HAL
open science

Spectral-Spatial Classification of Hyperspectral Data based on a Stochastic Minimum Spanning Forest Approach

Kevin Bernard, Yuliya Tarabalka, Jesus Angulo, Jocelyn Chanussot, Jon Atli Benediktsson

To cite this version:

Kevin Bernard, Yuliya Tarabalka, Jesus Angulo, Jocelyn Chanussot, Jon Atli Benediktsson. Spectral-Spatial Classification of Hyperspectral Data based on a Stochastic Minimum Spanning Forest Approach. IEEE Transactions on Image Processing, 2012, 21 (4), pp.2008-2021. 10.1109/TIP.2011.2175741 . hal-00728498

HAL Id: hal-00728498

https://inria.hal.science/hal-00728498v1

Submitted on 6 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spectral–Spatial Classification of Hyperspectral Data Based on a Stochastic Minimum Spanning Forest Approach

Kévin Bernard, Yuliya Tarabalka, *Member, IEEE*, Jesús Angulo, Jocelyn Chanussot, *Fellow, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—In this paper, a new method for supervised hyperspectral data classification is proposed. In particular, the notion of stochastic minimum spanning forest (MSF) is introduced. For a given hyperspectral image, a pixelwise classification is first performed. From this classification map, M marker maps are generated by randomly selecting pixels and labeling them as markers for the construction of MSFs. The next step consists in building an MSF from each of the M marker maps. Finally, all the M realizations are aggregated with a maximum vote decision rule in order to build the final classification map. The proposed method is tested on three different data sets of hyperspectral airborne images with different resolutions and contexts. The influences of the number of markers and of the number of realizations M on the results are investigated in experiments. The performance of the proposed method is compared to several classification techniques (both pixelwise and spectral–spatial) using standard quantitative criteria and visual qualitative evaluation.**

***Index Terms*—Classification, hyperspectral image, marker selection, minimum spanning forest (MSF), multiple classifiers, stochastic.**

I. Introduction

HYPERSPECTRAL imaging can be defined as the simultaneous acquisition of an image in many narrow contiguous spectral bands [1]. The advantage of this technique is that, considering that every element (water, tree, soil, etc) is defined by a precise spectrum (*spectral signature*), it should be possible to accurately classify every pixel of the image by considering their spectrum. The supplementary information brought by the spectrum of the pixels should help in the identification of objects in a scene acquired by a hyperspectral imaging sensor (i.e., hyperspectral image classification).

The high dimensionality of these data sets, implied by the large number of spectral channels, presents challenges to image analysis. A lot of methods for supervised classification of hyperspectral images have been proposed in the last decades. *Pixelwise classifiers* [2], [3], for instance, are considering only the spectral information of the pixels. Some of the most popular classifiers based on the spectral information alone are the maximum likelihood (ML) or Bayesian estimation technique [2], [4]–[6], neural networks [7]–[11], decision trees [12], [13], genetic algorithms [14], and kernel-based methods [3], [15]–[17]. More recently, support vector machine (SVM) classifier has been introduced [18]–[20]. The SVM method attempts to separate training samples belonging to different classes by tracing maximum margin hyperplanes in the space where the samples are mapped. This technique has been proven to be well suited to classify hyperspectral data, particularly when the available number of training samples is limited [19], [21].

However, when using these classifiers, the spatial information contained in the image is not taken into account, whereas in real images, adjacent pixels are actually related or correlated. Therefore, *spectral–spatial* or *context* classifiers, which will assign each pixel to one class regarding both its spatial and spectral information, must be developed to improve classification results [16]–[22]. David Landgrebe and his research team at Purdue University were pioneers in introducing spatial context into a multiband image classification with the Extraction and Classification of Homogeneous Objects (ECHO) classifier [2], [23]. Since then, many studies have been led to propose new algorithms for performing spectral–spatial classification. For instance, one approach consists in including the information from the closest neighborhood of a pixel. Such an approach, which can use morphological filtering [16], morphological leveling [24], [25], or Markov random fields [26], has been proven to be much more accurate than the pixelwise methods to classify hyperspectral images. Another recent approach to perform a spectral–spatial classification is based on a morphological scale–space associated to the hyperspectral image, which is then explored using tensor modeling and classification [27]. Nevertheless, the major problem of this approach is the scale selection, particularly if small or complex structures are present

Manuscript received April 06, 2011; revised August 11, 2011; accepted October 21, 2011. Date of publication November 11, 2011; date of current version March 21, 2012. This research was supported in part by the Research Fund of the University of Iceland. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yongyi Yang.

K. Bernard is with the University of Iceland, 101 Reykjavik, Iceland, and also with Heriot Watt University, Edinburgh EH14 4AS, U.K.

Y. Tarabalka is with the Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, MD 20771 USA.

J. Angulo is with the Center of Mathematical Morphology, Department of Mathematics and Systems, École des Mines de Paris (Mines ParisTech), 77305 Fontainebleau, France.

J. Chanussot is with the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble Institute of Technology, 38402 Saint Martin d'Hères, France.

J. A. Benediktsson is with the University of Iceland, 101 Reykjavik, Iceland.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2011.2175741

1057-7149/$26.00 © 2011 IEEE

in the image. Another approach to include spatial information in classification consists in doing a segmentation of the image. *Segmentation* of a given image can be defined as its exhaustive partitioning into homogeneous regions (the homogeneity criterion can be based, for instance, on the intensity or on the texture) [28].

The subject of image segmentation includes many different techniques and algorithms. In particular, graph-based segmentation algorithms have become quite popular and mature due to their excellent performance (for a complete recent review, see [29]). Examples of these graph-based algorithms are watershed [30], [31], minimum spanning forest (MSF) [32]–[34], graph cuts [35], random walker [36], and shortest paths [37]. In fact, as shown by Meyer in [38], watershed segmentation (hierarchies and segmentation with markers) and MSF segmentation are closely linked. This point will be developed below in Section II-A. Other recent works [39], [40] have also studied the links between both methodologies. A recent effort has been also reported in [29] to unify different graph-based optimization algorithms via the notion of power watershed. In previous works, some of these graph-based algorithms have been applied to the segmentation of hyperspectral images; for instance, Tarabalka *et al.* used watershed [41], partitional clustering [42], hierarchical segmentation [43], and MSF [33] techniques to perform unsupervised segmentation of hyperspectral images. The best classification accuracy values were obtained for the MSF grown from automatically selected markers. Markers are the pixels that are chosen to initiate, in the case of the MSF, the growing procedure. They are a collection of connected components with labels, each marker being assigned a distinct label (a marker can be split into several connected components with the same label).

The algorithm used by Tarabalka *et al.* [33] is depicted in Fig. 1. As the MSF segmentation is a marker-controlled region growing method (such as watershed segmentation), the choice of the seeds to initiate the growing procedure of the forest is crucial. In this case, a pixelwise classification of the hyperspectral data set is performed. At the same time, a probabilistic map is computed. By combining the information from the latter and from a connected component analysis of the classification map, only the most reliable pixels are kept as markers. From this map of markers, an MSF is built. For a more detailed description of the method, the reader can refer to [33] where this method provides very good results in terms of classification accuracy values. However, if the connected component analysis is easy to perform, generation of a probabilistic map is more demanding in terms of algorithmic complexity. Thus, we propose here a new way to select the markers that provide similar or better results with a lower computational cost.

This new proposed method was inspired by the recently proposed stochastic watershed approach in [44]. The classical paradigm of watershed segmentation lies on the appropriate choice of markers, which are the seeds to initiate the flooding procedure. In the stochastic watershed approach, an opposite direction is followed by selecting random germs for markers on the watershed segmentation. This arbitrary choice will be balanced by the use of a given number M of realizations in order to filter out the nonsignificant fluctuations. The original framework was then extended in [45] for unsupervised hyperspectral image segmentation. More recently, the approach has been adapted to a semisupervised framework in hyperspectral segmentation [46], i.e., by using a training data set for each spectral class, the random markers are generated according to a family of density functions that depends on each spectral class.

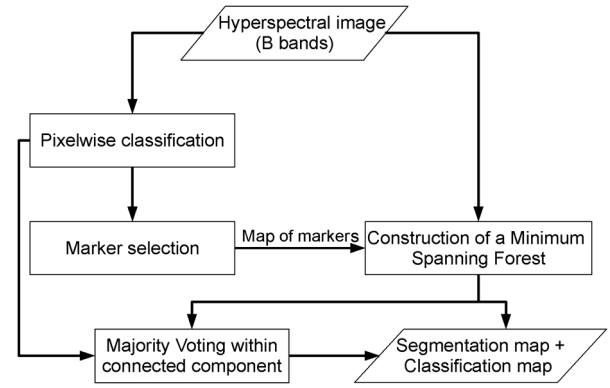

Fig. 1. Flowchart of the spectral–spatial classification approach using an MSF grown from automatically selected markers proposed in [33].

In this paper, we propose exporting this concept of stochasticity to the more general case of MSF segmentation. In particular, considering the case of hyperspectral segmentation and starting from a supervised classification of the spectral space, an original stochastic minimum spanning tree (MST) mechanism of simulation by generating random nodes from the classification and by a majority voting rule is introduced.

This paper is organized as follows. In the next section, the proposed spectral–spatial classification method based on MSF rooted on randomly selected markers is presented. Experimental results are depicted in Section III. Finally, conclusions are drawn in Section IV.

II. Stochastic MSF Approach

The aim of this section is to introduce the algorithm developed for hyperspectral image classification using the present stochastic MSF approach. It is, nevertheless, important to technically precise the relationships between this original approach and the intimately related watershed and MSF segmentations.

A. Connections Between Marker-Driven Watershed and MSF Segmentations

As already aforementioned, the underlying idea of the stochastic MSF is to incorporate the segmentation robustness shown by the stochastic watershed in the powerful framework of the graph-based MSF segmentation. In fact, this idea is a natural generalization since both segmentation approaches are strongly interwoven, as previously shown in the literature [40], [47].

Watershed segmentation is a well-known morphological methodology for image segmentation. The classical flooding algorithm on a gray-scale image works as follows [30]. During successive flooding of the gray value relief, watersheds with adjacent catchment basins are constructed. This flooding process is performed on the gradient image, i.e., the basins should emerge along the edges. Each local minimum in the image will be given a different label and will be then referred to as "marker." It is from these markers that the flooding

procedure will be initiated. The neighboring pixels of each marked area are inserted into a priority queue with a priority level corresponding to the gray level of the pixel. The pixel with the highest priority level is extracted from the priority queue. If the neighbors of the extracted pixel that have been already labeled all have the same label, then the pixel is labeled with their label. All nonmarked neighbors that are not yet in the priority queue are put into the priority queue. The labeling procedure is repeated until the priority queue is empty. The nonlabeled pixels are the watershed lines. Normally, considering all the minima will lead to an oversegmentation of the image. Marker-based watershed transformation makes use of specific marker positions as seeds for the flooding procedure.

Most efficient algorithms of marker-based watershed segmentation are implemented using graph representations. Let us review some of the basic results introduced in [47], which will be the basic ingredients for an efficient implementation of our stochastic MSF algorithm. The gradient image to be segmented is represented by a weighted neighborhood graph, whose nodes are the regional minima. Two minima are linked by an edge if the corresponding catchment basins are neighbors. The altitude of the lowest path found on the boundary between two catchment basins gives the valuation of the edge. Every time a catchment basin contains at most one marker or part of one marker, the corresponding node in the neighborhood graph will get the label of the marker and will be part of a graph marker. If a marker contains several connected nodes on the neighborhood graph, these may be replaced by a unique node with the same outside connections.

From this representation, constructing a watershed from a set of markers on a neighborhood graph yields a partition of the graph. It was proven in [30] that this partition is an MSF, with each tree rooted a marker. In fact, the MSF is associated to the notion of MST. Consider an undirected graph $G = (V, E, W)$, where V and E are the sets of (vertices) nodes and edges, respectively, and W is a mapping of the set of the edges E into $\mathbb{R}^+$; each edge $e \in E$ is associated to a positive number $w(e)$, which is the weight of the edge e. Given a partial graph $G' = (V, E', W')$ of G, the weight of G' is number $w(G') = \sum_{e \in E'} w(e)$. We remind that an MST is a tree (a connected graph without cycle, i.e., there exists a unique path between any pair of nodes) of minimum weight. It is unique if all weights are different. There exist several efficient algorithms for computing the MST [32], [48]. Let $M = \{m_i\}_{i \in I}$ be a set of markers. The construction of an MSF is generally based on adding a dummy node for each marker m_i of the graph and linking it to marker nodes with zero value. The MST of this new graph is constructed, and the dummy nodes and edges are suppressed, producing a forest, where each tree A_i is rooted in m_i. An example of construction of the MSF rooted on markers is depicted in Fig. 2. It can be concluded that, working on the neighborhood graph of catchment basins for a given set of image markers, the same segmentation is obtained with the watershed transform or with the MSF approach.

1) Efficient Computation of MSF From an MST: An MST of a neighborhood graph has far less edges than the graph taken as a whole. In addition, the MST can be taken as a starting point to speed up the construction of an MSF. Let $G = (V, E, W)$ be a graph and $\{A_i\}_{i \in I}$ be an MSF associated to the set of markers M. Consider now a new graph $G'' = (Y, E'', W'')$: the nodes

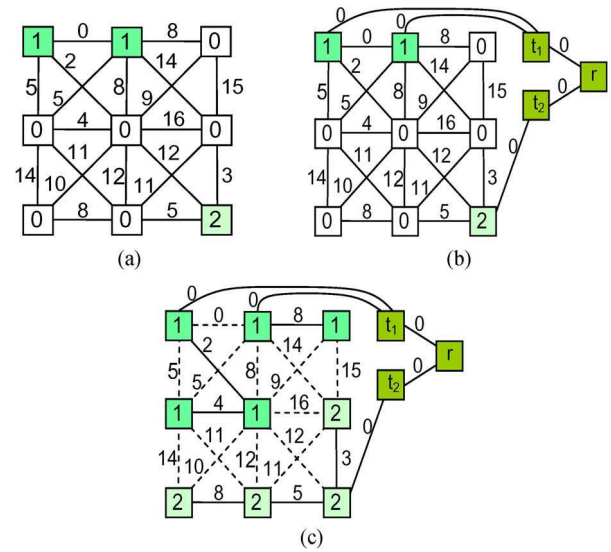

Fig. 2. Example of construction of an MSF rooted on markers. (a) Original image graph G, where colored vertices represent markers 1 and 2. Nonmarker pixels are denoted by "0." (b) Addition of extra nodes t_1, t_2, and r to the graph. (c) MST of the graph presented in (b). After removing node r, an MSF is obtained, where each tree grown from node t_i forms a region in the graph. The picture was taken from [33].

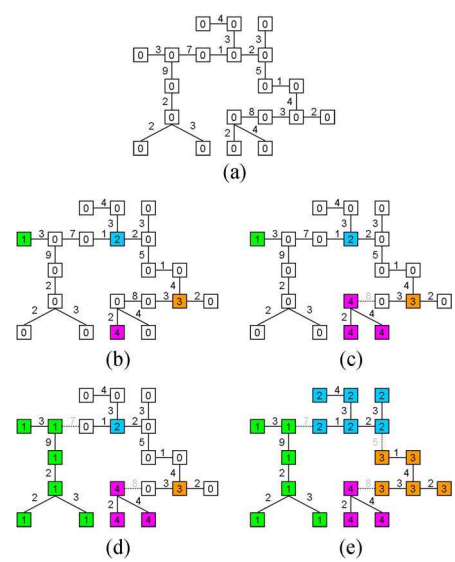

Fig. 3. Example of construction of an MSF rooted on markers from an MST. (a) Initial image MST T. (b) Four markers defined by the colored nodes. (c) and (d) Illustration of the construction of the MSF from the four markers by highest weight edge suppression (see the text). (e) Final MSF, where each tree, rooted in a marker, has the color of its marker.

represent the trees A_i of the forest, and the edges represent the edges of E having extremities in two different trees A_i and A_j of the forest. Let T' be an MST of graph G''. The union of the edges of the MSF $\{A_i\}_{i \in I}$ and of the MST T' constitutes an MST T of initial graph G. Consequently, the MST is an overset of any MSF associated with a set of markers; in addition, if the MST is available, the MSF can be efficiently computed. A simplification of the classical algorithm by Kruskal [32] was introduced by Meyer [30], [47] as the starting point for an efficient strategy of interactive segmentation (suppression and addition of markers). An example of the efficient computation of an MSF from an MST, borrowed from [47], is depicted in Fig. 3. Given

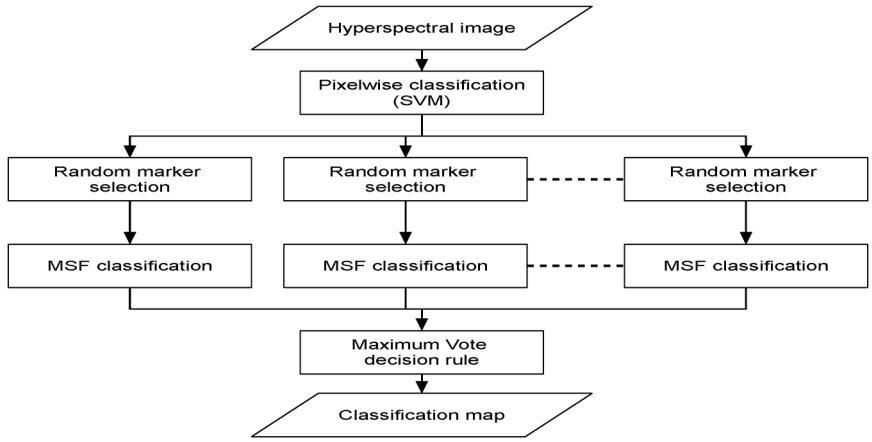

Fig. 4. Flowchart of the RD-MSF algorithm described in Section II-B.

the MST T of a graph $G = (V, E, W)$ and a set of markers $M = \{m_i\}_{i \in I}$, the edges of T are first ordered in the order of their decreasing weights and are considered one after the other. Suppose e is the edge currently under consideration. Edge e belongs to a subtree of the MST. Suppressing e will cut this tree into two smaller subtrees; if each of them contains at least one marker, then the suppression of e is validated (this is the case in Fig. 3(c)–(e) where an edge has been suppressed each time); if at least one of the subtrees does not contain a marker, then edge e is reintroduced (this is the case with the highest edge of the MST in Fig. 3(b), with a weight equal to 9, which cannot be suppressed since it does not separate two distinct markers). The process stops when each of the subtrees that has been created contains one and only one marker. The forest constructed like this obviously produces an MSF with one tree rooted in each marker.

B. Proposed Algorithm for Hyperspectral Classification

The introduced algorithm of stochastic MSF (RD-MSF), which is depicted in Fig. 4 and illustrated in Fig. 5, is composed of four main steps as follows:

1) *supervised pixelwise spectral classification*;
2) *random marker selection*;
3) *spatial–spectral classification of each realization using marker-controlled MSF*;
4) *maximum vote rule for class decision*.

1) Supervised Pixelwise Spectral Classification: The aim of this step is to first provide a supervised classification map from a hyperspectral image, from which we will create several realizations of marker maps. We notice that this initial classification is exclusively based on the spectral domain. This is done by performing an SVM classification using a training data set for each spectral class of the image. Basically, for a binary classification problem (B-dimensional space $\mathbb{R}^B$, with N training samples, $\mathbf{x}_i \in \mathbb{R}^B$, and their corresponding labels $y_i = \pm 1$, i.e., $\{(\mathbf{x}_i, y_i) | I \in [1, N]\}$ available), the SVM method consists in finding the hyperplane that maximizes the margin, i.e., the distance to the closest training data points in both classes. The parameters used to perform this classification will be detailed for each image later on. For more information about the SVM theory and its efficiency for hyperspectral image classification, we refer the reader to [18]–[21]. We chose to perform pixelwise classification by using an SVM classifier rather than another one (such as ML classifier) because of the effectiveness of the SVM classifier in terms of classification accuracy, computational complexity, and robustness to parameter setting. In addition, according to [49], the SVM classifier has been the most widely used to perform pixelwise classification of hyperspectral data in the last decade.

2) Random Marker Selection: In this step, we will produce several stochastic marker maps. Each map is built by randomly selecting some pixels and setting them as markers for the MSF-based image partition that follows. This step is very important as it is here that the number of markers nb_mark and the number M of realizations (also referred to as nb_map) will be chosen. Both parameters are the basic degrees of freedom of the stochastic algorithm, and their effects in practical examples will be considered in detail in the next section.

In this paper, the nb_mark markers are uniformly selected without any prior information. However, if the *a priori* probability of each spectral class, representing the likelihood for the appearance of the particular class in the image, is available, this information can be used to conditionally select a proportional number of uniform random markers for each class. On the other hand, the random markers can be also selected according to a regionalized density function that aims at favoring certain regions of the image, as it is done in [46].

3) Spatial–Spectral Classification of Each Realization Using Marker-Controlled MSF: For this step of the stochastic algorithm, we chose to use the MSF approach on pixel neighborhood graph $G = (V, E, W)$, where each edge $e_{i,j} \in E$ of the graph connects a couple of nodes i and j corresponding to the neighboring pixels; hence, each pixel k of the image is a node v_k of the graph. If an 8-neighborhood is used, every pixel is connected by an edge with each of its neighbors (in total, eight edges). Furthermore, a weight $w_{i,j}$ is assigned to each edge $e_{i,j}$, which indicates the degree of dissimilarity between two pixels connected by this edge. Consequently, the associated pixel graph MSF segmentation will not be the same segmentation obtained from a marker-controlled watershed. The choice of this pixel graph, instead of the catchment basin graph, was motivated by [33]. In this paper, it is shown that this pixel graph MSF gives significantly better results for hyperspectral image segmentation. Moreover, the simplicity of the algorithm involved and the speed of execution of an MSF classification are two important factors that outperform the watershed segmentation, and thus,

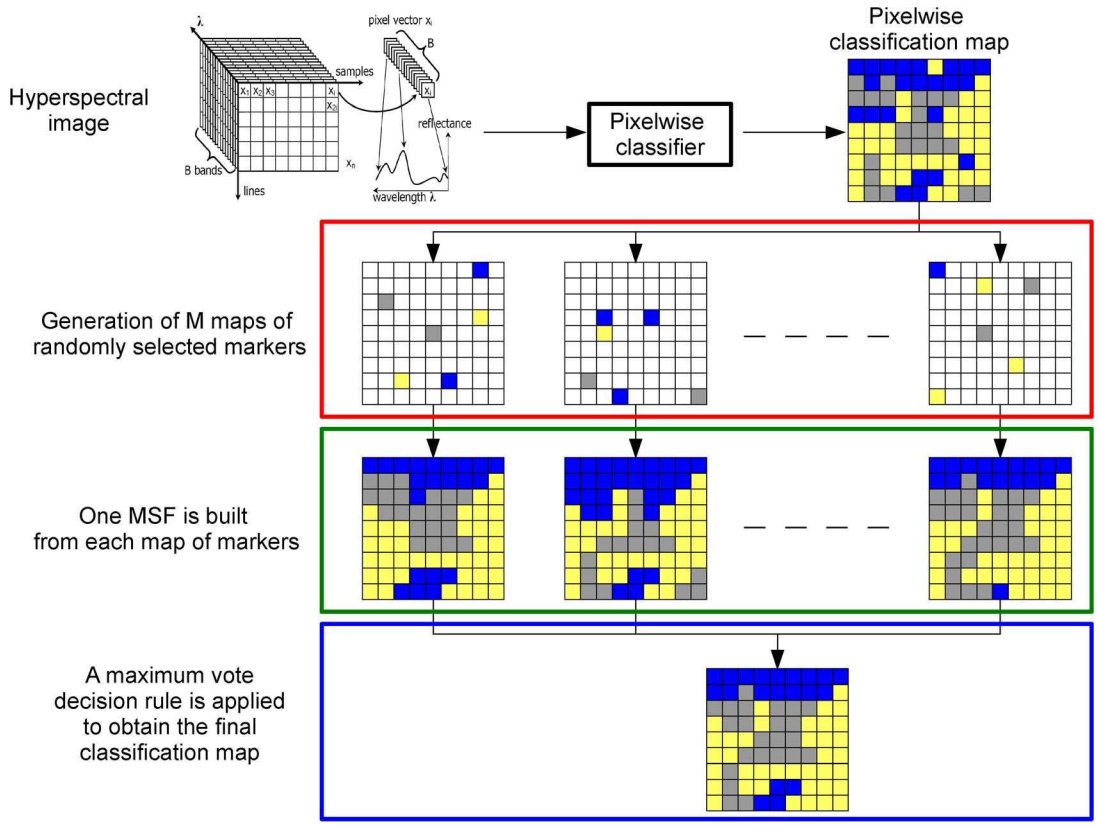

Fig. 5. Illustration of the stochastic MSF classification algorithm.

they are two strong arguments to choose the MSF segmentation algorithm, rather than the watershed approach. Generally speaking, the construction of an MSF using an arbitrary dissimilarity measure is a general, simple, flexible, and efficient region partition algorithm.

Due to the fact that, starting from the same image, i.e., from the same graph G, an MSF segmentation should be computed for each realization of random markers, a very efficient algorithm for the stochastic approach is based on the MST algorithm discussed above. That is, the MST T of the pixel neighborhood graph G is computed, which is the expensive part of the approach. Then, for each one of the M realizations of markers, an MSF is computed from T.

The only parameter to set in the construction of our pixel neighborhood graph-based MSF is the choice of the dissimilarity criterion. In the literature, we can find a lot of different dissimilarity measures [22], [50], [51]. Nevertheless, we chose the Spectral Angle Mapper (SAM) criterion that can be defined as follows. The SAM measure determines the spectral similarity between two vectors $\mathbf{x}_i = (x_{i1}, \ldots, x_{iB})^T$ and $\mathbf{x}_j = (x_{j1}, \ldots, x_{jB})^T$ by computing the angle between them. More precisely, it is defined as

$$\mathrm{SAM}(\mathbf{x}_i, \mathbf{x}_j) = \arccos\left(\frac{\sum_{b=1}^{B} x_{ib}x_{jb}}{\left[\sum_{b=1}^{B} x_{ib}^2\right]^{1/2}\left[\sum_{b=1}^{B} x_{jb}^2\right]^{1/2}}\right). \quad (1)$$

This distance has been proven to be very well suited to most of the cases in hyperspectral image classification [52]. However, one of the drawbacks is that this distance is not well suited to discriminate classes regarding the illumination of the scene. This is to say, if in an image we want to sort some pixel as class "shadows," we might have some problems. This is a parameter that we will discuss later in this paper when looking at the results for classification of urban areas.

4) Maximum Vote Rule for Class Decision: At this stage of the algorithm, we have M spatial–spectral classification maps. The value for the number of realizations M should be set by the user. In the following step, maximum vote, we will compare all the results of the different classifications previously obtained. This step is very important. It will allow us to get rid of the statistical fluctuations of the classification and, thus, improve both the efficiency and the robustness of the results. This approach is known in the literature as a multiclassifier approach. It means that, rather than keeping only the best realization, we will store all the classifications in order to compare them. It will allow us to keep the potentially relevant information contained in the less successful classifiers. The process is as follows. For a given pixel, its values obtained with the different classification map are put in a $1 \times M$ vector (M realizations). At the end, the pixel will be assigned to the class with the highest occurrence (in case of a tie, the final class of the pixel will be the class determined by the SVM classifier). The same process is applied to every pixel of the image. At the end, the final classification map is obtained. This classification function is named maximum vote.

III. Experimental Results

A. Evaluation of the Results

Here, we will present the results that we obtained with our proposed method on different sets of images. Three airborne hyperspectral images are used to test our method. They imply different contexts such as agricultural, volcanic, and urban areas. The efficiency of the method is evaluated in terms of the following two qualities:

1) mathematical efficiency: *classification accuracy measures* [*overall accuracy (OA)*, *class-specific accuracy (CA)*, and *average accuracy (AA)*];
2) visual accuracy: *visual comparison of the classification maps*.

Assuming the following definitions:

Overall Accuracy: The OA is the percentage of correctly classified pixels (K is the number of classes)

$$\mathrm{OA} = \frac{\sum_i^K C_{ii}}{\sum_{ij}^K C_{ij}} \times 100\% \tag{2}$$

where C_{ij} is the number of pixels classified to class j and referenced as class i.

Class Accuracy: The CA (or producer's accuracy) is the percentage of correctly classified pixels for a given class

$$\mathrm{CA}_i = \frac{C_{ii}}{\sum_j^K C_{ij}} \times 100\%. \tag{3}$$

Average Accuracy: The AA is the mean of CA values for all the classes

$$\mathrm{AA} = \frac{\sum_i^K \mathrm{CA}_i}{K} \times 100\%. \tag{4}$$

We compare our results to four other methods.

1) The SVM classifier. The parameters used in this case will be detailed for each image.
2) The spectral–spatial ECHO classifier. (Here, we used the MultiSpec software [2], [23] to classify the images. We tested several sets of parameters, and only the results for the best configuration are reported.)
3) A spectral–spatial classifier proposed by Tarabalka *et al.* using a watershed segmentation combined with a majority voting postprocessing step. (A detailed description of this classification method can be found in [41]. The method we compare our results with is referred to as "WHED + MV" in this paper.)
4) The final one is also a spectral–spatial classifier submitted by Tarabalka *et al.* in [33]. It is based on an MSF segmentation combined with a majority voting postprocessing step. The method used for comparison with the proposed approach is referred to as "SAM + MV" (as the SAM distance is used to build the MSF) in this paper for the *Indian Pines* and *Hekla* images and as "L1 + MV" (as the L1 distance was used to build the MSF) for the *Center of Pavia* image. The L1 distance was used to classify the Center of Pavia data set as it has been proven to be more efficient than the SAM distance to deal with classification of urban areas [33].

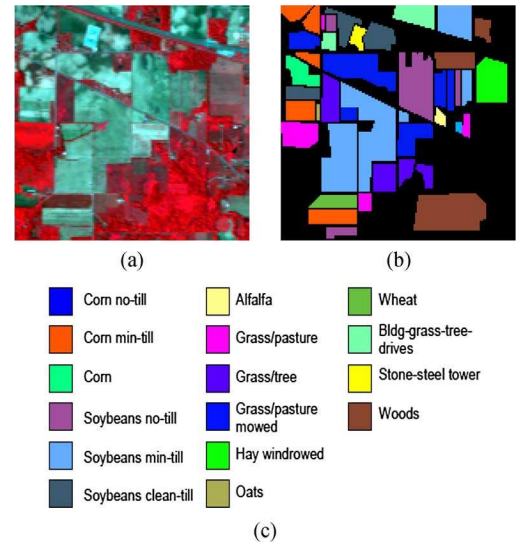

Fig. 6. *Indian Pines* image. (a) Three-band color composite (837, 636, and 537 nm). (b) Reference data.

B. Classification of the Indian Pines Image

The *Indian Pines* image is of a vegetation area that was recorded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines test site in Northwestern Indiana. The image has spatial dimensions of 145×145 pixels and a spatial resolution of 20 m/pixel. Twenty water absorption bands have been removed [53], and a 200-band image was used for the experiments. Sixteen classes of interest are considered. A three-band false color image and the reference data are presented in Fig. 6. We have randomly chosen 50 samples for each class from the reference data as training samples, except for classes " *alfalfa*," " *grass/pasture-mowed*," and " *oats*." These classes contain a small number of samples in the reference data. Therefore, only 15 samples for each of these classes were randomly chosen to be used as training samples. The remaining samples comprised the test set. The SVM classification was performed due to the Library for Support Vector Machines (LIBSVM) tool [54], and the optimal parameters, i.e., $C = 128$ and $\gamma = 0.0156$, were chosen by fivefold cross validation.

In Fig. 7(a) and (b), we can see the evolution of the OA and AA against the number of initial markers used to build the MSF for different values of nb_map. We can first notice the high robustness of the results regarding the value of parameter nb_map. However, the curve for nb_map $= 10$ is slightly under the others. Some experiments have been led in order to quantify the variation of the results to see if this difference was due to a singular result or if it is general. The conclusion is that, for nb_map $= 20$, 30, 40, 50, and 60, the accuracy values have some very similar values and the variation is very small. For lower values of nb_map, the accuracy values tend to decrease and the variation tends to increase. The second observation directly concerns the shape of the curves. Indeed, both OA and AA

TABLE I
CLASSIFICATION ACCURACY VALUES IN PERCENTAGES FOR THE INDIAN PINES IMAGE FOR SEVERAL APPROACHES: OA, AA, AND CA. RO IS THE RATIO BETWEEN THE NUMBER OF TRAINING SAMPLES AND THE NUMBER OF TESTING SAMPLES FOR EACH CLASS

Method		ML	SVM	ECHO	WHED+MV	SAM+MV	RD-MSF
OA		75,41	78,17	82,64	86,63	**91,80**	91,33
AA		79,61	85,97	83,75	91,61	**94,28**	93,73
Class-specific accuracies	Ro						
Corn − no till	0,036	71,4	78,2	83,5	94,2	93,2	**97,5**
Corn − min till	0,064	63,0	69,6	75,1	78,1	**96,6**	81,4
Corn	0,272	85,87	91,9	92,4	88,6	**95,7**	**95,7**
Soybeans − no till	0,055	79,4	82,0	90,1	96,3	93,9	**96,6**
Soybeans − min till	0,021	52,7	58,9	64,1	68,8	**81,9**	78,2
Soybeans − clean till	0,089	86,0	87,9	89,9	90,8	**97,2**	95,9
Alfalfa	0,385	48,7	74,4	48,7	**94,9**	**94,9**	**94,9**
Grass/pasture	0,112	93,5	92,2	94,2	95,1	94,6	**98,7**
Grass/trees	0,072	94,7	91,7	96,3	**98,0**	97,3	97,3
Grass/pasture-mowed	1,364	36,4	**100**	36,4	**100**	**100**	**100**
Hay−windrowed	0,114	97,7	97,7	97,7	99,5	**99,8**	**99,8**
Oats	3	100	100	100	100	100	100
Wheat	0,3086	98,2	98,8	98,2	**99,4**	**99,4**	**99,4**
Woods	0,04 95,4	93,0	94,2	97,1	**99,7**	99,3	
Bldg−Grass−Tree−Drives	0,152	73,0	61,5	**81,5**	69,4	68,8	68,5
Stone−steel towers	1,111	**97,8**	**97,8**	**97,8**	95,6	95,6	95,6

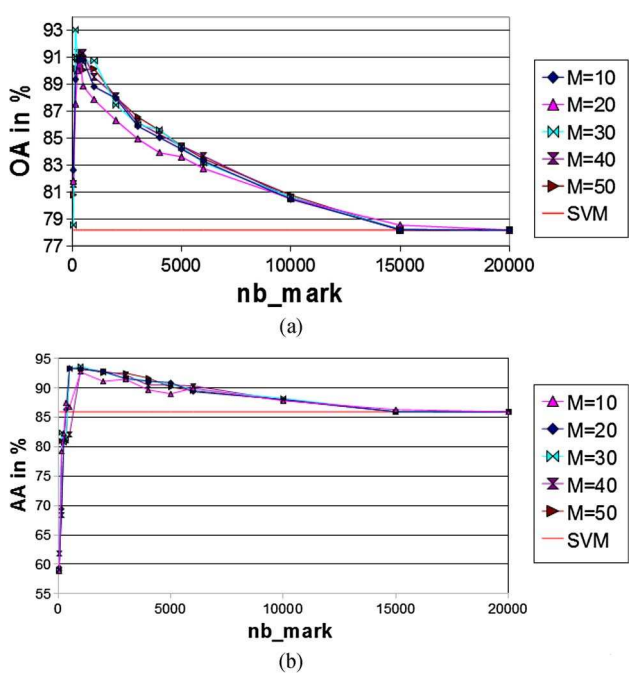

Fig. 7. (a) Evolution of the OA against the number of initial markers. (b) Evolution of the AA against the number of initial markers. In red, the OA and AA obtained for a single SVM classification are shown. Image: *Indian Pines*.

have a similar global behavior. First, when nb_mark increases, the accuracy values are increasing really quickly until reaching a peak. This peak can be more or less wide and can be roughly characterized by a plateau. After this maximum, the accuracy values will start to decrease up to the accuracy values of the initial classification map (here, the SVM classification map). Furthermore, we can note that the results are significantly improved (McNemar test shows that the difference between the SVM and RD-MSF classifiers is statistically significant at the 5% level of significance with $|z| = 4.4286$) for both OA and AA, in the area of the peak, when compared with the initial values of the SVM classification. Therefore, it could be very interesting to find a way to choose a value of nb_mark that will give the best results in terms of classification accuracy values in an automatic way. In this case, the best results are obtained for values of nb_mark between 500 and 1000, which correspond to 2.4% and 4.7% of the total number of pixels in the image, respectively. The idea is then to compare these percentages with what will be obtained for the other data sets.

Table I shows the OA, AA, and CA values for several classification approaches. Several observations can be made when looking at this chart. First, the significant increase in the accuracy values when the RD-MSF classifier is applied, as compared with the results of a pixelwise classifier such as ML or SVM, is obvious. It is also shown that, except for the last class *stone–steel towers*, all the CA values are enhanced. We also observe that the two spectral–spatial classifiers that are using an MSF-based segmentation (SAM + MV and RD-MSF) provide better results than all the others. Furthermore, even if the proposed method does not give the very best accuracy values, the values are almost similar to the ones obtained with the SAM + MV classifier. The point of this last observation is that, when an MSF is built, the most difficult task consists in finding the right seeds. In [33], Tarabalka *et al.* used a probabilistic map and a connected component analysis to determine the most reliable pixels that will be set as markers for the construction of an MSF. The problem of this method, even if it provides very good results, is the generation of the probabilistic map. This step is quite complex and time consuming. The proposed method allows choosing very easily the seed of the MSF, but the main problem is the choice of the good value for parameter nb_mark. Thus, it would be very interesting to find out a rule to automatically determine the number of markers.

Fig. 8 shows the classification maps for the *Indian Pines* image obtained by different classifiers. Here again, a striking evolution is shown between the first techniques for classification of hyperspectral data, which were very noisy [see Fig. 8(a)–(c)], and the new algorithms providing much more accurate classification results. It is also shown that the classification map for the WHED + MV method [see Fig. 8(d)] is still subject to a slight

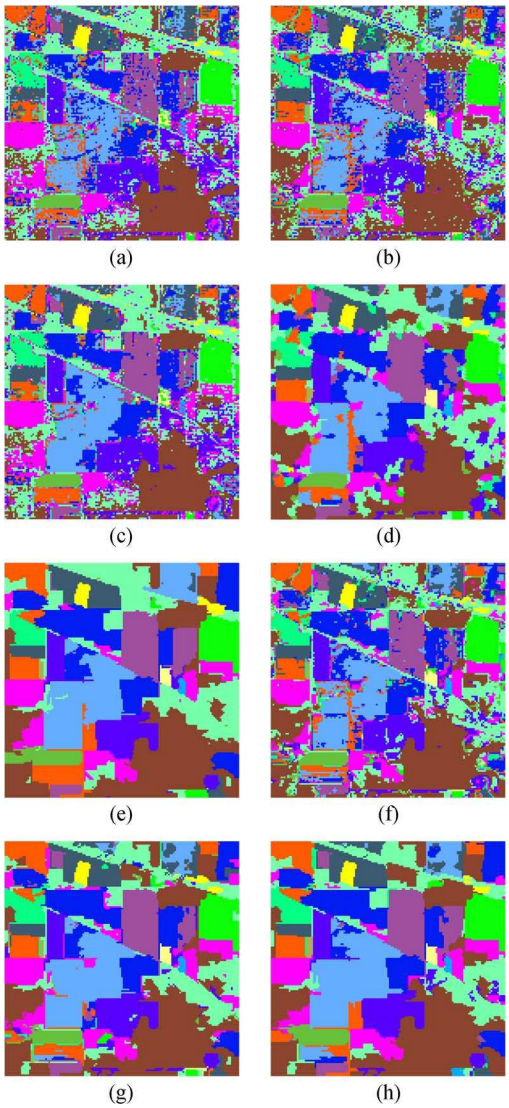

Fig. 8. Classification maps for the *Indian Pines* image using different classifiers. (a) ML. (b) SVM. (c) ECHO. (d) WHED + MV. (e) SAM + MV. (f) RD-MSF for nb_map = 30 and nb_mark = 6000. (g) RD-MSF for nb_map = 30 and nb_mark = 1000. (h) RD-MSF for nb_map = 30 and nb_mark = 500.

oversegmentation. There is also a very interesting point to highlight. If we carefully look at the image and, particularly, the area at the top in the middle, we can see that this region is classified, in most cases, as *soybeans-clean till*, except for the SAM + MV classification where it is classified as *bldg–grass–tree–drives*. From the three-band image in Fig. 6(a), it is clearly shown that this area is composed of two different kinds of soil. As a consequence, when analyzing an image, the visual appearance is also an important factor to take into account when assessing the efficiency of a method. Thus, even if a method provides very good results in terms of accuracy values, we must not forget that the calculations of the efficiency values are based on reference data. In addition, as it is shown in Fig. 6(b), the reference data do not cover the entire image. The classification map can be then a tool to estimate the visual accuracy, which can be characterized by the presence of noise, oversegmentation, or misclassified areas. It can be concluded that the proposed method, even if it is not

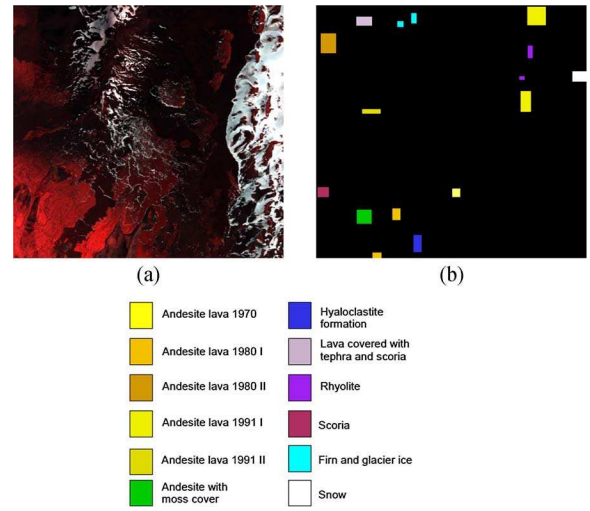

Fig. 9. *Hekla* image. (a) Three-band color composite (1125, 636, and 567 nm). (b) Reference data.

better from an accuracy value's point of view than the SAM + MV method, might be more accurate from a visual point of view.

Fig. 8(f) shows the phenomenon of oversegmentation that starts to appear when we take a too high proportion of pixels as markers. Fig. 8(g) and (h) displays two classification maps with the best results for the RD-MSF algorithm. It is shown that, as for the SAM + MV method, the areas are well defined, and the noise that was present in the SVM classification map has been removed.

C. Classification of the Hekla Image

The *Hekla* image was acquired by the AVIRIS sensor over the region surrounding the central-volcano Hekla in Iceland [55]. The AVIRIS sensor operates in the wavelength range from 0.4 to 2.4 μm, and it utilizes four spectrometers collecting 224 data channels. During data collection, spectrometer four was not properly working. The 64 data channels recorded by this spectrometer were deleted from the data, along with the first channels for the other three spectrometers (those channels were blank). Therefore, the 157 remaining data channels were used for experiments. The considered image has spatial dimensions of 560 $\times$ 600 pixels. Twelve land cover classes of interest are considered. Fig. 9 depicts a three-band false color image and the reference data. Fifty samples for each class were randomly chosen from the reference data as training samples, and the rest of the samples were used as the test set. SVM classification was performed with the MultiSpec software, and the optimal values for C and γ determined by fivefold cross validation are the following: $C = 100$ and $\gamma = 0.1$.

Fig. 10(a) and (b) describes the evolution of the OA and AA against the number of initial markers used to build the MSF for different values of nb_map. Here, nb_mark is ranging from 1000 to 50 000. It is shown on these curves that the results are very robust regarding the value of nb_map, as for the Indian Pines image. In addition, a significant improvement (McNemar test shows that the difference between the SVM and RD-MSF classifiers is statistically significant at the 5% level of significance

TABLE II
CLASSIFICATION ACCURACY VALUES IN PERCENTAGES FOR THE *Hekla* IMAGE FOR SEVERAL APPROACHES: OA, AA, AND CA. RO IS THE RATIO BETWEEN THE NUMBER OF TRAINING SAMPLES AND THE NUMBER OF TESTING SAMPLES FOR EACH CLASS

Method		ML	SVM	ECHO	WHED+MV	SAM+MV	RD-MSF
OA		96,18	88,56	96,63	96,93	**98,96**	97,89
AA		96,99	89,44	97,67	97,57	98,45	**98,57**
Class-specific accuracies	Ro						
Andesite lava 1970	0,171	99,0	88,4	99,7	97,3	**100**	99,3
Andesite lava 1980 I	0,076	98,9	87,3	99,2	97,7	**100**	**100**
Andesite lava 1980 II	0,035	94,3	88,2	94,3	95,9	99,9	**100**
Andesite lava 1991 I	0,019	94,0	84,9	94,4	95,3	**99,6**	94,8
Andesite lava 1991 II	0,139	96,4	93,3	96,9	99,7	88,9	**100**
Andesite lava with moss cover	0,051	98,2	94,2	98,3	**100**	98,5	**100**
Hyalocastite formation	0,079	98,7	87,5	99,4	97,2	99,7	**100**
Lava with tephra and scoria	0,077	96,2	91,7	96,3	**97,5**	97,4	95,1
Rhyolite	0,141	92,4	85,9	95,5	98,9	**100**	98,6
Scoria	0,1	97,6	74,2	**99,6**	93,8	97,6	95,0
Firn and glacier ice	0,123	100	100	100	100	100	100
Snow	0,075	98,2	98,5	98,5	97,6	**100**	**100**

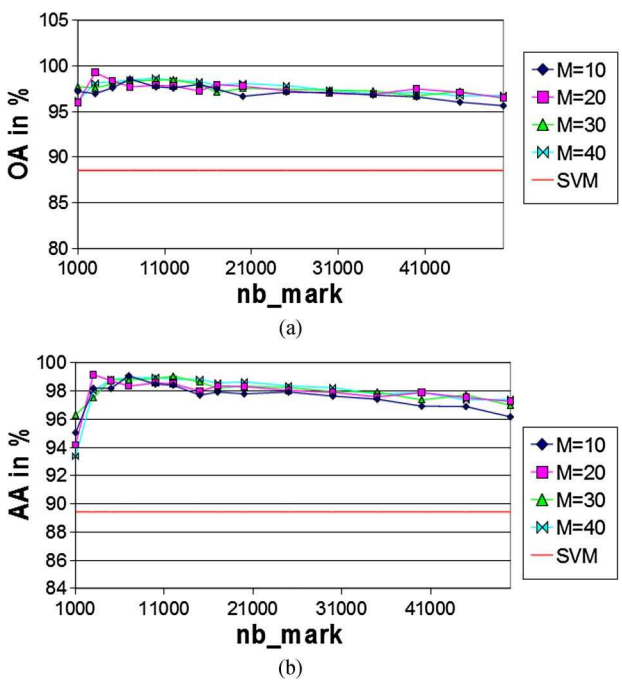

Fig. 10. Evolution of (a) OA and (b) AA against the number of markers. Image: *Hekla*.

with $|z| = 27.6602$) of both OA and AA can be observed. A similar shape of *"quick increase–plateau–slow decrease"* that was obtained for the previous image can be observed. The *increase* phase is not visible on the OA graph, but the end of this step can be seen on the AA graph. It can be also observed that the decrease is much slower than that for the previous image, which emphasizes the plateau aspect of the curves. As for the Indian Pines image, the values of nb_mark that are giving the best results have been sought. In the case of the Hekla image, the maximum accuracy values are obtained when using about 3% of the total number of pixels in the image as markers. This value is coherent with the results that are obtained for the Indian Pines image (between 2.4% and 4.7%), and thus, it seems that a rule to determine the number of pixels to set as markers could be established.

As for the Indian Pines data set, the performance of the RD-MSF are compared with other classifiers. The AA, OA, and CA values are depicted in Table II. Once again, a high improvement can be observed between the original SVM classification and the final RD-MSF classification (+9.33 percentage points for OA and +9.13 percentage points for AA). Furthermore, it is important to note that, in this case, the SVM classifier is far less accurate than the ML classifier. However, the results obtained after application of the introduced method are very good even if the initial classification map is not very accurate. The SAM + MV technique gives again the best result for the OA even if the RD-MSF is close behind (only one percent less), but here, the AA values are almost similar for both SAM + MV and RD-MSF (0.12%). It is also interesting to note that we find again a significant improvement of accuracy values between the spectral–spatial classifiers, which are both based on an MSF segmentation algorithm, and the other classifiers.

Fig. 11 shows the corresponding classification maps for the *Hekla* image. It can be noticed that, when looking at these classification maps, the classification maps for the ML and ECHO methods are very noisy [see Fig. 11(a) and (b)], although good classification accuracy values were obtained. In addition, in these two maps, the *hyaloclastite formation* (dark blue) seems to be overclassified (i.e., too many regions are detected, hence, the noisy aspect of the classification map). Thus, even if the numeric accuracy values are better for the ML and ECHO classifications, this may not be a reliable result. Nevertheless, we can still see the evolution between the initial SVM classification map, the WHED + MV classification map [see Fig. 11(d)], which looks more accurate, and both the RD-MSF classification map [see Fig. 11(f)] and the SAM + MV classification map [see Fig. 11(e)], where the noise has almost disappeared.

D. Classification of the Center of Pavia Image

The *Center of Pavia* image is of an urban area that was recorded by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor. The image used for the experiments is 900 × 300 pixels, with 102 spectral channels (the 13 most noisy channels have been removed). The reference data contain nine classes of interest. A three-band false color image and the reference data are presented in Fig. 12. Thirty samples for each class were randomly chosen from the reference data as training

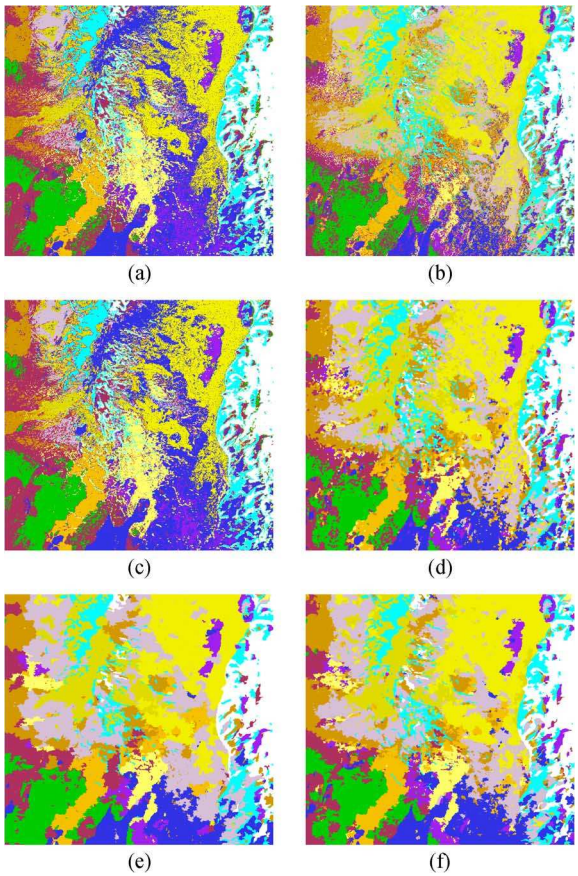

Fig. 11. Classification maps for the *Hekla* image for different classifiers. (a) ML. (b) SVM. (c) ECHO. (d) WHED + MV. (e) SAM + MV. (f) RD-MSF for nb_map = 20 and nb_mark = 10 000.

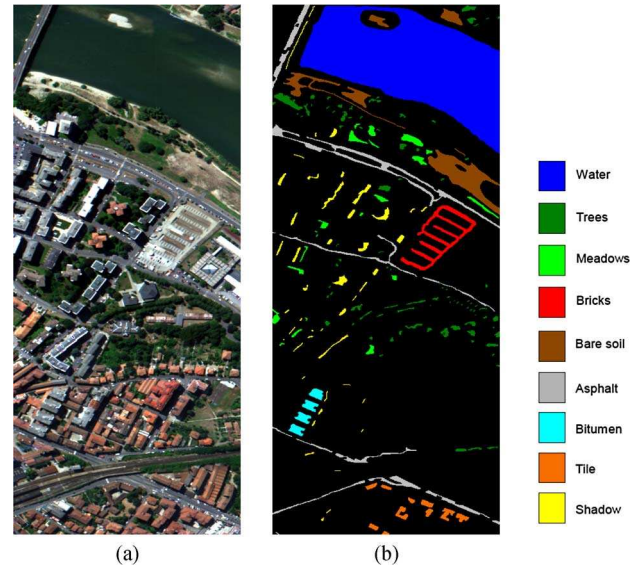

Fig. 12. *Center of Pavia* image. (a) Three-band color composite (650, 558, and 478 nm). (b) Reference data.

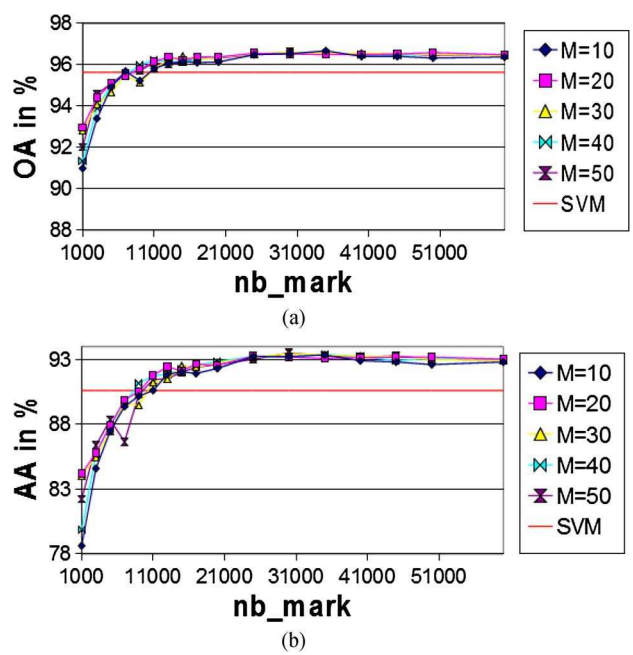

Fig. 13. Evolution of (a) the OA and (b) AA against the number of markers. Image: *Center of Pavia*.

samples. The remaining samples comprised the test set. SVM classification was performed using the MultiSpec software. The optimal values for C and γ were determined by fivefold cross validation ($C = 128$ and $\gamma = 0.0156$).

Fig. 13(a) and (b) describes the evolution of the OA and AA against the number of initial markers used to build the MSF for different values of nb_map. Here, nb_mark ranges from 1000 to 60 000. As for the first two images (Indian Pines and Hekla), high robustness regarding the variation of nb_map can be observed. Moreover, in this case, the shape of the curve is much closer to the expected *plateau* than for the other images. It might be explained by the lowest gain in accuracy values. If we consider that the limit value for the accuracy values when nb_mark tends to be very high (which means almost all the pixels of the image) is close to the maximum value, then the slope will be very low. If we now try to determine the range of values for nb_mark that gives the best results, we can see that it is centered around 30 000 pixels, which means about 11% of the total pixels. This difference of optimal value for nb_mark might be surprising, but it can be explained. In the case of the Hekla and Indian Pines images, the best classifications are obtained for 3% of the pixels set as markers. Moreover, for these two images, all the areas are quite big, and most of the time, they are well delimited. Thus, even if the number of seeds for the building of the MSF is low, the classification will be still reasonably accurate, and the underclassification phenomenon will only appear for very low values of nb_mark. It is quite different in the case of the Center of Pavia image. This data set, as previously mentioned, is one of the urban areas and, thus, is composed of many little entities. As a consequence, to avoid a severe undersegmentation during building of the MSF, the number of markers should be high enough to statistically have a chance to find at least one seed in each different region of the image. As a conclusion, an optimal value for nb_mark is higher for the Center of Pavia image.

In Table III, the results for the different classifiers are gathered. Here, we would like to draw the attention to several points regarding this table. As mentioned in Section III-A, the results

TABLE III
CLASSIFICATION ACCURACY VALUES IN PERCENTAGES FOR THE *Center of Pavia* IMAGE FOR SEVERAL APPROACHES: OA, AA, AND CA. RO IS THE RATIO BETWEEN THE NUMBER OF TRAINING SAMPLES AND THE NUMBER OF TESTING SAMPLES FOR EACH CLASS

Method		ML	SVM	ECHO	WHED+MV	L1+MV	RD-MSF(SAM)	RD-MSF(L1)
OA		95,75	95,64	96,22	96,26	96,62	96,53	**97,17**
AA		91,13	90,60	92,47	92,08	92,55	93,19	**94,03**
Class-specific accuracies	Ro							
Water	8,74e-04	99,4	99,8	99,4	99,8	**99,9**	99,8	**99,9**
Trees	0,012	92,7	90,3	93,9	92,0	90,3	**94,2**	94,0
Meadows	0,017	87,1	96,4	88,2	97,8	96,9	98,2	**98,6**
Bricks	0,014	79,3	64,0	**82,6**	64,2	61,1	71,2	69,4
Bare soil	0,006	90,0	88,4	91,1	90,7	**93,0**	90,4	92,0
Asphalt	0,005	89,7	90,5	90,5	91,9	93,9	91,5	**95,4**
Bitumen	0,032	88,0	87,5	92,4	93,3	**99,9**	96,5	97,4
Tile	0,028	95,7	98,7	95,9	99,0	**100**	**100**	99,6
Shadows	0,014	98,4	99,9	98,4	**100**	**100**	96,9	**100**

of the proposed approach are compared with those of the MSF classifier described in [33]. However, the L1 dissimilarity measure was used here instead of the SAM distance. The reasons for this choice are dual. Indeed, the SAM distance is designed for not taking into account the intensity of the light in an area, in other words, for ignoring the shadows. In an urban area, the shadow areas are quite numerous. Moreover, in the case of this image, the shadow class is among the ones to be classified. As a consequence, the SAM distance might not be the most suitable one for classifying such images. Furthermore, the L1 distance has been proven to be more adapted for classification of urban area images [33]. Nevertheless, because we wanted our method to be as robust as possible, whatever the nature of the input image, we still run the RD-MSF with the SAM distance (referred to as RD-MSF(SAM) in the chart). However, the results for an RD-MSF classification using an L1 dissimilarity distance are also shown in Table III [see RD-MSF(L1)].

If we consider the RD-MSF(SAM) classifier (the results given in Table III are obtained with nb_mark $= 30\,000$ and nb_map $= 20$), we can see that, even if the starting classification map (SVM) is the worst one in terms of accuracy values, the final classification results are among the best ones. However, on a CA level, the difference of classification of the shadow class between the L1 + MV and RD-MSF(SAM) classifiers is obvious (more than 3%). If now we consider the results obtained with the RD-MSF(L1) classifier, [1] we can see that, in addition to the fact that now the OA and AA are the best ones, the shadow class is now much more accurately classified. It is also interesting to note that, with this choice of dissimilarity measurement, the number of markers needed to perform a good classification is reduced (20 000 markers correspond to 7.4 of the total number of pixels).

In Fig. 14, the classification maps for the *Center of Pavia* image based on several classifiers are depicted. We can observe that the SVM classification map is a bit noisy and that this noise aspect has been removed by the application of the stochastic MSF. If we focus on the bridge (upper left corner), we can clearly see the improvement between the RD-MSF(SAM) and RD-MSF(L1) classifiers for classification of shadows. In addition, in the lowest part of the image, the railway is misclassified as water. The problem of this area is that there is no class existing to classify railways. Thus, all the pixels belonging to this potential class will be assigned to the closest class. Hence, a bad classification is obtained in this part of the image.

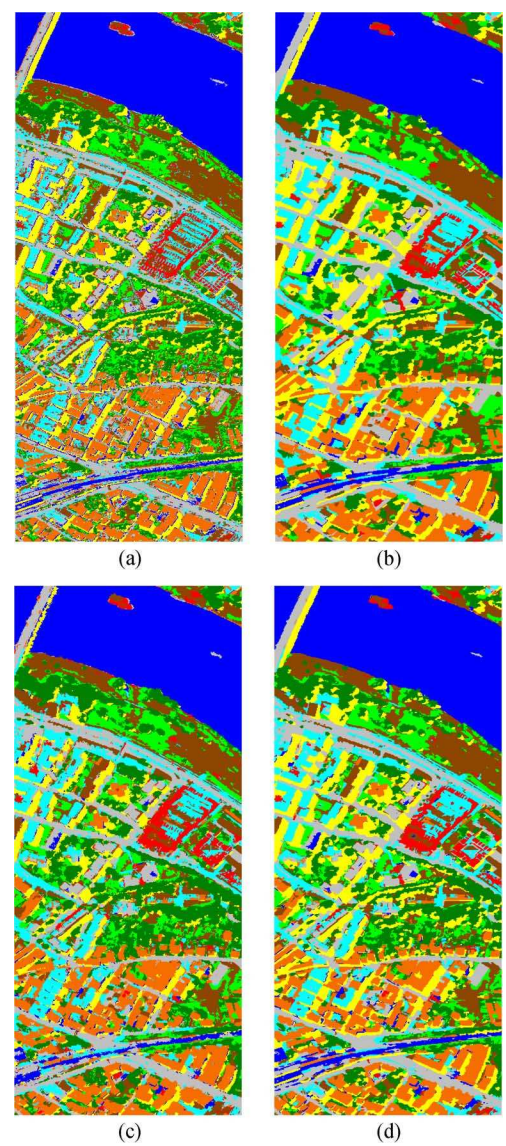

Fig. 14. Classification maps for the *Center of Pavia* image. (a) SVM. (b) L1 + MV. (c) RD-MSF(SAM) for nb_map $= 20$ and nb_mark $= 30\,000$. (d) RD-MSF(l1) for nb_map $= 20$ and nb_mark $= 20\,000$.

[1]In this case, nb_mark $= 20000$ and nb_map $= 20$.

IV. Discussion and Conclusion

In Section III, the results obtained using the introduced method of stochastic MSF are detailed. When analyzing these results, several conclusions can be drawn. First of all, for every image, the two classifiers that are based on an MSF approach yield the best results. Thus, we can assess that the use of an MSF to classify hyperspectral images is very efficient and is quite robust to the choice of the initial seeds for the growing procedure of the forest. Nevertheless, whatever the way to choose the markers (stochastic approach or selection of the most relevant pixels [33]), it is always a delicate step. In the method of Tarabalka *et al.* proposed in [33], the computation of an SVM probabilistic map [56] is needed. In the proposed method, this time-consuming step is avoided because the probability estimates are not needed to select the markers. As a result, the pixelwise classification execution time is reduced by 5%–10%. The random selection of the markers and the generation of the marker maps in the proposed RD-MSF method are also faster than the marker selection by connected component analysis in the previous MSF-based classification approach [33]. Moreover, construction of an MSF for each marker map can be easily executed in parallel [57], [58]. As a result, with an efficient implementation, the stochastic MSF runs about 5%–10% faster than the previous method of Tarabalka *et al.* [33].

The new approach requires less parameters to tune than the previous MSF-based technique. Apart from the SVM parameters, two parameters must be selected for the proposed approach, i.e., a number of initial markers and a number of realizations to fuse in the maximum vote step. Due to the experiments that were performed, default values can be proposed for these parameters. First, for nb_map, it has been observed in our experiments that above 20 maps, the improvement of the accuracy values is not significant. As increasing the value of nb_map will also increase the computation time without giving better results, we suggest to set this parameter to 20. For the right value of nb_mark, this is not as simple. It has been shown that, in the case of images mainly composed of wide areas (such as Indian Pines and Hekla), the optimal values for nb_mark are centered around 3.5% of the total number of pixels of the image to classify. However, in the case of images such as Center of Pavia, which are composed of many small areas and that can contain shadows areas, the number of markers needed to achieve good classification is highly increased and reaches 11% of the image's pixels. In addition, as a conclusion, it implies that, to have the best results, the user should have a prior knowledge of the image. Thus, to classify nonurban areas (e.g., agricultural fields and volcanic regions), the user should use 3.5% of the pixels as markers and 11% in the case of urban areas. Once these two parameters are set, the last choice that has to be made is the one of the dissimilarity measure that is used in the construction of the MSF. In our case, we used the SAM distance as it has been proven to be well suited for the classification of hyperspectral images [52]. However, in Section III-D, we have highlighted the fact that this distance is not well suited to classify urban areas, particularly because of the presence of numerous shadow areas. Thus, before trying to classify a hyperspectral image with the stochastic MSF method, the user should try to find which dissimilarity measure is the best suited with his needs.

It might be also interesting to investigate other classification functions than maximum vote. A weighted decision rule approach was also investigated where the OA of each classification map was used as a weight for the maximum vote procedure. The obtained results were slightly lower when compared with those of the proposed method. Some more sophisticated methods might give some better results but will probably imply to possess some knowledge of the image to analyze.

As a conclusion, the introduced method of classification of hyperspectral images by using a stochastic MSF approach has been proven to be very efficient. The results are better than the standard classifier such as ML, SVM, or ECHO. They also outrun the spectral–spatial classifier based on a watershed segmentation approach. Moreover, they are always very close to the results of the MSF segmentation method proposed in [33] and even slightly better in some cases. Regarding the computational complexity, it has been observed that the marker selection step is now simpler to implement than in [33], which is mainly due to the suppression of the computation of the probabilistic map. Furthermore, the new approach is 5%–10% faster than the previously proposed MSF-based technique [33].

Acknowledgment

The authors would like to thank P. Gamba from the University of Pavia, Italy, and D. Landgrebe from Purdue University, USA, for providing the hyperspectral data.

References

[1] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, Jun. 1985.

[2] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New York: Wiley, 2003.

[3] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[4] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York: Springer-Verlag, 1999.

[5] C. H. Chen and P. G. P. Ho, "Statistical pattern recognition in remote sensing," *Pattern Recognit.*, vol. 41, no. 9, pp. 2731–2741, Sep. 2008.

[6] N. Bali and A. Mohammad-Djafari, "Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 217–225, Feb. 2008.

[7] S. Subramanian, N. Gat, M. Sheffield, J. Barhen, and N. Toomarian, "Methodology for hyperspectral image classification using novel neural network," in *Proc. SPIE*, Aug. 1997, vol. 3071, pp. 128–137.

[8] J. A. Benediktsson, P. H. Swain, and O. K. Ersoy, "Conjugate gradient neural networks in classification of very high dimensional remote sensing data," *Int. J. Remote Sens.*, vol. 14, no. 15, pp. 2883–2903, 1993.

[9] E. Merényi, "Intelligent understanding of hyperspectral images through self-organizing neural maps," in *Proc. 2nd Int. Conf. CITSA*, Orlando, FL, 2005, pp. 30–35.

[10] H. Yang, F. van der Meer, W. Bakker, and Z. J. Tan, "A back-propagation neural network for mineralogical mapping from AVIRIS data," *Int. J. Remote Sens.*, vol. 20, no. 1, pp. 97–110, 1999.

[11] C. Hernandez-Espinoza, M. Fernandez-Renondo, and J. Torres-Sospedra, "Some experiments with ensembles of neural networks for classification of hyperspectral images," in *Proc. ISNN*, 2004, no. 1, pp. 912–917.

[12] P. K. Goel, S. O. Prasher, R. M. Patel, J. A. Landry, R. B. Bonnell, and A. A. Viau, "Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn," *Comput. Electron. Agr.*, vol. 39, pp. 67–93, 2003.

[13] H. Zhou, Z. Mao, and D. Wang, "Classification of coastal areas by airborne hyperspectral image," in *Proc. SPIE*, 2005, vol. 5832, pp. 471–476.

[14] C. Vaiphasa, "Innovative genetic algorithm for hyperspectral image classification," in *Proc. Int. Conf. Map Asia*, 2003, vol. 20, p. 45.

[15] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Evaluation of kernels for multiclass classification of hyperspectral remote sensing data," in *Proc. ICASSP*, May 2006, pp. II-813–II-816.

[16] M. Fauvel, "Spectral and spatial methods for the classification of urban remote sensing data," Ph.D. dissertation, Grenoble Inst. Technol., Grenoble, France, 2007.

[17] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Customizing kernel functions for SVM-based hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 622–629, Apr. 2008.

[18] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, Jun. 1998.

[19] J. A. Gualtieri and R. F. Cromp, "Support Vector Machines for hyperspectral remote sensing classification," in *Proc. SPIE*, 1998, vol. 3584, pp. 221–232.

[20] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[21] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[22] S. Makrogiannis, G. Economou, and S. Fotopoulos, "A region dissimilarity relation that combines feature-space and spatial information for color image segmentation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 1, pp. 44–53, Feb. 2005.

[23] R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Trans. Geosci. Electron.*, vol. GE-14, no. 1, pp. 19–26, Jan. 1976.

[24] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[25] M. Fauvel, J. Chanussot, J. A. Benediktsson, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 3804–3814, Oct. 2008.

[26] A. Farag, R. Mohamed, and A. El-Baz, "A unified framework for map estimation in remote sensing image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 7, pp. 1617–1634, Jul. 2005.

[27] C. Velasco-Forero and J. Angulo, "Morphological scale–space for hyperspectral images and dimensionality exploration using tensor modeling," in *Proc. IEEE WHISPERS (First IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Grenoble, France)*, Aug. 2009, pp. 1–4.

[28] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision*, ser. McGraw-Hill series in Computer Science. New York: McGraw-Hill, 1995.

[29] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power watersheds: A unifying graph-based optimization framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1384–1399, Jul. 2011.

[30] F. Meyer and S. Beucher, "Morphological segmentation," *J. Vis. Commun. Image Represent.*, vol. 1, no. 1, pp. 21–46, Sep. 1990.

[31] L. Vincent and P. Soille, "Watershed in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.

[32] J. Kruskal, "On the shortest spanning tree of a graph and the traveling salesman problem," in *Proc. Amer. Math. Soc.*, 1956, vol. 7, pp. 48–50.

[33] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1267–1279, Oct. 2010.

[34] J. Stawiaski, "Mathematical morphology and graphs: Application to interactive medical image segmentation," Ph.D. dissertation, Paris School Mines, Paris, France, 2008.

[35] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[36] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.

[37] X. Bai and G. Sapiro, "A geodesic framework for fast interactive image and video segmentation and matting," in *Proc. ICCV*, 2007, pp. 1–8.

[38] F. Meyer, "Minimal spanning forest for morphological segmentation," in *Proc. ISMM—Mathematical Morphology and its Applications to Signal Processing*, 1994, pp. 77–84.

[39] C. Allene, M. Couprie, J.-Y. Audibert, and R. Keriven, "Some links between extremum spanning forests, watersheds and min-cuts," *Image Vis. Comput.*, vol. 28, no. 10, pp. 1460–1471, Oct. 2010.

[40] L. N. J. Cousty, G. Bertrand, and M. Couprie, "Watershed cuts: Minimum spanning forests and the drop of water principle," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 8, pp. 1362–1374, Aug. 2009.

[41] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010.

[42] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[43] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Classification of hyperspectral data using support vector machines and adaptive neighborhoods," in *Proc. 6th EARSeL SIG IS Workshop*, Tel Aviv, Israel, 2009.

[44] J. Angulo and D. Jeulin, "Stochastic watershed segmentation," in *Proc. ISMM*, 2007, pp. 265–276.

[45] J. Angulo, S. Velasco-Forero, and J. Chanussot, "Multiscale stochastic watershed for unsupervised hyperspectral image segmentation," in *Proc. IEEE IGARSS*, 2009, pp. III-93–III-96.

[46] J. Angulo and S. Velasco-Forero, "Semi-supervised hyperspectral image segmentation using regionalized stochastic watershed," in *Proc. SPIE Symp. Defense, Security, Sens.—Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI*, Apr. 2010, vol. 7695, p. 769 51F.

[47] F. Meyer, "An overview of morphological segmentation," *Int. J. Pattern Recog. Artif. Intell.*, vol. 15, no. 7, pp. 1089–1118, Nov. 2001.

[48] R. Prim, "Shortest connection networks and some generalizations," *Bell Syst. Tech. J.*, vol. 36, pp. 1389–1401, 1957.

[49] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, J. Inglada, E. Christophe, J. Chanussot, and P. Gamba, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, Nov. 2009.

[50] C.-I. Chang, "An information theoretic-based approach to spectral variability, similarity and discriminability for hyperspectral image analysis," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.

[51] B. L. P. Corsini and F. Marcelloni, "A fuzzy relational clustering algorithm based on a dissimilarity measure extracted from data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 775–781, Feb. 2004.

[52] F. van der Meer, "The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 8, no. 1, pp. 3–17, Jan. 2006.

[53] S. Tadjudin and D. A. Landgrebe, "Covariance estimation with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 4, pp. 2113–2118, Jul. 1999.

[54] C.-C. Chang and C.-J. Lin, LIBSVM: A Library for Support Vector Machines 2001 [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[55] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1367–1377, May 1999.

[56] N. Cristianini and J. Shawe-Taylor, *Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[57] A. Plaza, D. Valencia, J. Plaza, and P. Martinez, "Commodity cluster-based parallel processing of hyperspectral imagery," *J. Parallel Distrib. Comput.*, vol. 66, no. 3, pp. 345–358, Mar. 2006.

[58] Y. Tarabalka, T. V. Haavardsholm, I. Kåsen, and T. Skauli, "Real-time anomaly detection in hyperspectral images using multivariate normal mixture models and GPU processing," *J. Real-Time Image Process.*, vol. 4, no. 3, pp. 287–300, Aug. 2009.

Kévin Bernard received the M.Sc. degree in electrical engineering from Grenoble Institute of Technology (INPG), Grenoble, France, in 2010. He is currently working toward the Ph.D. degree at the University of Iceland, Reykjavik, Iceland, and at Heriot Watt University, Edinburgh, U.K.

His research interests are in the areas of image processing, biomedical imaging and engineering, and more recently, eye tissue optics.

Yuliya Tarabalka (M'10) received the B.S. degree in computer science from Ternopil Ivan Pul'uj State Technical University, Ternopil, Ukraine, in 2005, the M.Sc. degree in signal and image processing from Grenoble Institute of Technology (INPG), Grenoble, France, in 2007, the Ph.D. degree (with European Honours) in signal and image processing from INPG, and the Ph.D. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 2010.

From July 2007 to January 2008, she was a Researcher with the Norwegian Defence Research Establishment, Norway. She is currently a Postdoctoral Research Fellow with the Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, MD. Her research interests are in the areas of image processing, pattern recognition, hyperspectral imaging, and development of efficient algorithms.

Dr. Tarabalka is a member of the IEEE Society and the IEEE Geoscience and Remote Sensing Society.

Jesús Angulo was born in Cuenca, Spain, in 1975. He received the degree in telecommunications engineering from Polytechnical University of Valencia, Valencia, Spain, in 1999, with Master's thesis on image and video processing and the Ph.D. degree in mathematical morphology and image processing from the École des Mines de Paris (Mines ParisTech), Fontainebleau, France, in 2003, under the guidance of Prof. J. Serra.

He is currently a Permanent Researcher (Chargé de Recherche) with the Center of Mathematical Morphology, Department of Mathematics and Systems, Mines ParisTech. His research interests are in the areas of multivariate image processing (color, hyper/multispectral, temporal series, polarimetric, and tensor imaging) and mathematical morphology (filtering, segmentation, shape and texture analysis, stochastic and geometry approaches, and PDE approaches), and their application to the development of theoretically sound and high-performance algorithms and software in the fields of biomedicine/biotechnology, remote sensing, and industrial vision.

Jocelyn Chanussot (M'04–SM'04–F'12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from Savoie University, Annecy, France, in 1998.

In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Generale de l'Armement (DGA–French National Defense Department). Since 1999, he has been with Grenoble INP, where he was an Assistant Professor from 1999 to 2005, an Associate Professor from 2005 to 2007, and is currently a Professor of signal and image processing. He is currently conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab). His research interests include image analysis, multicomponent image processing, nonlinear filtering, and data fusion in remote sensing.

Dr. Chanussot is the founding President of IEEE Geoscience and Remote Sensing French chapter (2007-2010) which received the 2010 IEEE GRS-S Chapter Excellence Award "for excellence as a Geoscience and Remote Sensing Society chapter demonstrated by exemplary activities during 2009." He was the recipient of the 2011 IEEE GRSS Symposium Best Paper Award. He was a member of the IEEE Geoscience and Remote Sensing Society AdCom (2009–2010), in charge of membership development. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS). He is the Chair (2009-2011) and was the Cochair of the GRS Data Fusion Technical Committee (2005–2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006–2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing, (2009). He was an Associate Editor for the IEEE *Geoscience and Remote Sensing Letters* (2005–2007) and for *Pattern Recognition* (2006–2008). Since 2007, he is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. Since 2011, he is the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984 and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 1987 and 1990, respectively.

He is currently the Pro-Rector for Academic Affairs and a Professor of electrical and computer engineering with the University of Iceland. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing, and he has extensively published in these fields.

Dr. Benediktsson is the President of the IEEE Geoscience and Remote Sensing Society for 2011–2012. He was an Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2003 to 2008. He was the recipient of the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing, the Icelandic Research Council's Outstanding Young Researcher Award in 1997, the IEEE Third Millennium Medal in 2000, the yearly research award from the Engineering Research Institute of the University of Iceland in 2006, and the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society in 2007. In 2004, he was also a corecipient of the University of Iceland's Technology Innovation Award. He is a member of Societas Scientiarum Islandica and Tau Beta Pi.