Tian, Z.; Liu, L.; Zhang, Z.; Fei, B. A superpixel-based segmentation for 3D prostate MR images. *IEEE Trans. Med. Imag.* **2016**, *35*, 791–801.
44. Lu, H.; Li, X.; Zhang, L.; Ruan, X.; Yang, M. Dense and sparse reconstruction error based saliency descriptor. *IEEE Trans. Image Process.* **2016**, *25*, 1592–1603.
45. Li, J.; Zhang, H.; Zhang, L. Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5338–5351.
46. Fang, L.; Li, S.; Duan, W.; Ren, J.; Benediktsson, J.A. Classification of hyperspectral images by exploiting spectra-spatial information of superpixel via multiple kernels. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6663–6674.
47. Saranathan, A.M.; Parente, M. Uniformity-based superpixel segmentation of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1419–1430.
48. Wang, Y.; Zhang, Y.; Song, H.W. A Spectral-texture kernel-based classification method for hyperspectral images. *Remote Sens.* **2016**, *8*, 919.
49. Wang, Q.; Lin, J.; Yuan, Y. Salient band selection for hyperspectral image classification via manifold ranking. *IEEE Trans. Neural Netw.* **2016**, *27*, 1279–1289.
50. Tong, N.; Lu, H.; Zhang, L.; Ruan, X. Saliency detection with multiscale superpixels. *IEEE Signal Process. Lett.* **2014**, *21*, 1035–1039.
51. Tan, N.; Xu, Y.; Goh, W.B.; Liu, J. Robust multi-scale superpixel classification for optic cup location. *Comput. Med. Imag. Graph.* **2015**, *40*, 182–193.
52. Neubert, P.; Protzel, P. Beyond holistic descriptors, keypoints, and fixed patches: multiscale superpixel grids for place recognition in changing environments. *IEEE Robot. Autom. Lett.* **2016**, *1*, 484–491.
53. Tropp, J.A.; Gilbert, A.C.; Strauss, M.J. Algorithms for simultaneous sparse approximation. part I: Greedy pursuit. *Signal Process.* **2006**, *86*, 572–588.
54. Jolliffe, I.T. *Principal Component Analysis*; Wiley: New York, NY, USA, 2005.
55. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*; Prentice Hall: Englewood Cliffs, NJ, USA, 2009.
56. Chacon, M.I.; Corral, A.D. Image complexity measure: a human criterion free approach. In Proceedings of the 2005 Annual Meeting of the North American Fuzzy Information Processing Society, Detroit, MI, USA, 26–28 June 2005; pp. 241–246.
57. Silva, M.P.D.; Courboulay, V.; Estrailier, P. Image complexity measure based on visual attention. In Proceedings of the eighteenth IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; pp. 3281–3284.
58. Cardaci, M.; Gesù, V.D.; Petrou, M.; Tabacchi, M.E. A fuzzy approach to the evaluation of image complexity. *Fuzzy Sets Syst.* **2006**, *160*, 1474–1484.
59. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 167–181.
60. Veksler, O.; Boykov, Y.; Mehrani, P. Superpixels and supervoxels in an energy optimization framework. In Proceedings of the eleventh European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; pp. 211–224.

61. Liu, M.Y.; Tuzel, O.; Ramalingam, S.; Chellappa, R. Entropy-rate clustering analysis via maximizing a submodular function subject to matroid constraint. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 99–112.
62. Nemhauser, G.L.; Wolsey, L.A.; Fisher, M.L. An analysis of approximations for maximizing submodular set functions. *Math. Prog.* **1978**, *14*, 265–294.
63. Fang, L.; Li, S.; Kang, X.; Benediktsson, J.A. Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7738–7749.
64. Fang, L.; Li, S.; Kang, X.; Benediktsson, J.A. Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4186–4201.
65. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27.



© 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/>).