

Article

# A Spectral-Spatial Classification of Hyperspectral Images Based on the Algebraic Multigrid Method and Hierarchical Segmentation Algorithm

Haiwei Song<sup>1,2</sup> and Yi Wang<sup>1,2,\*</sup>

<sup>1</sup> Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China; songchaocug@hotmail.com

<sup>2</sup> Subsurface Multi-Scale Imaging Laboratory of Hubei Province (SMIL), China University of Geosciences, Wuhan 430074, China

\* Correspondence: cug.yi.wang@gmail.com; Tel.: +86-27-6788-3251

Academic Editors: Lenio Soares Galvao, Parth Sarathi Roy and Prasad S. Thenkabal

Received: 31 December 2015; Accepted: 21 March 2016; Published: 31 March 2016

**Abstract:** The algebraic multigrid (AMG) method is used to solve linear systems of equations on a series of progressively coarser grids and has recently attracted significant attention for image segmentation due to its high efficiency and robustness. In this paper, a novel spectral-spatial classification method for hyperspectral images based on the AMG method and hierarchical segmentation (HSEG) algorithm is proposed. Our method consists of the following steps. First, the AMG method is applied to hyperspectral imagery to construct a multigrid structure of fine-to-coarse grids based on the anisotropic diffusion partial differential equation (PDE). The vertices in the multigrid structure are then considered as the initial seeds (markers) for growing regions and are clustered to obtain a sequence of segmentation results. In the next step, a maximum vote decision rule is employed to combine the pixel-wise classification map and the segmentation maps. Finally, a final classification map is produced by choosing the optimal grid level to extract representative spectra. Experiments based on three different types of real hyperspectral datasets with different resolutions and contexts demonstrate that our method can obtain 3.84%–13.81% higher overall accuracies than the SVM classifier. The performance of our method was further compared to several marker-based spectral-spatial classification methods using objective quantitative measures and a visual qualitative evaluation.

**Keywords:** algebraic multigrid methods; classification; hyperspectral images; marker selection; spectral-spatial; hierarchical segmentation

## 1. Introduction

Hyperspectral imaging systems can acquire numerous contiguous spectral bands throughout the electromagnetic spectrum. Therefore, hyperspectral imaging techniques are widely used for many applications, including environmental monitoring, mineralogy, astronomy, surveillance and defense [1]. Nevertheless, the high dimensionality of the pixels, undesirable noise, high spectral redundancy and spectral and spatial variabilities, in conjunction with limited ground truth data, present challenges for the analysis of hyperspectral imagery. In recent decades, many intensive hyperspectral image classification methods have been proposed, including ensemble learning [2–4], Bayesian approaches [5–7], kernel-based methods [8,9], neural networks [10], sparse representations [11,12] and manifold learning [13].

To further improve classification performance, many contributions have been dedicated to spectral-spatial classification, which combine spatial contextual information with spectral information

for hyperspectral image classification; the approaches include using co-occurrence [14], extended morphological profiles [15,16], extended morphological attribute profiles [17], Gabor filtering [18] and multihypothesis prediction [19] to extract spatial features and Markov random fields [20,21], graph kernels [22] and composite kernels [23] to perform spectral-spatial classification.

Alternatively, spatial information can also be integrated into the spectral-spatial classification process by performing image segmentation. For that purpose, many segmentation techniques have been proposed, including watershed, partitional clustering, the HSEG algorithm, and minimum spanning forest (MSF), to segment hyperspectral imagery into homogeneous regions according to a homogeneity criterion. A maximum vote decision rule is then employed to classify all the homogeneous regions according to a pixel-wise classification result. Because the technique can define large-scale neighborhoods for large homogeneous regions without missing small regions, more accurate classification results can be achieved than by traditional spectral-spatial methods. However, the automatic segmentation of hyperspectral imagery is still an continuing problem. To remedy this problem, a marker-controlled segmentation method was proposed to automatically select a single hierarchical segmentation level [24]. The idea behind this approach is to select at least one pixel for each spatial object and to grow regions from the selected seeds (also called markers) to guarantee that each region is associated with one marker in the segmentation maps. Meanwhile, the experimental results in [24,25] demonstrated that the classification accuracies of the HSEG algorithm and MSF segmentation algorithms using automatically selected markers can greatly outperform the SVM classifier. Nevertheless, the drawbacks of these approaches are twofold. First, the marker selection methods are based on the performance of pixel-wise classifiers, *i.e.*, the performance of different pixel-wise classifiers leads to different markers and uncertainties in the classification results. Even if the same classifier is applied, using different parameter settings may cause a similar problem. Second, the randomness of the training samples in the pixel-wise classification procedure always generates stochastic markers, which results in unstable classification accuracies.

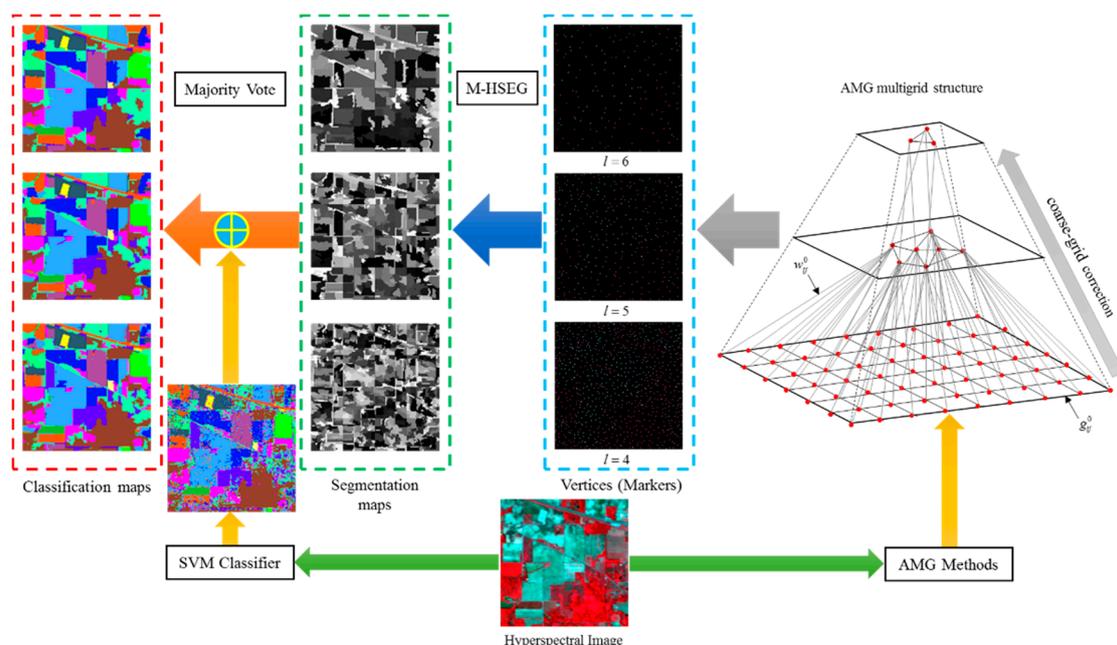
To solve these problems, a novel spectral-spatial classification method for hyperspectral imagery based on the AMG method and HSEG algorithm is presented. Our method includes four main steps. First, a multiscale representation of the hyperspectral imagery is obtained using the AMG method to solve an anisotropic diffusion PDE. Recently, PDE-based techniques have been applied to image processing and computer vision due to their outstanding edge-preservation smoothing properties and explicit accounting of intrinsic geometries. The scale-space representation of hyperspectral imagery can be obtained by solving a classical anisotropic diffusion PDE using the AMG method for higher accuracies, efficiencies and scalability. Following this idea, we applied the AMG method to construct a multigrid structure with different grid levels, starting with the finest grid (which corresponds to the original image) to the coarsest one. The vertices in the multigrid structure are then considered as markers for region growing and are clustered to obtain a sequence of segmentation results with different numbers of homogeneous regions. It should be remarked that the markers that are extracted from the multigrid structure are only determined by the structures of the hyperspectral imagery rather than training samples or the performance of pixel-wise classifiers, which is more robust than the traditional methods. In addition, by combining the pixel-wise SVM classification map and the unsupervised segmentation maps using a maximum vote decision rule, a series of spectral-spatial classification maps are obtained. Finally, a final classification map is produced by choosing the optimal grid level to extract representative pixels for our segmentation and classification. Multiscale representations for hyperspectral imagery using the AMG method always contain series of grid levels. In fact, each grid level can be utilized to extract the required representative pixels. As a consequence, a sequence of segmentation maps is produced for the subsequent classification process. In this paper, we will discuss an optimization problem to automatically select an optimal grid level to achieve the best classification accuracies.

The remainder of this paper is organized as follows. Section 2 presents the proposed spectral-spatial classification of hyperspectral imagery based on the AMG method and HSEG algorithm.

Section 3 describes the experimental results, Section 4 is a discussion of our method, and Section 5 presents our concluding remarks.

## 2. Materials and Methods

A flow-chart of the proposed classification method is illustrated in Figure 1. Because the AMG method is more robust for PDE-based image analysis [26], we utilized it to solve the hyperspectral anisotropic diffusion PDE. First, the hyperspectral anisotropic diffusion PDE is briefly discussed in this section. The AMG method should construct a multigrid structure, which is visualized in Figure 1 as a pyramid, starting with the finest grid (which corresponds to the original image) on its base and the coarsest grid on its top. The detailed introduction of the AMG algorithm is then presented. Once the multigrid structure based on the AMG method is constructed for the hyperspectral image, we used the representative spectra in Figure 1 as markers for region growing in the HSEG algorithm. A step-by-step procedure of the AMG-derived M-HSEG method is also provided. As shown in Figure 1, a multiscale representation of hyperspectral imagery using the AMG method always contains a series of grid levels, each of which can be utilized to extract the required representative pixels. The selection of an optimal grid level to achieve the best classification accuracies is finally introduced.



**Figure 1.** Spectral-spatial classification of hyperspectral images based on the AMG method and marker-based HSEG algorithm (M-HSEG). The original Indian Pines data set is visualized by an RGB composite of bands 47, 23, and 13; the marker maps are visualized by an RGB composite of bands 47, 23, and 13 using the representative spectra, which are used as the initial seeds (markers) for the HSEG algorithm. The segmentation maps correspond to the marker maps, and the different grey levels represent different region labels.

### 2.1. Hyperspectral Anisotropic Diffusion PDE

It is well known that the linear scale-space can be represented by the heat diffusion equation. However, the disadvantage of the equation is that the edge features are smeared and distorted after a few iterations of the diffusion evolution. To overcome the problem of the linear scale-space, Perona and Malik [27] presented an anisotropic diffusion PDE to encourage intraregional smoothing while

preventing interregional smoothing. Let  $u^0 : \Omega \subset \mathbb{R}^2$  denote an original grey-level image, and  $\Omega$  is the image domain. The anisotropic diffusion PDE is

$$\frac{\partial u}{\partial t} = \operatorname{div} (g(|\nabla u|) \nabla u) \quad (1)$$

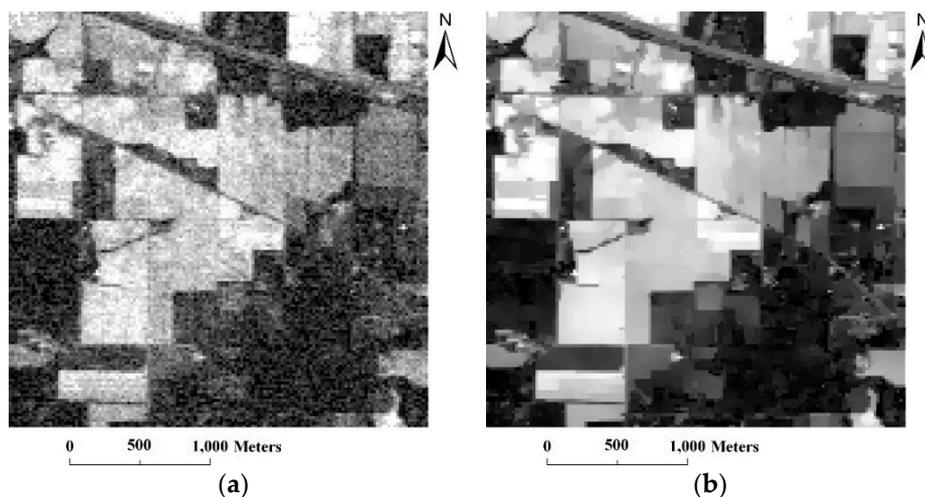
where  $u$  is the smoothed image at time  $t$ , and  $\nabla$  and  $\operatorname{div}$  are the gradient and divergence operators, respectively.  $g(s)$  is the diffusion coefficient, which is defined as a non-negative monotonically decreasing function of the local gradient magnitude  $s = |\nabla u|$ , and several forms of  $g(\cdot)$  have been widely exploited for Equation (1) [28–30]. We can obtain a series of smoothed images by iteratively evolving Equation (1) starting from the observed image at  $t = 0$ , which constitutes a nonlinear scale space. To construct a scale-space representation of the hyperspectral imagery, a vector-valued anisotropic diffusion PDE is used:

$$\frac{\partial u_i}{\partial t} = \operatorname{div} (g(\theta(\nabla \mathbf{u}_\sigma)) \nabla u_i), i = 1, 2, \dots, N \quad (2)$$

where  $N$  is the number of bands,  $\mathbf{u} = (u_1, u_2, \dots, u_N)$ , and  $\mathbf{u}_\sigma$  is obtained by convolving  $\mathbf{u}$  with a Gaussian kernel of standard deviation  $\sigma$ .  $g(\cdot)$  is the diffusion function of  $|\nabla \mathbf{u}_\sigma|$ . In our computations, the diffusion coefficient proposed in [31] is utilized because it can produce segmentation-like images:

$$g(\theta) = \begin{cases} 1, & \theta = 0 \\ \frac{3.31488}{1 - e^{-(\theta/K)^8}}, & \theta > 0 \end{cases} \quad (3)$$

where  $\theta$  is a measure of the image edge strength and can be discretized as the Euclidean distance (ED) or the spectral angle mapper (SAM) between two pixel vectors.  $K$  denotes a threshold that modulates the amount of diffusion with respect to  $\theta$ . A simple example of the hyperspectral anisotropic diffusion process is illustrated in Figure 2. It can be observed that sensor noise in the hyperspectral image is effectively removed, and edge features are better preserved. In addition, the homogeneous regions are very smooth, which is greatly preferable for segmentation and classification.



**Figure 2.** Example of the hyperspectral anisotropic diffusion PDE process for the Indian Pines data set: (a) original noisy band 185; and (b) diffused result.

To solve PDE Equation (2), a first order discretization in time is used to approximate  $\partial u_i / \partial t$  by  $(u_i^{n+1} - u_i^n) / \tau$ , where  $\tau$  is the time step size, and  $u_i^n$  is the solution of Equation (2) at time  $t_n = n \cdot \tau$ .

An explicit (or Euler forward difference) scheme is widely used to solve Equation (2) and control the diffusion by the diffusion coefficients computed from the previous time step using

$$\frac{u_i^{n+1} - u_i^n}{\tau} = \text{div} (g(\theta(\nabla \mathbf{u}_\sigma^n)) \nabla u_i^n) \quad (4)$$

The solution  $\mathbf{u}^{n+1}$  is obtained explicitly from  $\mathbf{u}^n$  by simply rearranging Equation (6) as

$$u_i^{n+1} = u_i^n + \tau \cdot \text{div} (g(\theta(\nabla \mathbf{u}_\sigma^n)) \nabla u_i^n) \quad (5)$$

Using space discretization, the explicit scheme can be given by

$$u_i^{n+1} = (\mathbf{I} + \tau \mathbf{A}(\mathbf{u}_\sigma^n)) u_i^n \quad (6)$$

where  $\mathbf{I}$  and  $\mathbf{A}$  are the identity matrix and the matrix of diffusion coefficients, respectively. However, this scheme requires a maximum possible time step size  $\tau \leq 1/4$  to ensure its stability, which means the computation efficiency is very low. An alternative discretization of Equation (2) employs semi-implicit schemes, which are stable for all time step sizes. The traditional semi-implicit scheme of the diffusion PDE can be calculated by a series of linear equations,

$$(\mathbf{I} - \tau \mathbf{A}(\mathbf{u}_\sigma^n)) u_i^{n+1} = u_i^n \quad (7)$$

The tridiagonal matrix  $\mathbf{I} - \tau \mathbf{A}(\mathbf{u}_\sigma^n)$  can be inverted using the Gauss-Seidel or preconditioned conjugate gradient method. However, it is necessary to solve large linear systems at each iteration step. As shown in [32], semi-implicit schemes such as additive operator splitting (AOS) and alternative direction implicit (ADI) schemes can significantly speed up the evolution of the PDE Equation (2).

## 2.2. Multiscale Representation of Hyperspectral Imagery

Following the work of [33], we construct a multigrid structure that is obtained by applying the AMG method to solve Equation (7). The multiscale representation for hyperspectral imagery by the AMG method is based on graph theory. Initially, the first graph  $(\mathbf{V}^0, \mathbf{E}^0)$  is built from the original hyperspectral image, where  $\mathbf{V}^0$  denotes the set of vertices and has the same size as the hyperspectral image, and  $\mathbf{E}^0$  represents the set of edges that connect each vertex to its four neighborhoods with weights. In our method, the initial weights  $g_{ij}^0$  of  $(i, j) \in \mathbf{E}^0$  are computed using the diffusion coefficient Equation (3). To build an AMG multigrid structure, the authors of [33] introduced a mass  $m_i$  for each vertex, which is a measure of the number of pixels assigned to a given vertex selected for the next grid and can be initialized as  $m_i^0 = 1$ .

The construction of the multigrid structure mainly consists of the two following steps.

- The first step is the consecutive selection of a new set of  $\mathbf{V}^{l+1}$  from  $\mathbf{V}^l$ , where  $l$  is used to denote the current grid level with  $0 \leq l \leq S$ , where  $S$  is the coarsest grid. The set of vertices  $\mathbf{V}^l$  is sorted in decreasing order according to  $m_i^l$ . Then, the first vertex of  $\mathbf{V}^{l+1}$  is initialized as the vertex in  $\mathbf{V}^l$  that has the highest mass. Finally, the set of vertices at grid  $l+1$  can be obtained:

$$\frac{\sum_{j \in \mathbf{V}^{l+1}} g_{ij}^l}{\sum_{(i,j) \in \mathbf{E}^l} g_{ij}^l} \leq v \Rightarrow \mathbf{V}^{l+1} = \mathbf{V}^l \cup \{i\}, \text{ for each } i \in \mathbf{V}^l \setminus \mathbf{V}^{l+1} \quad (8)$$

where  $0 < v < 1$  is a threshold value, and  $\mathbf{V}^l \setminus \mathbf{V}^{l+1}$  denotes the set difference between  $\mathbf{V}^l$  and  $\mathbf{V}^{l+1}$ . Note that  $m_i^0, g_{ij}^0$  and  $\mathbf{E}^0$  in the finest grid are initialized as before, and we can compute the coarser grid according to the previous algorithm. Once the vertices of the finest coarse grid are

constructed, we compute the masses in the next grid and the dependence degrees of the vertices in  $\mathbf{V}^l \setminus \mathbf{V}^{l+1}$  to the vertices in  $\mathbf{V}^l$  as

$$\forall i \in \mathbf{V}^{l+1} : m_i^{l+1} = m_i^l + \sum_{j \in \mathbf{V}^{l+1}} w_{ij}^l \text{ and} \tag{9}$$

and

$$\forall i \in \mathbf{V}^l \setminus \mathbf{V}^{l+1}, j \in \mathbf{V}^{l+1} : w_{ij}^l = w_{ji}^l = \frac{g_{ij}^l}{\sum_{k \in \mathbf{V}^{l+1}} g_{ik}^l} \tag{10}$$

where  $w_{ij}^l$  is a measure of how much vertex  $i \in \mathbf{V}^l \setminus \mathbf{V}^{l+1}$  depends on vertex  $j \in \mathbf{V}^{l+1}$ .

- The second step is performed by connecting the vertices in  $\mathbf{V}^{l+1}$  to obtain  $\mathbf{E}^{l+1}$ , which can be realized first by computing  $g_{ij}^{l+1}$  for all the vertices in grid level  $l + 1$  and then connecting the vertices to obtain  $\mathbf{E}^{l+1}$ . Based on the Garlekin principle [34], the matrix of diffusion coefficients can be defined as the Garlekin operator  $\mathbf{G}^{l+1} = I_f^c \mathbf{G}^l I_c^f$ , where  $I_f^c$  denotes the restriction operator that maps vectors in a fine grid into a coarser one and is given by

$$\left[ I_f^c \right]_{ij} = \frac{w_{ij}^l}{1 + \sum_{j \in \mathbf{V}^l \setminus \mathbf{V}^{l+1}} w_{ij}^l} \tag{11}$$

where  $I_c^f$  is the interpolation operator, which is used to interpolate the intensity back to the finer grids and is given by

$$\left[ I_c^f \right]_{ij} = w_{ij}^l \tag{12}$$

By combining Equations (11) and (12) into the Garlekin operator, we obtain

$$g_{ij}^{l+1} = \frac{1}{1 + \sum_{j \in \mathbf{V}^l \setminus \mathbf{V}^{l+1}} w_{ij}^l} \sum_{p,q \in \mathbf{V}^l} w_{ip}^l g_{pq}^l w_{qj}^l \tag{13}$$

We can then connect the vertices to obtain  $\mathbf{E}^{l+1}$ :

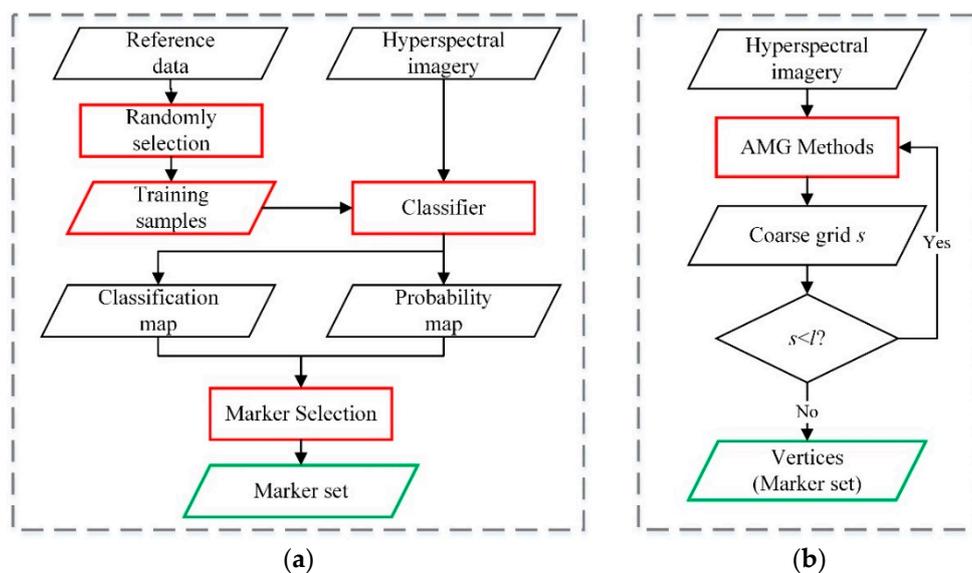
$${}^{l+1} = \left\{ (i, j) : i, j \in \mathbf{V}^{l+1} \wedge g_{ij}^l > 0 \right\} \tag{14}$$

Finally, the set of vertices and edges in grid level  $l + 1$  are constructed. Therefore, the representative spectra in the coarse grid are determined by the position of the vertices that correspond to grid level  $l + 1$ .

Anisotropic diffusion PDEs are always used for multiscale image analyses to construct nonlinear scale spaces. As time elapses, noise in the hyperspectral image is effectively removed, and edge features are better preserved. In addition, homogeneous regions become very smoothed, which is greatly preferable for segmentation and classification, as shown in Figure 2. However, the solution of the hyperspectral anisotropic diffusion PDE imposes a heavy computation burden. The AMG method can construct a hierarchical representation of the PDE solver from the finest grid (which corresponds to the original image) to the coarsest grid, *i.e.*, the linear system can be solved in the coarser grid with higher accuracy, efficiency and scalability. On the other hand, with respect to graph  $(\mathbf{V}^l, \mathbf{E}^l)$ , the diffusion coefficient Equation (3) is used as a similarity function to assign a weight to each edge in  $\mathbf{E}^l$ . According to the definition of the diffusion coefficient, if  $g_{ij}$  is close to 1, the pixel is inside a homogeneous region, whereas if  $g_{ij}$  is close to 0, the pixel is near the edges. As a consequence, when the set of vertices  $\mathbf{V}^l$  is sorted in decreasing order of mass, the first vertex can correspond to a pixel

inside a spectrally uniform region. Therefore, the anisotropic diffusion PDE is suitable for marker selection, and the acquired representative spectra is more reliable for the HSEG algorithm.

In the multigrid structure, each vertex selected by the AMG method can be exploited as a pioneer for a certain area in the hyperspectral image. The main problem is the use of the selected vertices for marker-based segmentation. In fact, we can use vertices in any grid level as the initial seeds for region growing. In [24], two automatic marker selection techniques were introduced to obtain the most reliable classified pixels as markers, *i.e.*, marker selection approaches using morphological filtering (Morpho-MS) and probabilistic SVM (Proba-MS). To compare those two classification-derived marker selection methods with ours, flow-charts of the two schemes are illustrated in Figure 3. It can be observed in Figure 3a that the choice of markers in the classification-derived methods is always strongly dependent on the performance of the pixel-wise classifiers. Even if the same classifier is used, using different parameter settings can also produce different classification maps, which impacts the final markers. In addition, the selected markers can be very different due to the randomness of the training samples, which causes uncertainty in the classification maps. However, the final markers produced by our method are only determined by the structure of the hyperspectral image, as shown in Figure 3b, which is more robust than the traditional marker selection methods. Furthermore, the AMG-derived marker selection method requires fewer adjustment parameters than classification-derived marker selection techniques.



**Figure 3.** Flow-charts of the classification-derived marker selection method and our method: (a) the classification-derived marker selection scheme; and (b) the AMG-derived marker selection scheme.

### 2.3. AMG-Derived M-HSEG Method

Once the multigrid structures are constructed by the AMG method, we integrate the multiscale representation with the HSEG algorithm to obtain unsupervised segmentation maps. The basic idea of our segmentation method is that vertices in the multigrid structure can be considered as initial seeds (markers) for region growing and are clustered to obtain a sequence of segmentation results. The AMG-derived M-HSEG segmentation algorithm is briefly described as follows.

---

**Algorithm 1: AMG-derived M-HSEG Segmentation Algorithm**


---

**Input:** An original hyperspectral image  $\mathbf{u}$  and the coarsest grid level  $S$ .

**Output:** Segmentation maps

---

1. Input a hyperspectral image and construct an undirected graph as the finest grid.
  2. AMG Relaxation
    - At the finest grid level, perform a Gauss-Seidel relaxation to solve  $(\mathbf{I} - \tau\mathbf{G}^0)\mathbf{X}^0 = \mathbf{u}$  with an initial guess image  $\mathbf{u}$  and compute the error  $\mathbf{X}^0 = (\mathbf{I} - \tau\mathbf{G}^0)\mathbf{X}^0 - \mathbf{u}$ .
    - At the coarser grid level  $l$  ( $0 < l \leq S$ ), perform a Gauss-Seidel relaxation to solve the residual equation  $(\mathbf{I} - \tau\mathbf{G}^l)\mathbf{X}^l = \mathbf{F}^l$  with an initial guess  $\mathbf{0}$ , and compute the error  $\mathbf{X}^l = \mathbf{F}^l - (\mathbf{I} - \tau\mathbf{G}^l)\mathbf{X}^l$  and then the residual  $\mathbf{F}^l = (\mathbf{I} - \tau\mathbf{G}^l)\mathbf{X}^l$ .
  1. AMG Coarse-Grid Correction: Select the set of vertices in  $\mathbf{V}^l$  for  $\mathbf{V}^{l+1}$  to obtain  $\mathbf{F}^{l+1}$  for the coarser grid level  $l + 1$ .
  2. Compute  $\mathbf{G}^{l+1}$  and connect the nodes in  $\mathbf{V}^{l+1}$  to obtain  $\mathbf{E}^{l+1}$ .
  3. If  $l \leq S$ , go to step 2; otherwise, go to the next step.
  4. Initialize the vertices in grid  $l$  as markers for the subsequent HSEG algorithm by assigning each vertex a non-zero marker label and each pixel as a separate region.
  5. Perform the M-HSEG algorithm by using the markers obtained from grid  $l$  of the hyperspectral image:
    - (a) Calculate the dissimilarity criterion (DC) values between all pairs of spatially adjacent regions. It should be noted that we only calculate the DC value between a marked pixel and a non-marked pixel and merge the pair of adjacent pixels that has the smallest DC value.
    - (b) Merge the pair of adjacent pixels that has the smallest DC value.
    - (c) Stop when there is no more merging, which means that the DC value is NaN.
  3. Obtain the resultant segmentation maps for the subsequent classification.
- 

The AMG-derived M-HSEG algorithm can be divided into three procedures. (i) In the coarsening grid procedure, the error  $\mathbf{X}^l$  ( $0 < l \leq S$ ) is first estimated by solving  $(\mathbf{I} - \tau\mathbf{G}^l)\mathbf{X}^l = \mathbf{F}^l$  using the Gauss-Seidel relaxation, where  $\mathbf{F}^l$  is the residual. The masses, the matrix of diffusion coefficients and the set of edges are then updated for grid level  $l + 1$ . We perform this procedure until the  $S$ -level multigrid structure is constructed; (ii) In the marker selection procedure, all the vertices in grid  $l$  are considered as the initial seeds (markers) for region growing. Because the scale-space can provide several coarse grid levels, we can obtain a series of segmentation maps; (iii) In the HSEG algorithm procedure, we only compute the DC value between a marked pixel and a non-marked pixel and merge the pair of adjacent pixels that has the smallest DC value, which can reduce the computational burden. To further improve the computational efficiency, the RHSEG strategy [35] is also employed. As a consequence, the number of regions in the resultant segmentation map is equal to the number of marker sets.

#### 2.4. Selection of the Optimal Grid Level

With respect to the AMG-based scale space, if we choose the finest grid for marker extraction, a large number of representative spectra will be selected. Consequently, a segmentation map is produced with many small regions, which is very difficult to extract in large homogeneous areas. On the contrary, if we choose the very coarse grid, the number of markers is very small, and only a few large regions

exist in the segmentation map. The key problem is choosing the optimal grid level from the multigrid structure for classification. According to the above analyses, we can achieve the best classification accuracies by choosing a certain grid level  $l$  ( $0 \leq l \leq S$ ) for marker extraction from the AMG multigrid structure. All of our experiments on hyperspectral imagery, including those not reported here, confirm that the classification accuracy achieved by our method can be considered as a concave function of the number of markers. Inspired by this rule, we can provide a strategy for automatically selecting an optimal value of  $l$ . Therefore, the parameter tuning can be converted into an optimization problem:

$$l_{\text{opt}} = \underset{l}{\operatorname{argmax}} \text{OA}(l) \tag{15}$$

where  $l_{\text{opt}}$  is the index of the optimal grid level, and OA denotes the overall accuracy of our classification. It is possible to compare the resultant images by different values of  $l$  to obtain a solution to Equation (11) with the best OA value, which corresponds well with a parameter tuning procedure for marker selection.

### 3. Results

#### 3.1. Parameter Settings and Evaluation Measures

To enable a better understanding of the maximum vote decision rule in our method, an illustrative example of the integration process using the majority voting step is depicted in Figure 4. As mentioned in Section 2.3, the AMG-derived M-HSEG algorithm is used to segment the hyperspectral image into different regions with region labels, as shown in Figure 5b. To assign each region an information label, we integrate the unsupervised segmentation map and the pixel-wise SVM classification map by applying majority voting within this region in the segmentation map. For each region in Figure 5b, its class label is assigned to the most frequent class in the pixel-wise classification map in Figure 5a within this region. In this way, the advantages of both the pixel-wise SVM classification and the AMG-derived M-HSEG algorithm are combined; the resultant classification map is shown in Figure 5c. It should be noted that a marker may be assigned to the wrong class by most of the marker-based classification methods. Therefore, all pixels within the region grown from this marker are at risk of being wrongly classified. To tackle this problem, the majority voting step is widely used. Nevertheless, the purpose of the step in our method is to perform spectral-spatial classification on all the homogeneous regions in the unsupervised segmentation map by combining spectral and spatial information because we have no idea which class a marker should belong to with respect to the ground truth.

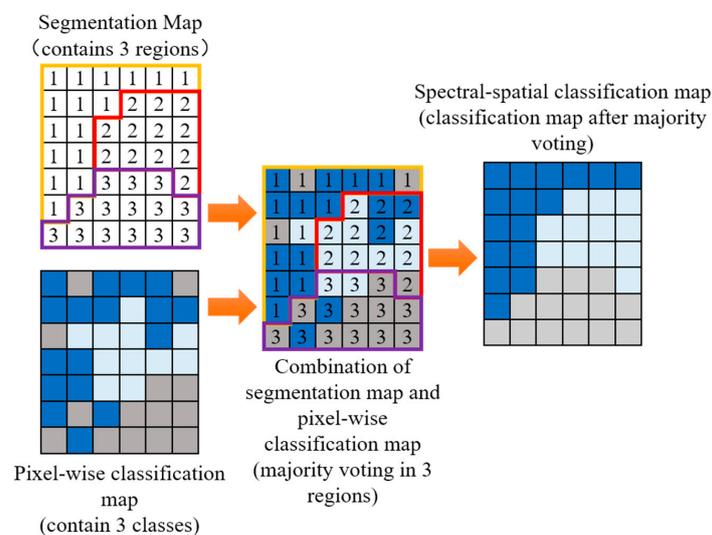
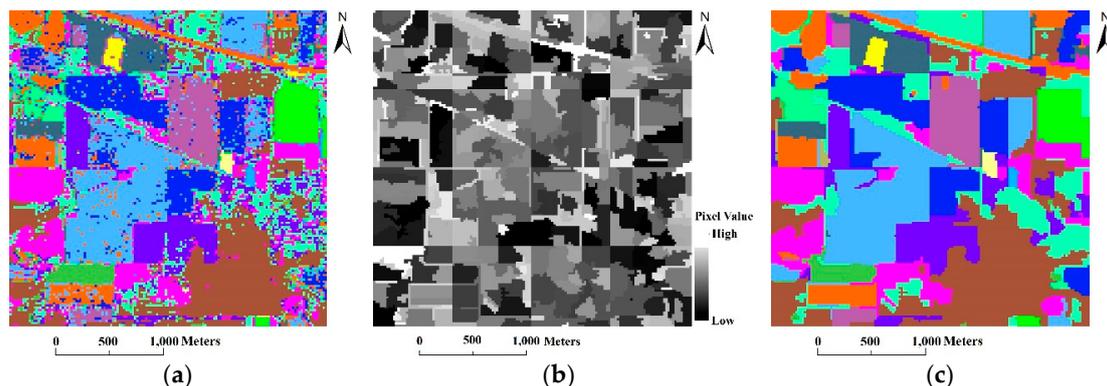


Figure 4. Schematic example of using majority voting within segmentation regions.



**Figure 5.** Spectral-spatial classification for the Indian Pines image using majority voting within segmentation regions: (a) the pixel-wise SVM classification map; (b) the AMG-derived M-HSEG segmentation map; and (c) the final classification map.

In this section, several hyperspectral classification methods are compared to the proposed AMG-derived M-HSEG (AMG-M-HSEG) classification method. (1) In the pixel-wise SVM classification method, the parameters are optimally set for each data set. (2) Two marker-based spectral-spatial classification methods proposed by Tarabalka *et al.* [36] used a HSEG algorithm following the classification-derived marker selection methods; these methods were used both without and with the optional majority voting under the rule that the class label of each region is given to the class with the maximum pixels within this region in the classification map. These methods are named “Morph-M-HSEG”, “Morph-M-HSEG + MV”, “Proba-M-HSEG” and “Proba-M-HSEG + MV”, respectively. (3) Another marker-based spectral-spatial classification method proposed by Tarabalka *et al.* [25] uses an MSF construction following the Proba-MS marker selection method, which are also used without and with the optional majority voting step. These methods are named “Proba-M-MSF” and “Proba-M-MSF + MV”, respectively.

(1) Because the merging of spatially non-adjacent regions always creates a large computational burden, the optional parameter  $S_{weight}$  is set as  $S_{weight} = 0.0$  to improve the computational efficiency for all the hyperspectral images in our experiments, which means that only spatially adjacent regions are merged in the HSEG step.

(2) To increase the computational efficiency, four-neighborhood connectivity is exploited in the HSEG algorithm and the MSF construction algorithm.

(3) Because two different similarity metric measures are commonly used for hyperspectral images to discretize  $\theta$  in Equation (3), *i.e.*, the ED and the SAM between spectral vectors, we apply those two measures for computing the DCs between the regions for the HSEG and the weights of the edges for the MSF construction, respectively. It should be remarked that our experiments on the images used in the paper demonstrate that both the Proba-M-HSEG (+MV) and Morph-M-HSEG (+MV) methods using the ED measure always result in inaccurate or false segmentation and classification maps because the ED measure cannot provide a satisfactory dissimilarity measure between the region mean vectors for the M-HSEG algorithm. Therefore, these two classification methods using the ED measure, both without and with the optional majority voting step, are not considered in our following experiments for comparison. In addition, other similarity metric measures such as the  $L1$  vector norm and the spectral information divergence (SID) can be used as well.

(4) The parameters for our method are set according to former research on the AMG method [33]. In our experiments, we set  $\tau = 1$ ,  $v = 0.2$ , and  $K = 0.01$ . In addition, we used the method described in [33] to determine the coarsest grid  $S$ , *i.e.*, if the number of vertices in any grid starting with the finest grid is equal or less than  $\log_2 U$  ( $U$  is equal to the original image size), the construction of the multigrid structure stops, and the coarsest grid  $S$  can be automatically obtained.

(5) The multiclass one-*vs.*-one SVM classifier with a Gaussian radial basis function (RBF) kernel is used on the hyperspectral data sets. SVM has been the most frequently used method and can achieve higher classification accuracies than traditional pixel-wise techniques when a limited number of training data sets are available. Refer to [37–39] for details on SVM. As a consequence, information classes are defined for the hyperspectral image, and each pixel is given a unique class label. The performance of the methods is objectively evaluated in terms of global accuracy (GA) measures that include the OA, average accuracy (AA), the kappa coefficient  $\kappa$  [40], and class-specific accuracy (CA). Note that these objective measures can be obtained from the confusion matrix.

### 3.2. The Indian Pines Image (AVIRIS)

The first Indian Pines hyperspectral image, which was acquired with the AVIRIS sensor, has  $145 \times 145$  pixels and 220 bands in the 400–2500 nm range, which represent a 2 mile by 2 mile area with a spatial resolution of 20 m. A spectral subset of 185 bands was used for our experiments. Sixteen classes of interest, which are shown in Table 1, were used in our experiments. To perform the supervised classification, 10% of the labeled pixels in each class in the ground truth data were randomly selected as training samples, and the remaining 90% were used as test samples. It can be observed from the ground truth data that some classes only include a very small number of samples, such as *Alfalfa*, *Grass/pasture-mowed* and *Oats*. For each of those classes, we randomly selected 10 training samples, and the remainder of the samples were used for testing. In addition, the SVM classifier parameters  $C$  and  $\gamma$  were optimally obtained using a five-fold cross-validation, and  $C = 8192$ ,  $\gamma = 0.5$ .

**Table 1.** Number of training and test samples and the GAs and CAs (percent) for all the hyperspectral data sets using the SVM classification method.

Class	Indiana Pines			Washington DC				Centre of Pavia				
	Name	Training	Test	SVM	Name	Training	Test	SVM	Name	Training	Test	SVM
1	Alfalfa	10	44	81.82	Grass	52	1000	93.00	Water	39	1940	100
2	Corn-no till	143	1291	76.61	Street	76	1463	94.05	Road	80	3968	90.93
3	Corn-min till	83	751	72.7	Roofs	129	2468	88.29	Trees	64	3145	89.41
4	Corn	23	211	46.45	Shadow	38	741	74.63	Grass	54	2659	92.59
5	Grass/pasture	49	448	86.16	Path	42	812	88.67	Soil	67	3293	98.45
6	Grass/trees	74	673	89.75	Trees	44	840	98.57	Bricks	40	1984	90.57
7	Grass/pasture-mowed	10	16	87.50					Bitumen	49	2421	88.76
8	Hay-windrowed	48	441	97.28					Shadow	30	1503	95.94
9	Oats	10	10	100					Roof-1	39	1959	93.72
10	Soybeans-no till	96	872	83.03					Roof-2	77	3817	74.33
11	Soybeans-min till	246	2222	87.62								
12	Soybeans-clean till	61	553	66.55								
13	Wheat	21	191	96.34								
14	Woods	129	1165	93.30								
15	Bldg-Grass-Trees-Drives	38	342	61.40								
16	Stone-steel towers	10	85	63.53								
	Total	1051	9315		Total	381	7324		Total	539	26,689	
	OA			82.51	OA			89.92	OA			90.39
	AA			80.63	AA			89.54	AA			91.47
	$\kappa$			79.96	$\kappa$			87.31	$\kappa$			89.23

In the Proba-MS method, there are three adjustable parameters ( $M$ ,  $P$  and  $T$ ). Because the maximum size of the connected components for *oats* in the SVM classification map was 19, we set  $M = 19$  for the Proba-MS procedure in the Proba-M-MSF (+MV) and Proba-M-HSEG (+MV) methods. In addition, in the Proba-M-MSF (+MV) method, the parameter  $P$  was computed as  $P = 6\%$ , given the condition that each marker for a large region should have at least one pixel. The last parameter,  $T$ , was set equal to the lowest probability within the highest 2% of the probability estimates, whereas in the Proba-M-HSEG (+MV) method, the parameter settings were set to  $P = 40\%$  and  $T = 50\%$ . In addition, the size of the structuring element in the Morph-M-HSEG (+MV) method was  $3 \times 3$ .

To build the multigrid structure, 10 coarse grid levels ( $S = 10$ ) were constructed according to Section 3.1. Figure 6 shows the impact on the objective quantitative assessments of the GAs caused

by varying  $l$  from 0 to 10. We can observe from these plots the high robustness of the results with respect to values of  $l$  from 1 to 5 when compared to the SVM classification result ( $l = 0$ ). In addition, it can be observed that the shapes of the plots share a similar global behavior. As  $l$  is increased, all the GAs rise gradually until reaching a peak. After the maximum, the OA and  $\kappa$  values drop from 96.32% and 95.80% ( $l = 5$ ) to 51.05% and 42.61% ( $l = 10$ ), respectively. However, the AA value drops more quickly than the other two measures, from a high value of 95.95% to 28.6%. This can be explained because for  $l \geq 6$ , all the pixels in the classification maps, which should belong to *Grass/pasture-mowed* in the ground truth data (refer to Figure 7b), are assimilated with their neighboring structures. As a consequence, the CA of this class is 0, and the corresponding AA is lower. In this experiment, which had a coarse grid level of  $l = 5$  estimated using Equation (15), 397 vertices, which occupy 1.9% of the total number of pixels in the image, are used as markers. It is worth noting that other GAs such as AA and  $\kappa$  can also be used in Equation (15), and the same value of  $l_{opt}$  ( $l_{opt} = 5$ ) will be obtained.

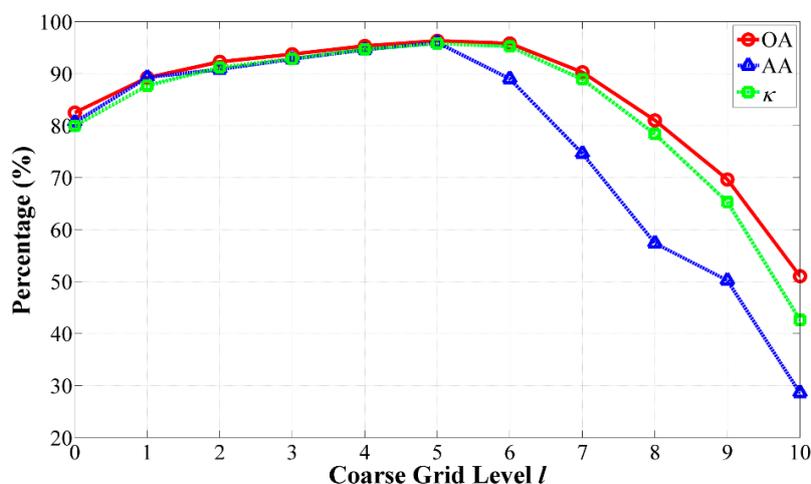


Figure 6. Sensitivity analysis of the GAs using the AMG-M-HSEG method with the Indian Pines image for  $l$  from 0 to 10.

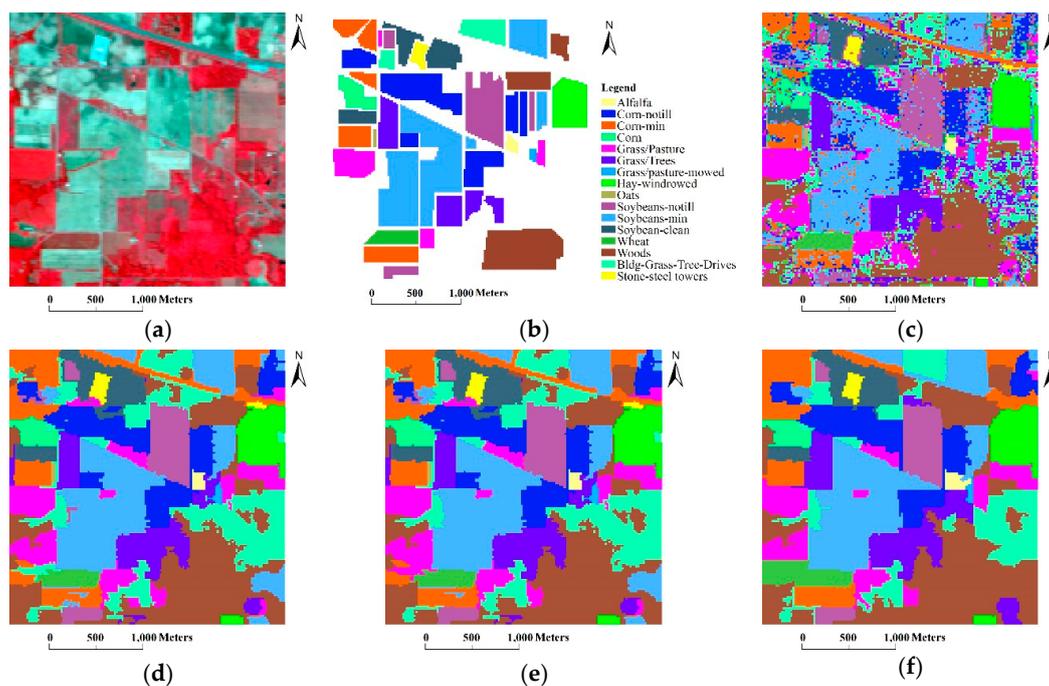
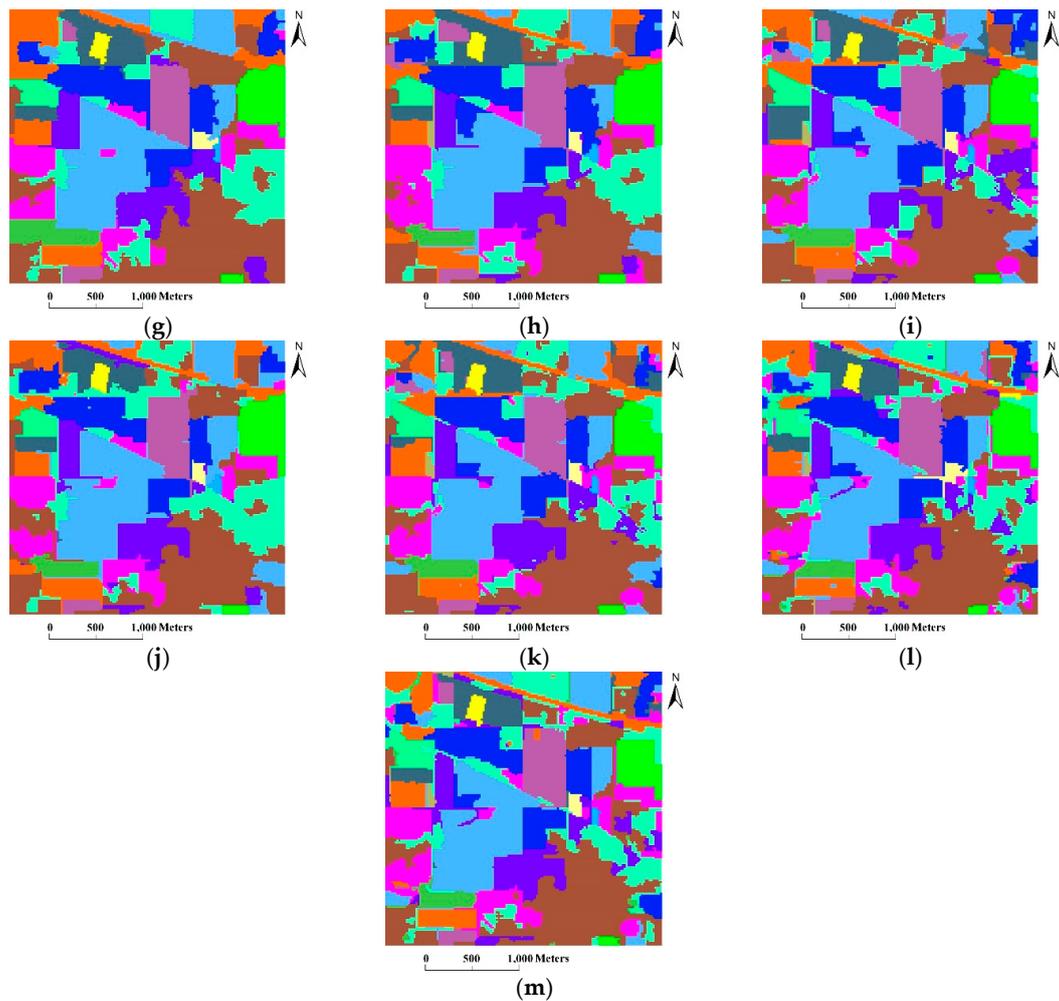


Figure 7. Cont.



**Figure 7.** Classification maps for the Indian Pines image: (a) RGB composite (bands 47, 23, and 13); (b) Ground truth data; (c) SVM; (d) Proba-M-HSEG; (e) Proba-M-HSEG +MV; (f) Morph-M-HSEG; (g) Morph-M-HSEG +MV; (h) Proba-M-MSF using ED; (i) Proba-M-MSF using SAM; (j) Proba-M-MSF +MV using ED; (k) Proba-M-MSF +MV using SAM; (l) AMG-M-HSEG using ED; and (m) AMG-M-HSEG using SAM.

Table 1 lists the number of training and test samples for each class in the ground truth data and the classification accuracies of the SVM classification, and Table 2 lists the classification accuracies of all the marker-based classification methods used here. The RGB composite map from bands 47, 23 and 13 of the AVIRIS image and its ground truth data are depicted in Figure 7a,b, respectively. Figure 7c–m illustrates the corresponding classification maps. From those results, we reached the following conclusions.

**Table 2.** The GAs and CAs (percent) for the Indian Pines image using all the spectral-spatial classification methods for comparison. The highest accuracies are indicated in bold in each category.

	AMG-M-HSEG		Proba-M-HSEG	Proba-M-HSEG + MV	Morph-M-HSEG	Morph-M-HSEG + MV	Proba-M-MSF		Proba-M-MSF + MV	
DC	ED	SAM	SAM	SAM	SAM	SAM	ED	SAM	ED	SAM
OA	95.16	<b>96.32</b>	94.49	94.86	95.04	95.04	93.17	91.44	93.33	94.04
AA	95.21	<b>95.97</b>	93.16	93.35	86.88	86.88	92.95	91.46	86.26	93.91
κ	94.48	<b>95.80</b>	93.71	94.13	94.33	94.33	92.21	90.24	92.37	93.20
Alfalfa	86.36	86.36	86.36	86.36	86.36	86.36	<b>88.64</b>	<b>88.64</b>	<b>88.64</b>	<b>88.64</b>
Corn-no till	92.10	92.33	87.06	87.06	86.91	86.91	<b>94.42</b>	87.61	87.76	87.61
Corn-min till	93.74	96.67	92.81	95.07	<b>97.74</b>	<b>97.74</b>	82.96	77.23	94.54	94.67
Corn	91.47	<b>99.33</b>	82.46	82.46	79.15	79.15	79.15	97.16	79.15	97.16
Grass/pasture	95.76	<b>96.43</b>	94.64	94.64	93.75	93.75	93.75	93.75	92.41	93.75
Grass/trees	97.18	96.43	95.99	95.99	<b>98.07</b>	<b>98.07</b>	96.88	92.27	96.88	96.43
Grass/pasture-mowed	<b>93.75</b>	<b>93.75</b>	<b>93.75</b>	<b>93.75</b>	75.00	75.00	<b>93.75</b>	<b>93.75</b>	<b>93.75</b>	<b>93.75</b>
Hay-windrowed	99.55	93.42	<b>99.77</b>	<b>99.77</b>	<b>99.77</b>	<b>99.77</b>	<b>99.77</b>	99.55	99.55	97.51
Oats	<b>100</b>	<b>100</b>	90.00	90.00	0	0	<b>100</b>	<b>100</b>	0	<b>100</b>
Soybeans-no till	85.44	<b>91.97</b>	88.76	88.76	89.33	89.33	91.86	91.06	85.21	90.25
Soybeans-min till	96.22	98.51	98.92	99.68	<b>99.01</b>	<b>99.01</b>	93.43	94.82	98.47	97.70
Soybeans-clean till	<b>98.73</b>	95.3	97.29	97.29	97.11	97.11	86.44	97.29	90.78	85.35
Wheat	<b>100</b>	98.95	99.48	99.48	99.48	99.48	99.48	<b>100</b>	99.48	<b>100</b>
Woods	99.48	98.88	99.66	99.66	99.83	99.83	99.57	<b>100</b>	99.31	<b>100</b>
Bldg-Grass-Tree-Drives	97.08	<b>98.83</b>	84.80	84.80	89.77	89.77	89.47	52.63	76.61	82.16
Stone-steel towers	96.47	97.65	<b>98.82</b>	<b>98.82</b>	<b>98.82</b>	<b>98.82</b>	97.65	97.65	97.65	97.65

(1) The Morph-M-HSEG and Proba-M-HSEG methods, both with or without the optional majority voting step, can achieve better GAs when compared with the SVM classification. Meanwhile, the highest CAs for 6 of the 16 classes were achieved when using those four methods, including *Corn-min till*, *Grass/trees*, *Grass/pasture-mowed*, *Hay-windowed*, *Soybeans-clean till* and *Stone-steel towers*. However, those methods always resulted in a slight under-segmentation in the HSEG step. For example, it can be observed in Figure 7 that some small regions of the *Corn-no till*, *Oats* and *Grass/pasture* classes were merged by their adjacent regions that belonged to the other classes, or some small regions of the other classes were merged by large regions of the *Corn-min* and *Soybean-clean* classes. In contrast, in the results of our method shown in Figure 7l,m, most of the small regions that belonged to different classes were better preserved. In addition, the majority voting step did not improve the GAs and CAs of the Morph-M-HSEG method using the SAM distance because almost all the pixels in each region in Figure 7f have the same class label.

(2) The Proba-M-MSF method, both with or without the optional majority voting step, can obtain better GAs than the SVM classifier. However, the highest OA by our method is 2%–5% higher when compared with this method. To clearly demonstrate the difference between the two methods, one region at the top-middle of the image was used for comparison. This region should be classified as *Bldg-Grass-Trees-Drives* according to the ground truth data, but a large number of pixels in that region were classified as *Woods* by the Proba-M-MSF method, as shown in Figure 7j,k. By comparison, our method can achieve more accurate classification maps. Apart from these observations, another small region in the top left of the image was used for comparison. It can be observed that this region was correctly classified as *Grass/pasture* by our method, which is consistent with the ground truth data. Nevertheless, the entire region was merged by its spatial adjacent region, which belonged to *Bldg-Grass-Trees-Drives*, by the Proba-M-MSF method.

(3) The GAs achieved by our method using the SAM distance were the best among all the classification methods used for comparison. In this case, the OA and  $\kappa$  increased by 13.81% and 15.84%, respectively. Meanwhile, the highest CAs for 8 of the 16 classes were achieved when using our method. On that occasion, the AA was improved by 15.34%. It is very important to preserve material boundaries and edge structures in classification maps. From the classification maps, we can observe that our method was better than the other marker-based classification methods in terms of region homogenization and edge preservation.

### 3.3. The Washington DC Image (HYDICE)

In the next example, the benchmark Washington DC image from the HYDICE sensor contains 1208 scan lines with 307 pixels in each scan line and 224 bands, and it has a spatial resolution of approximately 2.8 m. A sub-image was produced for our experiments by spatially and spectrally subsetting to include  $200 \times 225$  pixels and 191 bands. Because the image also has a high spatial resolution, we obtained a ground truth image with six labeled classes by identifying the different materials. In the SVM classification algorithm, 5% of the labeled pixels for each class in the ground truth data were randomly chosen for training, and the remaining labeled pixels were used for testing. The optimal parameters for the SVM classifier were estimated as  $C = 2084$  and  $\gamma = 2$  by five-fold cross-validation. The GAs and CAs of the classification of the data set using the SVM classification are listed in Table 1. The parameters for the Proba-MS algorithm were fixed as follows:  $M = 20$ ,  $P = 5\%$  and  $T = 2\%$  for the Proba-M-MSF (+MV) method and  $M = 20$ ,  $P = 40\%$  and  $T = 50\%$  for the Proba-M-HSEG (+MV) method. Additionally, a  $3 \times 3$  structuring element was used in the Morph-M-HSEG (+MV) method.

In our method, 12 coarse grid levels were constructed in the AMG structure, and the optimal coarse grid level  $l = 5$  can be obtained from Figure 8 using Equation (15). In this grid, 1125 vertices, which occupy 2.5% of the total number of pixels in the image, were utilized as markers. To objectively compare the classification results, the GAs and CAs of the Washington DC image are shown in Table 3, and all the corresponding classification maps are illustrated in Figure 9. As can be observed from

these results, we can obtain similar conclusions as for the Indian Pines image. In particular, it can be observed from Table 3 that the GAs can be better than the pixel-wise SVM by most of the spectral-spatial classification methods used here, except for the Proba-M-MSF method, which could not effectively differentiate the *Street*, *Roofs* and *Path* classes. Furthermore, the best GAs among all the classification methods were obtained by our method with the SAM distance. In this case, the OA and  $\kappa$  were better by 3.84% and 4.83%, respectively, compared with the SVM results. In addition, the highest CAs for 3 of the 6 classes were achieved by our method. In this example, the increase in AA values was as large as 3.58%.

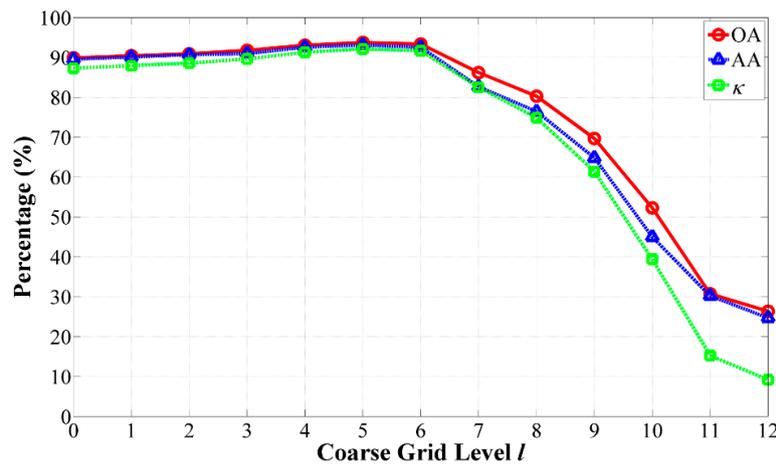
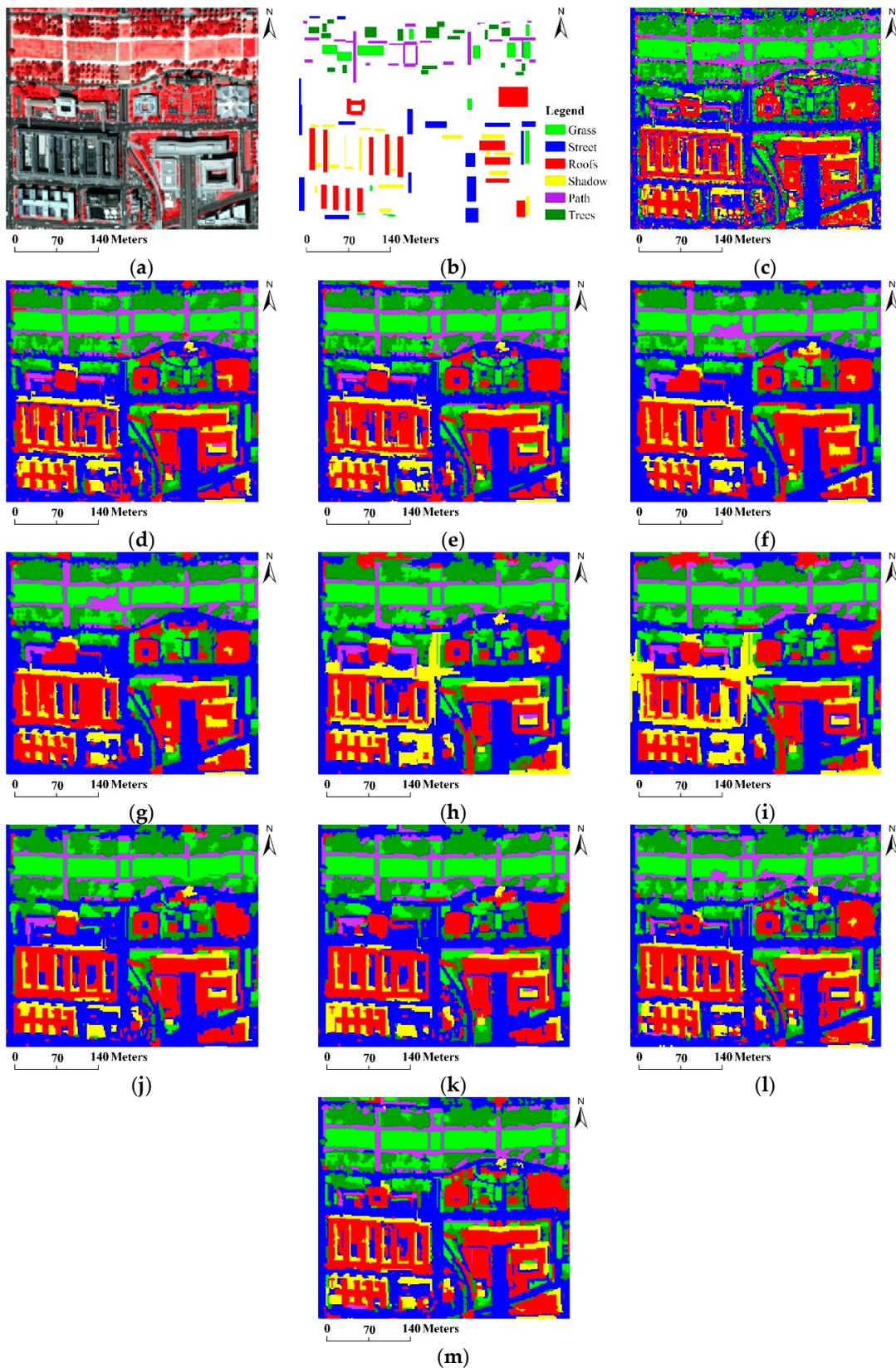


Figure 8. Sensitivity analysis of the GAs achieved by the AMG-M-HSEG method using the Washington DC image for  $l$  from 0 to 12.

Table 3. The GAs and CAs (percent) for the Washington DC image using all the spectral-spatial classification methods for comparison. The highest accuracies are indicated in bold in each category.

	AMG-M-HSEG		Proba-M-HSEG	Proba-M-HSEG + MV	Morph-M-HSEG	Morph-M-HSEG + MV	Proba-M-MSF		Proba-M-MSF + MV	
DC	ED	SAM	SAM	SAM	SAM	SAM	ED	SAM	ED	SAM
OA	92.80	<b>93.76</b>	90.36	91.79	92.91	92.93	85.89	84.18	91.08	92.69
AA	92.80	<b>93.12</b>	91.05	91.75	92.48	92.49	87.41	86.44	90.45	91.75
$\kappa$	90.95	<b>92.14</b>	87.92	89.69	91.08	91.10	82.42	80.39	88.79	90.80
Grass	94.20	94.5	94.90	95.20	95.30	95.30	96.10	95.30	<b>96.7</b>	95.30
Street	97.88	<b>97.88</b>	97.33	97.33	<b>98.29</b>	<b>98.29</b>	86.67	82.09	98.22	98.22
Roofs	91.73	<b>93.40</b>	84.68	88.86	91.33	91.37	79.94	77.07	88.90	92.26
Shadow	81.78	83.40	84.48	82.32	84.08	84.08	<b>89.2</b>	86.77	80.43	77.73
Path	88.30	<b>90.39</b>	86.58	88.42	87.19	87.19	75.62	87.68	78.69	87.59
Trees	<b>99.52</b>	99.17	98.33	98.33	98.69	98.69	96.9	89.76	99.76	99.40

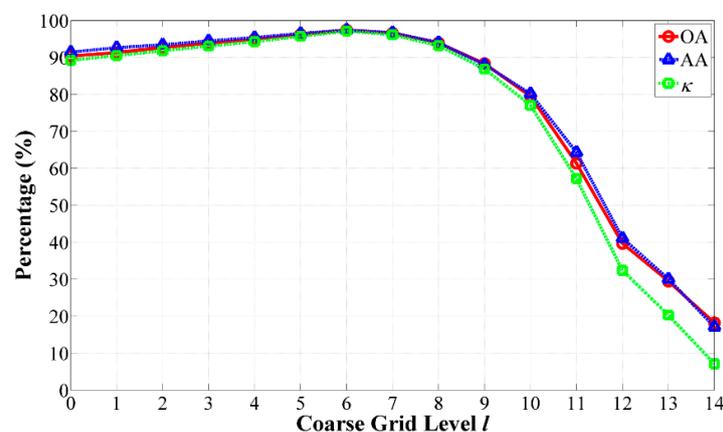


**Figure 9.** Classification maps for the Washington DC data set: (a) RGB colour composite (bands 63, 52, and 36); (b) Ground truth data; (c) SVM; (d) Proba-M-HSEG; (e) Proba-M-HSEG +MV; (f) Morph-M-HSEG; (g) Morph-M-HSEG +MV; (h) Proba-M-MSF using ED; (i) Proba-M-MSF using SAM; (j) Proba-M-MSF +MV using ED; (k) Proba-M-MSF +MV using SAM; (l) AMG-M-HSEG using ED; and (m) AMG-M-HSEG using SAM.

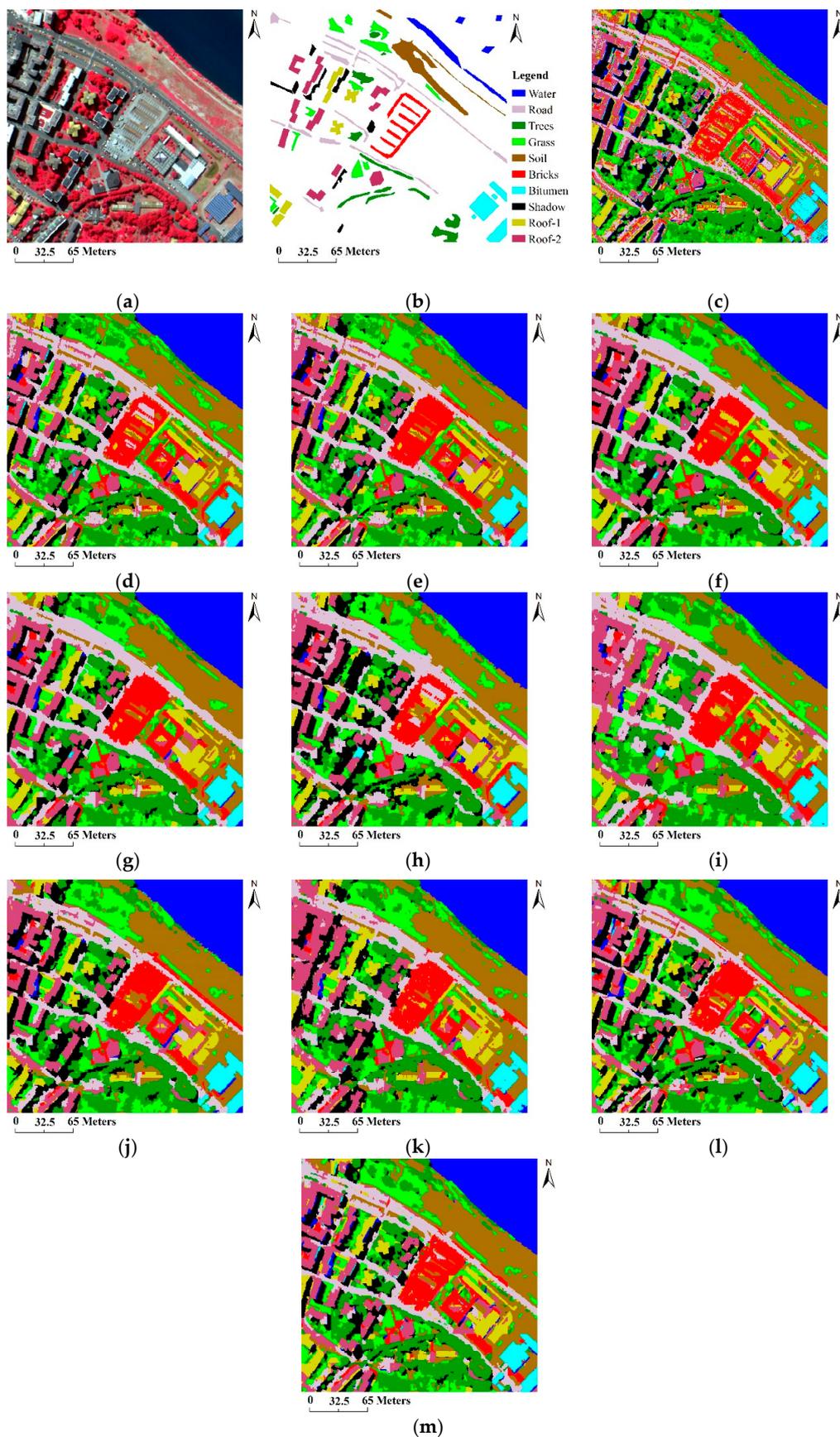
### 3.4. The Centre of Pavia Image (ROSI)

The third hyperspectral data set was the Centre of Pavia image, which was acquired by the ROSIS-03 optical sensor. The image has  $400 \times 400$  pixels, 102 spectral channels and a spatial resolution of 1.3 m. To evaluate our method, we manually generated a ground truth data for the image by visual interpretation that included ten material classes of interest, and 2% of the labeled pixels of each class from the ground truth data were selected as training samples. The remaining ones were used for testing. A pixel-wise SVM classification was performed on the image, and the following parameters were chosen by five-fold cross-validation:  $C = 1.31072 \times 10^5$  and  $\gamma = 2$ . The training and test samples for each class and the corresponding classification accuracies by the SVM classifier are reported in Table 1. The Proba-MS algorithm parameters for the Proba-M-MSF (+MV), Proba-M-HSEG (+MV) and Morph-M-HSEG (+MV) methods were the same as the second hyperspectral data set.

To obtain accurate markers for the following segmentation, 14 coarse grid levels of the Centre of Pavia image were constructed for building the AMG structure, in which 3105 vertices, which occupied 1.9% of the total number of pixels in the image for the optimal grid level of  $l = 6$ , were utilized as markers, as shown in Figure 10. For comparison, we show the GAs and CAs of the applied classification methods used here in Table 4, and the corresponding classification maps are displayed in Figure 11. It can be observed that almost all the spectral-spatial classification methods achieved higher GAs when compared with the pixel-wise SVM classification, except for the Proba-M-MSF method with the SAM distance. As shown in Figure 11i, most of the building shadows were not appropriately recognized by the Proba-M-MSF method with the SAM distance, and the CAs of this method confirm that conclusion. For example, the *Shadow* CA was only 27.15%, which was much lower than the 95.94% achieved by the SVM classifier. In addition, the best GAs were achieved using our method. For example, the increases in OA and  $\kappa$  were as large as 7.06% and 7.91%, respectively, compared with the SVM results. Apart from that, the highest CAs for 5 of the 10 classes were achieved by our method, and the improvement in AA was as large as 5.96%.



**Figure 10.** Sensitivity analysis of the GAs achieved by the AMG-M-HSEG method using the center of the Pavia image for  $l$  from 0 to 14.



**Figure 11.** Classification maps for the Centre of Pavia data set: (a) CIR colour composite (bands 63, 52, and 36); (b) Ground truth data; (c) SVM; (d) Proba-M-HSEG; (e) Proba-M-HSEG +MV; (f) Morph-M-HSEG; (g) Morph-M-HSEG +MV; (h) Proba-M-MSF using ED; (i) Proba-M-MSF using SAM; (j) Proba-M-MSF +MV using ED; (k) Proba-M-MSF +MV using SAM; (l) AMG-M-HSEG using ED; and (m) AMG-M-HSEG using SAM.

**Table 4.** The GAs and CAs (percent) for the center of the Pavia image using all the spectral-spatial classification methods for comparison. The highest accuracies are indicated in bold in each category.

	AMG-M-HSEG		Proba-M-HSEG	Proba-M-HSEG + MV	Morph-M-HSEG	Morph-M-HSEG + MV	Proba-M-MSF		Proba-M-MSF + MV	
DC	ED	SAM	SAM	SAM	SAM	SAM	ED	SAM	ED	SAM
OA	96.06	<b>97.45</b>	95.32	95.64	95.32	95.64	90.98	87.14	95.29	92.74
AA	96.58	<b>97.43</b>	95.42	95.92	95.42	95.92	90.38	83.88	95.87	92.63
$\kappa$	95.58	<b>97.14</b>	94.74	95.1	94.74	95.1	89.87	85.49	94.2	91.86
Water	99.85	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	99.18	99.85
Road	95.79	<b>98.36</b>	93.37	92.67	93.37	92.67	97.43	91.94	91.68	87.15
Trees	93	<b>98.12</b>	93.83	89.35	93.83	89.35	74.15	84.61	85.79	89.57
Grass	98.5	96.8	85.41	93.23	85.41	93.23	<b>99.14</b>	78.9	98.91	89.51
Soil	99.97	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	99.94	<b>100</b>	99.79	<b>100</b>
Bricks	97.53	96.88	97.23	99.5	97.23	<b>99.5</b>	81.85	94.71	95.21	98.44
Bitumen	95.95	93.27	96.61	96.28	96.61	96.28	90.62	<b>96.86</b>	95.46	94.96
Shadow	99.07	96.14	92.88	93.15	92.88	93.15	99.07	27.15	98.27	74.52
Roof-1	96.32	<b>99.29</b>	98.11	98.11	98.11	98.11	67.79	70.04	99.08	99.29
Roof-2	89.83	95.44	96.75	<b>96.91</b>	96.75	<b>96.91</b>	93.82	94.63	95.34	93

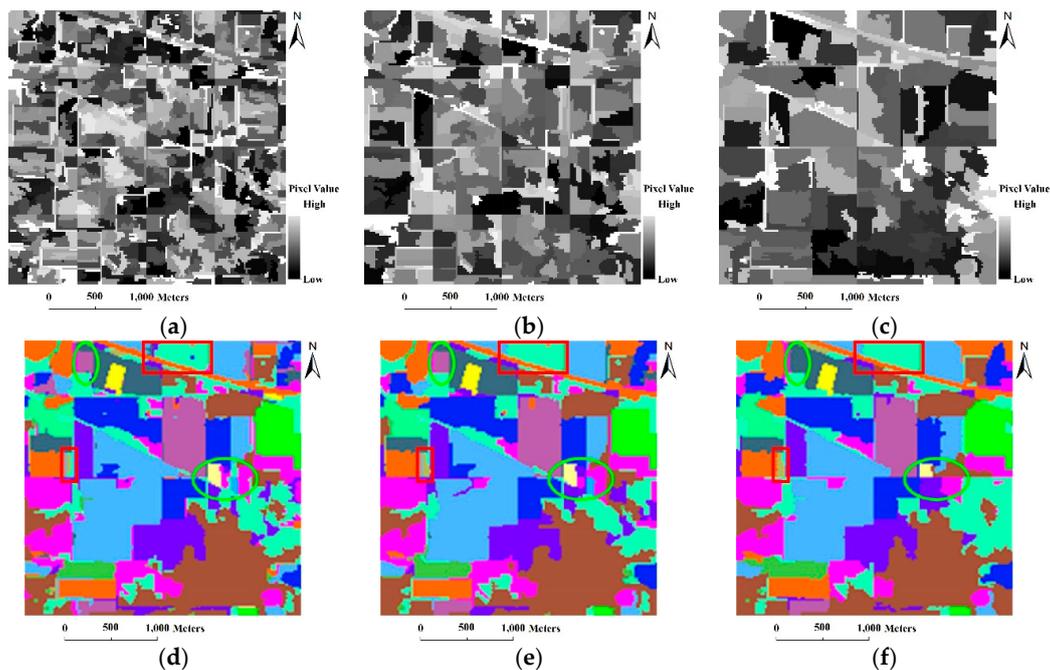
#### 4. Discussion

In our method, the coarse grid level  $l$  critically determines the AMG-M-HSEG classifier. In this section, the influence of this parameter on the classification accuracies of our method is investigated experimentally for the Indian Pines image. To clearly demonstrate the difference in the visual interpretation, the segmentation and corresponding classification maps with respect to  $l$  from 4 to 6 are shown in Figure 12. The GAs and CAs are also listed in Table 5. From these results, we can draw the following conclusions.

- (1) At the coarse grid level of  $l = 5$ , our algorithm achieved the best GAs.
- (2) As the coarse grid level  $l$  increased, the number of regions in the HSEG maps was reduced mainly due to the decrease of the selected markers. In addition, the CAs of classes with large connected regions were improved, such as the *Bldg-Grass-Tree-Drives* and *Oats* classes (refer to the red solid line rectangles in Figure 12). On the other hand, the CAs of the classes with small connected regions were greatly reduced, such as the *Grass/pasture-mowed*, *Soybeans-no till* and *Alfalfa* classes (refer to the green solid line ellipses in Figure 12). It is very interesting that different values of  $l$  produced homogeneous regions with different sizes. From this point of view, the adjustability of the grid level in the multigrid structure can provide flexibility for generating different segmentation maps to satisfy various applications.

**Table 5.** The GAs and CAs (percent) produced by our method using different values of  $l$  from 4 to 6 for the Indian Pines image. The highest accuracies are indicated in bold in each category.

Class	$l = 4$	$l = 5$	$l = 6$
OA	95.35	96.32	95.81
AA	94.59	95.97	88.99
$\kappa$	94.70	95.80	95.22
Alfalfa	<b>86.36</b>	<b>86.36</b>	70.45
Corn-no till	87.14	<b>92.33</b>	91.94
Corn-min till	93.74	96.67	<b>96.94</b>
Corn	96.21	<b>99.33</b>	97.63
Grass/pasture	<b>98.21</b>	96.43	93.08
Grass/trees	<b>97.77</b>	96.43	97.62
Grass/pasture-mowed	<b>93.75</b>	<b>93.75</b>	0
Hay-windrowed	<b>99.77</b>	93.42	<b>99.77</b>
Oats	80.00	<b>100</b>	<b>100</b>
Soybeans-no till	91.17	<b>91.97</b>	85.44
Soybeans-min till	96.98	98.51	<b>99.01</b>
Soybeans-clean till	<b>97.65</b>	95.30	96.93
Wheat	<b>100</b>	98.95	99.48
Woods	98.80	<b>98.88</b>	96.93
Bldg-Grass-Tree-Drives	98.25	98.83	<b>99.48</b>
Stone-steel towers	97.65	97.65	<b>99.40</b>



**Figure 12.** The segmentation and classification maps by our method with respect to different values of  $l$ . (a) the segmentation map ( $l = 4$ ); (b) the classification map ( $l = 4$ ); (c) the segmentation map ( $l = 5$ ); (d) the classification map ( $l = 5$ ); (e) the segmentation map ( $l = 6$ ); (f) the classification map ( $l = 6$ ).

## 5. Conclusions

In this paper, we proposed a spectral-spatial classification method for hyperspectral imagery based on the AMG method and the HSEG algorithm. Specifically, a multiscale representation for hyperspectral imagery is obtained using the AMG method to solve the hyperspectral anisotropic diffusion PDE. Once the multigrid structure is constructed, vertices in any grid level can be considered as markers for HSEG and clustered to obtain a sequence of unsupervised segmentation results. A maximum vote decision rule is then employed to classify all the homogeneous regions in the segmentation maps according to a pixel-wise classification result. Finally, to obtain the best classification accuracies, a final classification map is developed by choosing the optimal grid level to extract representative spectra. The major advantage of the proposed method is that it can obtain reliable markers that are only determined by the structures in the hyperspectral imagery, not the performance of the pixel-wise classifiers or the training samples in the pixel-wise classification step. The proposed classification method, *i.e.*, AMG-M-HSEG, was compared with the pixel-wise SVM classifier and marker-based spectral-spatial classification methods. The results of the experiments demonstrated that our method can achieve improvements of 13.81%, 3.84% and 7.06% in OA over the pixel-wise SVM classifier for the Indian Pines, Washington DC and Centre of Pavia datasets, respectively. Furthermore, for all three datasets, the OA and  $\kappa$  by our method were the best among all the traditional spectral-spatial classification methods. It can be concluded that the proposed method can effectively improve classification accuracy for real hyperspectral datasets with different resolutions and contexts. Therefore, our method is useful in real-world applications. In the future, a further improvement can be achieved by exploring more valuable information from the AMG-based scale-space of hyperspectral imagery.

**Acknowledgments:** This work was supported by the National Natural Science Foundation of China (61271408 and 61372153); the Natural Science Foundation of Hubei Province, China (2014CFA052); and the Fundamental Research Funds for the Central Universities, China University of Geosciences (CUGL140410). The authors would 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:** Haiwei Song and Yi Wang implemented all the classification methods in the C++ programming language, conducted the experiments and finished the first draft. Yi Wang supervised the research and contributed to the editing and review of the manuscript.

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

## References

1. Landgrebe, D.A. Multispectral land sensing: Where from, where to? *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 414–421. [[CrossRef](#)]
2. Ham, J.; Chen, Y.; Crawford, M.M.; Ghosh, J. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 492–501. [[CrossRef](#)]
3. Samat, A.; Du, P.; Liu, S.; Cheng, L. E<sup>2</sup>lms: Ensemble extreme learning machines for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2014**, *7*, 1060–1069. [[CrossRef](#)]
4. Bazi, Y.; Alajlan, N.; Melgani, F.; AlHichri, H.; Malek, S.; Yager, R.R. Differential evolution extreme learning machine for the classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1066–1070. [[CrossRef](#)]
5. Mianji, F.A.; Zhang, Y. Robust hyperspectral classification using relevance vector machine. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2100–2112. [[CrossRef](#)]
6. Bazi, Y.; Melgani, F. Gaussian process approach to remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 186–197. [[CrossRef](#)]
7. Sun, S.; Zhong, P.; Xiao, H.; Wang, R. Active learning with gaussian process classifier for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1746–1760. [[CrossRef](#)]
8. Camps-Valls, G.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1351–1362. [[CrossRef](#)]
9. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1780. [[CrossRef](#)]
10. Ratle, F.; Camps-Valls, G.; Weston, J. Semisupervised neural networks for efficient hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2271–2282. [[CrossRef](#)]
11. 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. [[CrossRef](#)]
12. Yuan, H.; Tang, Y.Y.; Lu, Y.; Yang, L.; Luo, H. Hyperspectral image classification based on regularized sparse representation. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2014**, *7*, 2174–2182. [[CrossRef](#)]
13. Ma, L.; Crawford, M.M.; Yang, X.; Guo, Y. Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2832–2844. [[CrossRef](#)]
14. Huang, X.; Zhang, L. A comparative study of spatial approaches for urban mapping using hyperspectral rosis images over pavia city, northern Italy. *Int. J. Remote Sens.* **2009**, *30*, 3205–3221. [[CrossRef](#)]
15. 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. [[CrossRef](#)]
16. Falco, N.; Benediktsson, J.A.; Bruzzone, L. Spectral and spatial classification of hyperspectral images based on ica and reduced morphological attribute profiles. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6223–6240. [[CrossRef](#)]
17. Mura, M.D.; Benediktsson, J.A.; Waske, B.; Bruzzone, L. Morphological attribute profiles for the analysis of very high resolution images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3747–3762. [[CrossRef](#)]
18. Shen, L.; Zhu, Z.; Jia, S.; Zhu, J.; Sun, Y. Discriminative gabor feature selection for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 29–33. [[CrossRef](#)]
19. Chen, C.; Li, W.; Su, H.; Liu, K. Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine. *Remote Sens.* **2014**, *6*, 5795–5814. [[CrossRef](#)]
20. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 809–823. [[CrossRef](#)]

21. Xia, J.; Chanussot, J.; Du, P.; He, X. Spectral-spatial classification for hyperspectral data using rotation forests with local feature extraction and markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2532–2546. [[CrossRef](#)]
22. 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. [[CrossRef](#)]
23. Li, J.; Marpu, P.R.; Plaza, A.; Bioucas-Dias, J.; Benediktsson, J.A. Generalized composite kernel framework for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4816–4829. [[CrossRef](#)]
24. Tarabalka, Y.; Tilton, J.C.; Benediktsson, J.A.; Chanussot, J. A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2012**, *5*, 262–272. [[CrossRef](#)]
25. Tarabalka, Y.; Chanussot, J.; Benediktsson, J.A. Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers. *IEEE Trans. Syst. Man Cybern. B Cybernet.* **2010**, *40*, 1267–1279. [[CrossRef](#)] [[PubMed](#)]
26. Kimmel, R.; Yavneh, I. An algebraic multigrid approach for image analysis. *SIAM J. Sci. Comput.* **2003**, *24*, 1218–1231. [[CrossRef](#)]
27. Perona, P.; Malik, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 629–639. [[CrossRef](#)]
28. Black, M.J.; Sapiro, G.; Marimont, D.H.; Heeger, D. Robust anisotropic diffusion. *IEEE Trans. Image Process.* **1998**, *7*, 421–432. [[CrossRef](#)] [[PubMed](#)]
29. Mrázek, P.; Navara, M. Selection of optimal stopping time for nonlinear diffusion filtering. *Int. J. Comput. Vis.* **2003**, *52*, 189–203. [[CrossRef](#)]
30. Voci, F.; Eiho, S.; Sugimoto, N.; Sekibuchi, H. Estimating the gradient in the perona-malik equation. *IEEE Signal Process. Mag.* **2004**, *21*, 39–65. [[CrossRef](#)]
31. Weickert, J.; Romeny, B.M.T.H.; Viergever, M.A. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Trans. Geosci. Remote Sens.* **1998**, *7*, 398–410. [[CrossRef](#)] [[PubMed](#)]
32. Duarte-Carvajalino, J.M.; Castillo, P.E.; Velez-Reyes, M. Comparative study of semi-implicit schemes for nonlinear diffusion in hyperspectral imagery. *IEEE Trans. Image Process.* **2007**, *16*, 1303–1314. [[CrossRef](#)] [[PubMed](#)]
33. Duarte-Carvajalino, J.M.; Sapiro, G.; Vélez-Reyes, M.; Castillo, P.E. Multiscale representation and segmentation of hyperspectral imagery using geometric partial differential equations and algebraic multigrid methods. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2418–2434. [[CrossRef](#)]
34. Briggs, W.L.; Henson, V.E.; McCormick, S.F. *A Multigrid Tutorial*, 2nd ed.; SIAM: Philadelphia, PA, USA, 2000; pp. 7–48.
35. Tilton, J.C. RHSEG User's Manual: Including HSWO, HSEG, HSEGExtract, HSEGReader and HSEGViewer, Version 1.50. Available online: <http://opensource.gsfc.nasa.gov/projects/HSEG/RHSEG%20User's%20Manual%20V1.47.pdf> (accessed on 24 March 2016).
36. Mathieu, F.; Jocelyn, C.; Atli, B.J. A spatial-spectral kernel-based approach for the classification of remote-sensing images. *Pattern Recognit.* **2012**, *45*, 381–392.
37. Vapnik, V.N. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998.
38. Chang, C.-C.; Lin, C.-J. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27. [[CrossRef](#)]
39. Hsu, C.-W.; Chang, C.-C.; Lin, C.-J. *A Practical Guide to Support Vector Classification*; Technical Report; Department of Computer Science, National Taiwan University: Taipei, Taiwan, 2003; pp. 1–12.
40. Richards, J.A.; Jia, X. *Remote Sensing Digital Image Analysis: An Introduction*; Springer-Verlag: New York, NY, USA, 1999.

