Spectral-Spatial Classification of Hyperspectral Images Based on Hidden Markov Random Fields

Pedram Ghamisi, *Student Member, IEEE,* Jon Atli Benediktsson, *Fellow, IEEE*
Magnus O. Ulfarsson, *Member, IEEE,*

***Abstract*—Hyperspectral remote sensing technology allows one to acquire a sequence of possibly hundreds of contiguous spectral images from ultraviolet to infrared. Conventional spectral classifiers treat hyperspectral images as a list of spectral measurements and do not consider spatial dependencies, which leads to a dramatic decrease in classification accuracies. In this work, a new automatic framework for the classification of hyperspectral images is proposed. The new method is based on combining Hidden Markov Random Field segmentation with Support Vector Machine (SVM) classifier. In order to preserve edges in the final classification map, a gradient step is taken into account. Experiments confirm that the new spectral and spatial classification approach is able to improve results significantly in terms of classification accuracies compared to the standard SVM method and also outperforms other studied methods.**

***Index Terms*—Hyperspectral Image Analysis, Image Segmentation, Hidden Markov Random Field, Support Vector Machine Classifier.**

I. Introduction

DUE to recent advances in hyperspectral sensor technology, it is possible to capture hundreds of spectral channels for each image pixel from ultraviolet to infrared. By increasing the amount of spectral information, the accurate discrimination of different materials of interest is possible. In addition, the fine spatial resolution of the sensors enables the analysis of small spatial structures in the image. Furthermore, the high spectral resolution allows detailed physical analysis of the structures [1].

Classification plays a key role in the analysis of hyperspectral images. Examples of applications where it plays a key role are land-use and land-cover mapping, crop monitoring, forest applications, urban development, mapping, tracking and risk management.

For hyperspectral images, several hundreds of spectral bands of the same scene are typically available, while for multispectral images up to ten bands are usually available. By increasing the dimensionality of the images in the spectral domain, theoretical and practical problems arise. For instance, with a limited training set, beyond a certain limit, the classification accuracy actually decreases as the number of features increases [2]. For the purpose of classification, these problems are related to the curse of dimensionality.

P. Ghamisi, J. A. Benediktsson and M. O. Ulfarsson are with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (corresponding author, e-mail: peg6@hi.is)

This research was supported in part by the Icelandic Research Fund for Graduate Students.

Conventional spectral classifiers treat hyperspectal images as a list of spectral measurements [3]. For instance, Support Vector Machine (SVM) classifiers have received significant attention lately because of their remarkable generalization capability for the classification of high dimensional data sets [4] and their considerable capability for handling big data sets with few number of training samples. The efficiency of SVM classifiers have been shown in terms of achieving very accurate results in a wide variety of applications [5], [6]. However, SVM classifiers do not consider spatial dependencies and classify images only based on their spectral information. Therefore, this approach discards information associated with the correlations among distinct pixels in the image and is considered as the most vital limitation of SVM classifiers for the analysis of remote sensing images in which pixel neighborhoods provide important information [7].

To address the above-mentioned problem, joint spectral and spatial classification techniques have recently received considerable attention. Consideration of spatial information helps to overcome the salt and pepper appearance of the classification. More importantly, other relevant information can be extracted from the spatial domain: for a given pixel it is possible to extract the size and the shape of the structure to which it belongs. Therefore, the combination of spectral and spatial information can improve the result of the classification stage. The goal of considering spatial context in the classification step can partially be achieved by using methods such as morphological filters (e.g., [1]), morphological leveling (e.g., [8]) and Markov random fields (MRFs) (e.g., [9]).

MRFs are a family of probabilistic models that can be described as 2-D stochastic processes over discrete pixels lattices [10]. They can be considered as a powerful tool for incorporating spatial and contextual information into the classification framework [11]. More recently [12], Hidden MRF (HMRF) was introduced as a special case of the Hidden Markov Model (HMM). In HMRF, the underlying stochastic process is MRF, instead of Markov Chains in HMM. Therefore, HMRF is not restricted to 1D and can be used in order to extract spatial information from 2D and 3D images.

There is extensive literature on the use of MRFs for increasing the accuracy of classification. For instance, in [13], the result of the Probabilistic SVM was regularized by a MRF. In [9] the mean field based SVM regression was used for image classification. Also, in [14], [15], [11], [7] and [16], MRFs were taken into consideration for modeling spatial and contextual information for improving the accuracy of the classification. Furthermore, a generalization of MRF, called

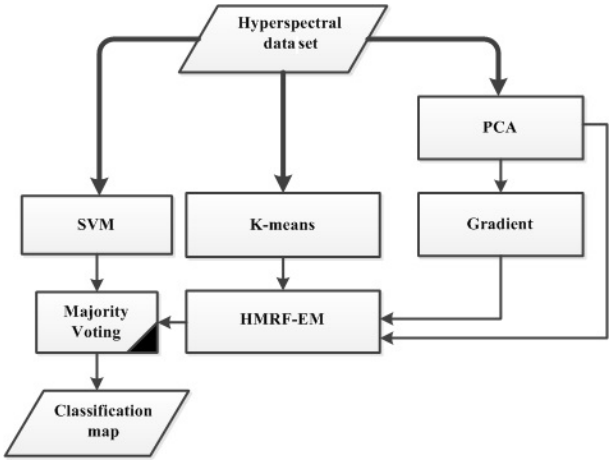

Fig. 1. A flowchart of the proposed method.

conditional MRF, was investigated in [17] for the spectral and spatial classification of remote sensing images. In [18], the concept of HMM was used for incorporating spectral and contextual information into a framework for performing unsupervised classification of remote sensing multispectral images. In addition, Gaussian MRF was employed in [19] for the purpose of segmentation and anomaly detection.

Based on the above, the integration of SVM classifiers and MRFs for the accurate classification of remote sensing images by considering both spectral and spatial information into the same framework is completely obvious. In this work, a novel fully automatic spectral and spatial approach is introduced for the classification of hyperspectral images. The new approach is based on the HMRF and SVM. In order to preserve the edges in the classification map, a gradient step based on the Sobel edge detector is taken into account. In addition, to our knowledge, this is the first time that HMRF is used in the field of remote sensing.

The paper is organized as follows: the proposed methodology is discussed in Section II. Then, Section III is devoted to experimental results. Finally, Section IV outlines the main conclusions.

II. METHODOLOGY

Fig. 1 illustrates the flowchart of the proposed method. In the following, specific parts of the proposed framework will be discussed in detail.

A. Notation

In the following we let $\mathbf{y} = (y_1, ..., y_N)^T$ denote the first principal component map where N is the number of pixels and $S = \{1, 2, ..., N\}$ is the set of pixel indices. Associated with pixel i is a class label x_i. A vector containing these labels is denoted by $\mathbf{x} = (x_1, ..., x_N)^T$.

B. HMRF-EM segmentation by preserving edges

B. 1. Finite Gaussian Mixture

For better understanding the concept of HMRF, we begin with the Finite Gaussian Mixture (FGM) model. For a pixel i we have:

$$q(l) = q(x_i = l)$$
$$p(y_i|l) = g(y_i; \theta_l)$$

where $p(y_i|l)$ is a conditional probability of the intensity y_i given the class label l ($l \in L$ and L is regarded as the set of all possible labels). $q(l)$ is probability mass function of the class label and $g(y_i; \theta_l)$ is a Gaussian probability density function (pdf) with parameter $\theta_l = (\mu_l, \sigma_l^2)$. The marginal distribution of $y = y_i$ dependent on the parameter set $\boldsymbol{\theta} = \{\boldsymbol{\theta}_l, l \in L\}$ can be written as:

$$p(y; \boldsymbol{\theta}) = \sum_{l \in L} g(y; \theta_l) q(l) \cdot \quad (1)$$

Although the FGM model is mathematically simple, it is not able to take the spatial information into consideration since all the data points are considered individually and are independent from the other neighborhood points. To overcome this limitation the HMRF was proposed in [12].

B. 2. HMRF Model

HMRF is a generalization of HMM. While HMM is based on 1D Markov chains, HMRF are based on hidden random fields. Due to its ability to handle 2D structure HMRF is more suitable for image segmentation than HMM.

The Gaussian HMRF is given by:

$$\begin{aligned} p(\mathbf{x}, \boldsymbol{y}; \boldsymbol{\theta}) &= f(\mathbf{x}) \prod_{i=1}^{N} p(y_i|x_i) \\ p(y_i|x_{N_i}; \boldsymbol{\theta}) &= \sum_{l \in L} g(y_i; \theta_l) q(l|x_{N_i}) \end{aligned} \quad (2)$$

where $f(\mathbf{x})$ is a pdf for x which follows the so-called Gibbs densities [20] and $q(l|x_{N_i})$ is a conditional probability mass function (pmf) for the class label l given x_{N_i} denotes a neighborhood for each pixel x_i. The difference between HMRF and FGM is the term $q(l|x_{N_i})$ in (2) and the term $q(l)$ in (1). If we do not consider the relationship between pixels in the neighboring system, HMRF and FGM are the same. In other words, spatial dependencies can be modeled in HMRF which are discarded in FGM. Therefore, the FGM model is a special case of HMRF. As a result, it can be concluded that HMRF is more flexible than FGM since it is able to model both the statistical and spatial properties of the image.

The model fitting procedure [12] involves an initialization and an iteration between two steps: Maximum a Posteriori (MAP) estimation of the class labels, and an Expectation-Maximization (EM) algorithm [21] for estimating $\boldsymbol{\theta}$. Now we consider these three steps.

B. 2. 1. Initialization

The output of this step provides the initial label $\mathbf{x}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$ for the MAP and EM algorithm respectively. In this paper,

K-means was used to provide the initial labels and initial parameters $\boldsymbol{\theta}$ were computed for the initialization step. The initial parameters are obtained by estimating the mean and the standard deviation of the pixels within each cluster.

K-means [22] is as one of the best-known clustering methods which was introduced by MacQueen (1967). This method starts with a random initial partition of the pixel vectors into candidate clusters and then reassigns these vectors to clusters by reducing the squared error in each iteration, until a convergence criterion is met.

B. 2. 2. MAP

From one point of view, image segmentation can be split into two categories: structural and statistical. The former is based on boundaries and regions. On the other hand, the latter is mostly based on the probability distribution function of image intensities and their associated class labels. Statistical approaches try to find the class label x, when only the intensity y for each pixel is given. Maximum a Posteriori (MAP) or Maximum Likelihood (ML) are widely used criteria for this kind of estimation. Using the MAP criterion, $\hat{\mathbf{x}}$ should be estimated based on:

$$\hat{\mathbf{x}} = \arg\max_{x\in\chi} \{p\left(\boldsymbol{y}\left|\mathbf{x}\,;\boldsymbol{\theta}\right.\right) f\left(\mathbf{x}\right)\}\cdot \tag{3}$$

It is assumed that y_i and x_i are pairwise independent so

$$p(\boldsymbol{y}|\mathbf{x},\boldsymbol{\theta}) = \prod_{i=1}^{N} p(y_i|x_i)\cdot$$

MRF can be completely explained by a Gibbs distribution using the Hammersley-Clifford theorem which describes the relation between MRF and Gibbs distribution [20]. Thus,

$$f(\mathbf{x}) = \frac{1}{Z}\exp(-U(\mathbf{x}))$$

where Z is a normalizing constant and

$$U(\mathbf{x}) = \sum_{c\in C} V_c(x_i, x_j)$$

where $V_c(x_i, x_j)$ are the so-called clique potentials and C is the set of all possible cliques; see more details in [20]. A clique c is a subset of S where every pair of distinct sites is neighbors, except for single-site cliques. Fig. 2 depicts all possible cliques for the predefined neighborhood system. The general idea behind the HMRF model is that if a pixel has a certain label, the pixels of its neighborhood system are also of that type. In this paper, it is assumed that each pixel has at most 4 neighbors in the image domain. Then, on pairs of neighboring pixels, the clique potentials are calculated by:

$$V_c(x_i, x_j) = \frac{1}{2}(1 - I_{x_i,x_j}) \tag{4}$$

$$I_{x_i,x_j} = \begin{cases} 0 & \text{if } x_i \neq x_j \\ 1 & \text{if } x_i = x_j. \end{cases}$$

MAP can be rewritten as a minimization problem

$$\hat{\mathbf{x}} = \arg\min_{x\in\chi} \{U(\boldsymbol{y}|\mathbf{x}) + U(\mathbf{x})\} \tag{5}$$

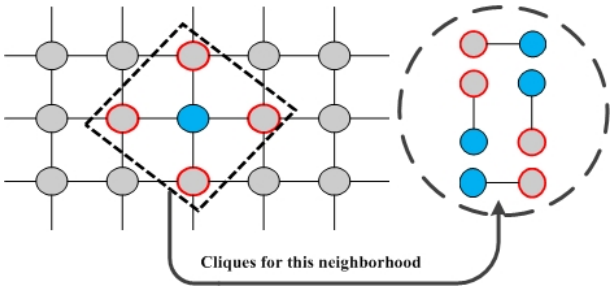

Fig. 2. All possible cliques for the predefined neighborhood system.

where $U(\boldsymbol{y}|\mathbf{x}) = \sum_i \left[\frac{\left(y_i-\mu_{x_i}\right)^2}{2\sigma_{x_i}^2} + \frac{1}{2}\log\sigma_{x_i}^2\right]$ measures the fit, and $U(\mathbf{x})$ can be viewed as a penalty term that encourages spatial smoothness. The iterative MAP algorithm stops when the relative change in the cost function is below a pre-specified threshold. There exist efficient algorithms for solving the MAP problem. Here, we use the same algorithms as in [12].

B. 2. 3. EM algorithm

A statistical model is complete if and only if both its functional forms and parameters are determined. In HMRF, the parameter set $\boldsymbol{\theta} = \{\theta_l, l \in L\}$ should be estimated. If the Gaussian density function is assumed for the pixel intensity value y, the parameters of each Gaussian class are $\theta_l = (\mu_l, \sigma_l)$. Since both the class labels and parameters are unknown, the calculation of the parameters is not straightforward. One reliable way to solve this issue is the EM algorithm [21]. We use the EM algorithm to estimate the parameters $\boldsymbol{\theta}$. Below, the EM algorithm is briefly explained:

1) E-step: We compute the EM functional:

$$Q\left(\boldsymbol{\theta}\left|\boldsymbol{\theta}^{(k)}\right.\right) = E\left[\log p(\boldsymbol{y},\mathbf{x};\boldsymbol{\theta})|\boldsymbol{y},\boldsymbol{\theta}^{(k)}\right]\cdot \tag{6}$$

2) M-step: For obtaining the next estimate we maximize the EM functional

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta}\left|\boldsymbol{\theta}^{(k)}\right.\right)\cdot \tag{7}$$

Then, let $\boldsymbol{\theta}^{(k)} \longrightarrow \boldsymbol{\theta}^{(k+1)}$ and return to the E-step.

The EM functional can be written as

$$Q = \sum_i \sum_l q^{(k)}(j|y_i)\left\{\ln q(l|x_{N_i}) - \frac{1}{2}\ln\sigma_l^2 - \frac{1}{2}\frac{(y_i-\mu_j)^2}{\sigma_l^2}\right\} \tag{8}$$

where the posterior $q^{(k)}(j|y_i)$ is obtained from the MAP step. The M-step yields the following updates

$$\mu_l^{(k+1)} = \frac{\sum_i q^{(k)}(j|y_i)y_i}{\sum_i q^{(k)}(j|y_i)} \tag{9}$$

$$\sigma_l^{2(k+1)} = \frac{\sum_i q^{(k)}(j|y_i)(y_i-\mu_l^{(k+1)})}{\sum_i q^{(k)}(j|y_i)}. \tag{10}$$

The iterative algorithm will stop when the relative change in the cost function is less than a predefined threshold.

C. Gradient

Image segmentation provides a smoothing process. Provided that one image has strong discontinuities, MRFs may cause over smoothing [23]. One way for addressing this issue is to combine the underlying label with an additional line process [23]. In order to preserve edges in the segmentation map, the input image is first transformed by Principal Component Analysis (PCA) and the PCs which have dominant variance (more than 99 percent of the total variation) are kept. Sobel edge detection is performed on each PC and then the output of Sobel edge detected PCs are summed together. Finally, the output is transformed to a binary format. Let us assume that we have a binary edge map z; $z_i = 1$ if the i-th pixel is edge and $z_i = 0$ if not. In this case, (5) is modified to

$$\hat{\mathbf{x}} = \arg\min_{x \in \chi} \left\{ U\left(\boldsymbol{y}\,|\mathbf{x}\right) + \sum_{j \in N_i, z_i = 0} V_c(l, \mathbf{x}_N^{(k)}) \right\}. \quad (11)$$

This shows that the clique potentials are only estimated for the pixels which are not edge pixels.

D. SVM

The general idea behind SVM is to separate training samples belonging to different classes by tracing maximum margin hyperplanes in the space where the samples are mapped [24]. SVMs were originally introduced for solving linear classification problems. However, they can be generalized to non-linear decision functions by considering the so-called kernel trick [25]. A kernel-based SVM is being used to project the pixel vectors into a higher dimensional space and estimate maximum margin hyperplanes in this new space, in order to improve linear separability of data [25]. The sensitivity to the choice of the kernel and regularization parameters can be considered as the most important disadvantages of SVM. The latter is classically overcome by considering cross-validation techniques using training data [26]. The Gaussian radial basis function (RBF) is widely used in remote sensing [25].

E. Majority Voting

In this work, Majority Voting (MV) is used for combining the result of the segmentation and classification steps. Fig. 3 shows the general idea of MV. The output of the segmentation methods is a number of objects where each object consists of several pixels with the same label. In other words, pixels in each object share the same characteristics. For performing MV on the output of the segmentation and classification steps, first, the number of pixels with different class labels in each object is counted. Then, the set of pixels in each object is assigned to the most frequent class label (coming from the classification step) in the object. Thus, each region from the segmentation map is considered as an adaptive homogeneous neighborhood for all the pixels within this region. The described technique leads to a considerable improvement in terms of classification accuracies. In addition, MV provides more homogeneous classification maps in comparison with classification methods which use local neighborhoods in order to take into account spatial information in a classifier [27]. For better understanding, the work flow of MV is given below:

1) The output of SVM (Classification Map (CM)) and HMRF-EM (Segmentation Map (SM)) are considered as the inputs for MV. SM consists of several object (in Fig. 3 we have 3 different objects 1, 2 and 3) and CM consists of different classes (in Fig. 3 we have 3 different classes blue, grey and white).
2) In each object, all the pixels are assigned to the most frequent class within this object.

III. EXPERIMENTAL RESULTS

Two hyperspectral data sets were used in experiments. They are described below.

A. Data description

1) Indian Pines data: The first data set is the well-known AVIRIS data set captured on NW Indian Pines in 1992 presenting 16 classes, mostly related to land covers. The data set consists of 145 by 145 pixels with a spatial resolution of 20 m/pixel. In this paper, we used 200 data channels, i.e., after the elimination of the bands affected by atmosphere absorption. The number of training and test samples are displayed in Table I. Fig. 4. a), b) and c) illustrate one band of Indian Pines and its corresponding training and test sets, respectively.
2) Salinas data: This data set was captured by AVIRIS over Salinas Valley, California, and is characterized by high spatial resolution (3.7-meter pixels) consisting of 512 by 217 samples. The original data set consists of 224 data channels but here 20 water absorption bands are discarded. It includes vegetations, bare soils, and vineyard fields. The Salinas reference data contains 16 classes. Fig. 5. a) and b) show the Salinas data set and its corresponding reference map.

B. General description

For the gradient step, the input image is transformed by PCA and the first PCs with cumulative variance more than 99 percent are selected as the most effective components since they explain almost all of the variance in the data. Then, Sobel edge detection is performed on each component. Following that, the components are summed up and the resulting image is transformed to binary format in order to create the gradient image.

Then, both data sets are classified by K-means and 16 and 20 are selected as the number of classes. Those numbers are selected in such a fashion that the former is equal to the number of classes in a reference map and the latter superior to the minimum number, in order to compare the efficiency of different methods in terms of different number of clusters in K-means. Ten iterations are chosen for this step and the output of this step and the edge detected image are regularized by HMRF-EM for providing the spatial information.

In parallel, for extracting spectral information, the data sets are classified by SVM with a Gaussian kernel. The hyper

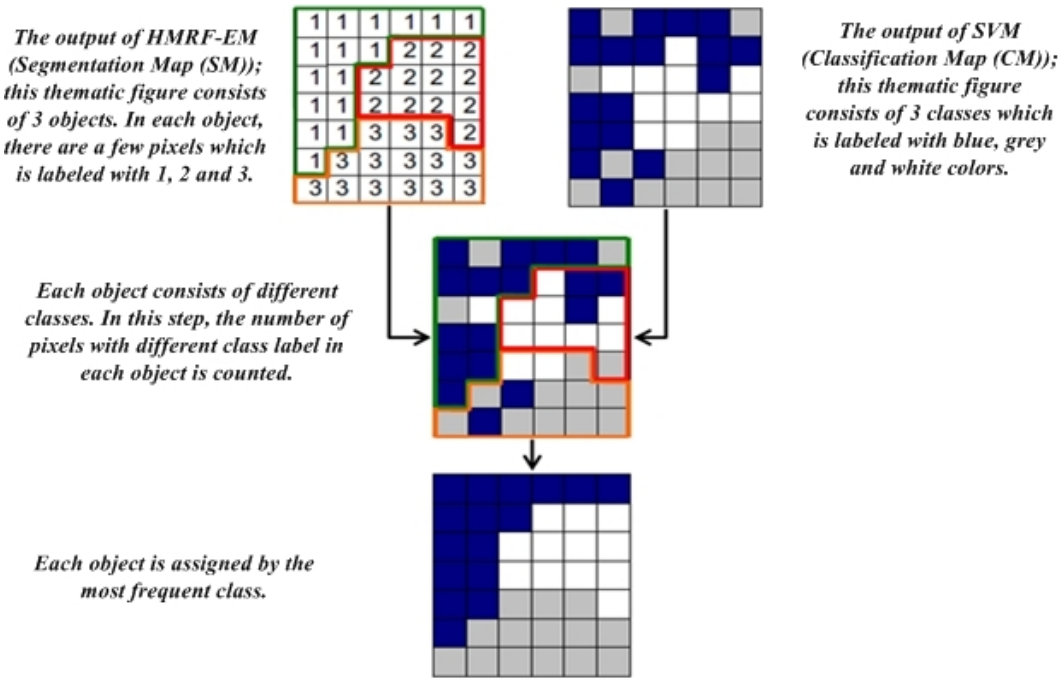

Fig. 3. The procedure of MV for combining the spectral and spatial information (based on [1]).

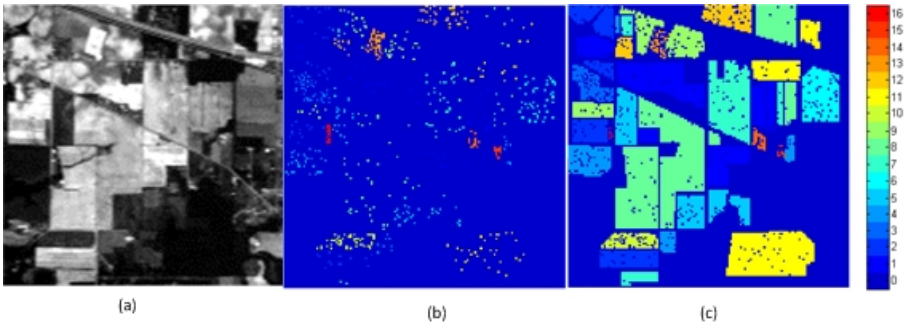

Fig. 4. An example of the Indian Pines test case. a) Data channels 27; b) training samples, c) test samples, each color represents a specific information class. The information classes are listed in Table I

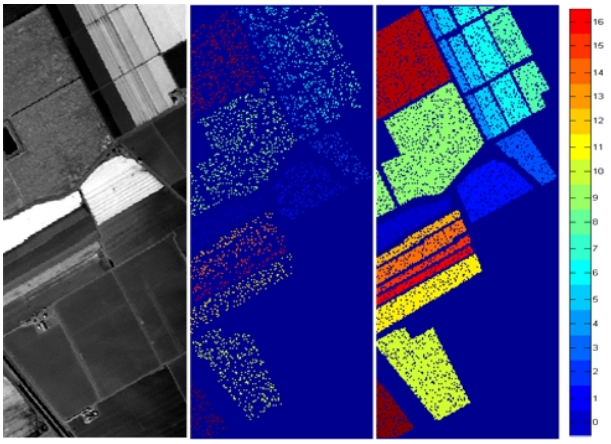

Fig. 5. An example of the Salinas test case. a) Data channels 57 b) training samples, c) test samples, each color represents a specific information class. The information classes are listed in Table III

tuning parameters are selected using 5-fold cross validation. To make the comparison as fair as possible, SVM is performed on each data set only once, and the classification map of this step, is directly used for other methods. In other words, the spectral part of all methods is the same and only the spatial part is changed for each method.

In the final step, the results of the spectral and spatial steps are combined using MV method and the output of this step is the final classification map.

In this paper, we use McNemar's test to assess our classification result. The aforementioned test is calculated as follows:

$$M = \frac{d_{12} - d_{21}}{\sqrt{d_{12} + d_{21}}} \tag{12}$$

where d_{12} is the number of pixels which are erroneously classified by the proposed method and not by the compared method, and d_{21} has a dual meaning [28]. The differences between the proposed method and others are statistically significant at 5-percent significant level if $|M| > 1.96$.

In this paper, SVM denotes the traditional SVM, HM-

RFSVM is the proposed method which is the combination of HMRF and SVM. HMRFSVM-E and HMRFSVM-NE are HMRFSVM with and without including the gradient step, respectively, and 16 and 20 depict the number of predefined clusters for K-means clustering. KmeansSVM denotes a combination of K-means and SVM by using MV.

C. Results

1) Indian Pines: For the classification of Indian Pines, all the available data channels are taken into consideration without performing feature reduction. It should be noted that all 16 classes were considered in order to evaluate the efficiency of different methods. The result of the classification for each class along with the overall accuracy and the Kappa coefficient are given in Table I. Fig. 6 shows the classification maps for SVM, KmeansSVM-16, HMRFSVM-NE-16, HMRFSVM-E-16, KmeansSVM-20, HMRFSVM-NE-20 and HMRFSVM-E-20, respectively.

The low spatial resolution of this data set adds more complexity, since it leads to the presence of the highly mixed pixels. In this case, the unsupervised clustering (or/and clustering based segmentation) might be degraded by spectrally mixed pixels in the image. In addition, the significant differences in the number of pixels in the reference data for different classes add more complexities on the data set and make the classification and the segmentation tasks more complicated [29].

As can be seen from Table I, the overall accuracy and Kappa coefficient increase when the number of clusters increases from 16 to 20. For instance, the overall accuracy of KmeansSVM, HMRFSVM-NE and HMRFSVM-E are improved by almost 1.9, 1.1 and 2.8 percent, respectively, when the number of clusters increases from 16 to 20. The main reason behind is undersegmentation which occurs when the number of predefined clusters is not sufficient. In this case several regions are detected as one and merged together which is not desired. This issue is easily solved by increasing the number of predefined clusters in the K-means.

Results confirm that the spectral and spatial classification approach, using majority voting is able to improve the pixel-wise classification accuracy considerably, in particular for the classification of large spatial structures in the data set. This fact helps to reduce the noisy behavior of the pixel-wise classification significantly. However, for small structures, when the spatial information from adjacent neighbors are taken into account, the small structures are in danger of disappearing and merging with bigger structures. Accurate segmentation can improve the spatial part of the spectral and spatial classification techniques and help to overcome the above-mentioned problem.

Due to the fact that the data set contains large spatial structures and the reference data does not comprise region edges, the advantage of considering the gradient step for HMRFSVM-E compared to HMRFSVM-NE is not obvious. With reference to Table I, HMRFSVM-E-16 improves SVM and KmeansSVM by 5.1 and 1.3 percent, respectively. In the same way, HMRFSVM-E-20 increases the overall accuracy of the classification of SVM and KmeansSVM by 8.2 and 2.2 percent, respectively.

Table II shows the results from McNemar's test. As can be seen from the table, the differences in classification accuracy between the proposed method and others are statistically significant using 5 percent level of significance. In this case HMRFSVM-20 is statistically different from SVM, KmeansSVM-20 and HMRFSVM-NE-20 by almost 22.15, 7.11 and 4.71 respectively.

2) Salinas: Table III shows the classification accuracies for the approaches applied to the Salinas data. As can be seen from the table, HMRFSVM-E gives the best performance in terms of classification accuracies when compared with the other methods. For 16 clusters, HMRFSVM-E-16 improves the classification accuracies of KmeansSVM-16 and SVM by 5.7 and 2.7 percent, respectively. In the same way, when the number of clusters was selected as 20, the proposed method showed improvement over all studied methods. Results confirm that considering majority voting helps different methods to decrease the noisy behavior of the traditional SVM. The main assumption behind HMRF is that in a predefined neighborhood structure, any given pixel is more likely to be allocated to a given cluster type if its neighboring pixels are also of that type. Therefore, it is easy to conclude that HMRF can be effective for images containing big structures.

KmeansSVM-16 shows the worst performance in terms of classification accuracies when compared to other methods. The main reasons for the bad performance of KmeansSVM-16 might be: 1) the spectral signature of Grapes-untrained and Vinyard-untrained are close to each other; in particular, considering only 16 clusters leads to a merging of the clusters which have a close spectral response, 2) KmeansSVM-16 does not consider spatial dependencies of the image and clustering is done by only considering the spectral information. In other words, since spatial dependencies are not taken into account and the number of predefined clusters is not enough, majority voting is not able to determine the correct class within each segment.

As can be seen from Table IV, the differences between the proposed method which considers edges and others were significantly different when the Salinas data were clustered with 16 and 20 clusters.

D. Comparison of the proposed method with the state-of-the-art

In this section the proposed method is compared with some recent approaches in terms of classification accuracy in order to provide a brief vision regarding the capability of HMRFSVM-E. Since Indian Pines is considered as one of best known data sets which many researchers have tested their algorithms on, that data set is used here for comparison. Table V reports the overall accuracy and Kappa coefficient for the-state-of-the-art. In the following we only analyze methods which have shown better classification accuracies than the proposed approach. The methods with better results than the proposed approach are shown in bold. For better understanding of the methods used for comparison, we refer readers to the

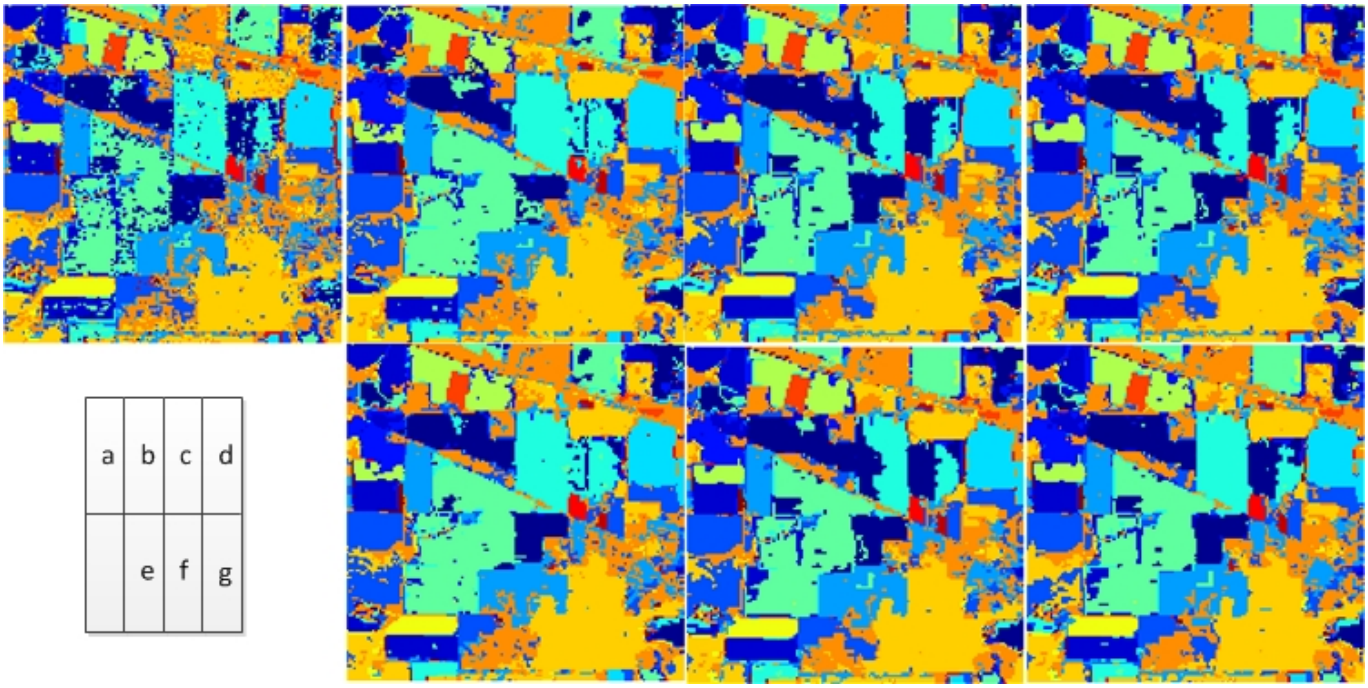

Fig. 6. Classification maps of different methods for Indian Pines: a) SVM, b) KmeansSVM-16, c) HMRFSVM-NE-16, d) HMRFSVM-E-16, e) KmeansSVM-20, f) HMRFSVM-NE-20 and, g) HMRFSVM-E-20.

TABLE I
INDIAN PINES: THE NUMBER OF TRAINING AND TEST SAMPLES; CLASSIFICATION ACCURACIES OF TEST SAMPLES IN PERCENTAGE FOR SVM, KMEANSSVM-16, HMRFSVM-NE-16, HMRFSVM-E-16, KMEANSSVM-20, HMRFSVM-NE-20 AND HMRFSVM-E-20.

No.	Class Name	No. of Samples Training	Test	SVM	KmeansSVM 16	KmeansSVM 20	HMRFSVM-NE 16	HMRFSVM-NE 20	HMRFSVM-E 16	HMRFSVM-E 20
1	Corn-notill	50	1384	79.1	63.7	73.0	**89.5**	88.3	85.5	89.4
2	Corn-mintill	50	784	83.4	89.4	93.3	96.1	96.0	**96.3**	95.9
3	Corn	50	184	92.9	97.2	**96.7**	92.9	94.0	96.2	94.5
4	Grass-pasture	50	447	**96.6**	95.7	95.9	95.5	93.9	96.2	95.1
5	Grass-trees	50	697	91.9	90.5	92.8	93.1	**93.8**	93.1	**93.8**
6	Hay-windrowed	50	439	96.8	**99.3**	98.4	98.4	98.6	98.4	97.7
7	Soybean-notill	50	918	84.6	**91.9**	91.2	72.6	89.4	66.9	90.8
8	Soybean-mintill	50	2418	69.1	**83.7**	83.5	81.5	80.6	82.3	82.4
9	Soybean-clean	50	564	87.2	83.5	85.6	89.1	88.6	**89.3**	88.3
10	Wheat	50	162	**99.3**	**99.3**	**99.3**	**99.3**	**99.3**	**99.3**	98.7
11	Woods	50	1244	88.5	96.1	**97.1**	96.4	96.3	96.5	96.3
12	Bldg-Grass-Tree-Drives	50	330	81.2	**93.9**	93.0	92.4	90.9	93.3	90.9
13	Stone-Steel-Towers	50	45	97.7	**100**	**100**	**100**	**100**	**100**	**100**
14	Alfalfa	15	39	89.7	76.9	92.3	**94.8**	**94.8**	89.7	**94.8**
15	Grass-pasture-mowed	15	11	90.9	**100**	90.9	90.9	90.9	90.9	90.9
16	Oats	15	5	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	Total Overall Accuracy	–	–	82.56	86.38	88.34	88.69	89.78	87.74	**90.50**
	Kappa Coefficient	–	–	0.8019	0.8446	0.8672	0.8709	0.8836	0.8601	**0.8917**

TABLE II
INDIAN PINES: THE RESULT OF MCNEMAR'S TEST TO VALIDATE WHETHER THE DIFFERENCE BETWEEN CLASSIFICATION ACCURACIES OF THE PROPOSED METHOD WITH BOTH PREDEFINED 16 AND 20 CLUSTERS IS SIGNIFICANTLY DIFFERENT FROM OTHER METHODS.

Indian Pines	M
HMRFSVM-E_16 vs. SVM	14.06
HMRFSVM-E_16 vs. KmeansSVM_16	4.24
HMRFSVM-E_16 vs. HMRFSVM-NE_16	4.41
HMRFSVM-E_20 vs. SVM	22.15
HMRFSVM-E_20 vs. KmeansSVM_20	7.11
HMRFSVM-E_20 vs. HMRFSVM-NE_20	4.71

references which can be found in front of each method in Table V.

As can be seen from Table V, the proposed method has an acceptable result in comparison with the other methods. Below, the proposed method is compared in more detail to SVMMRF-E [13], SVMMSF+MV [30] and MSSC-MSF [27].

1) HMRFSVM-E vs. SVMMRF-E: In SVMMRF-E the input data set is at first classified by a probabilistic SVM and then regularized by MRF using a gradient step. The most important disadvantage of SVMMRF-E is that the parameter β must be carefully set but that parameter controls the importance of the spatial energy terms versus the spectral

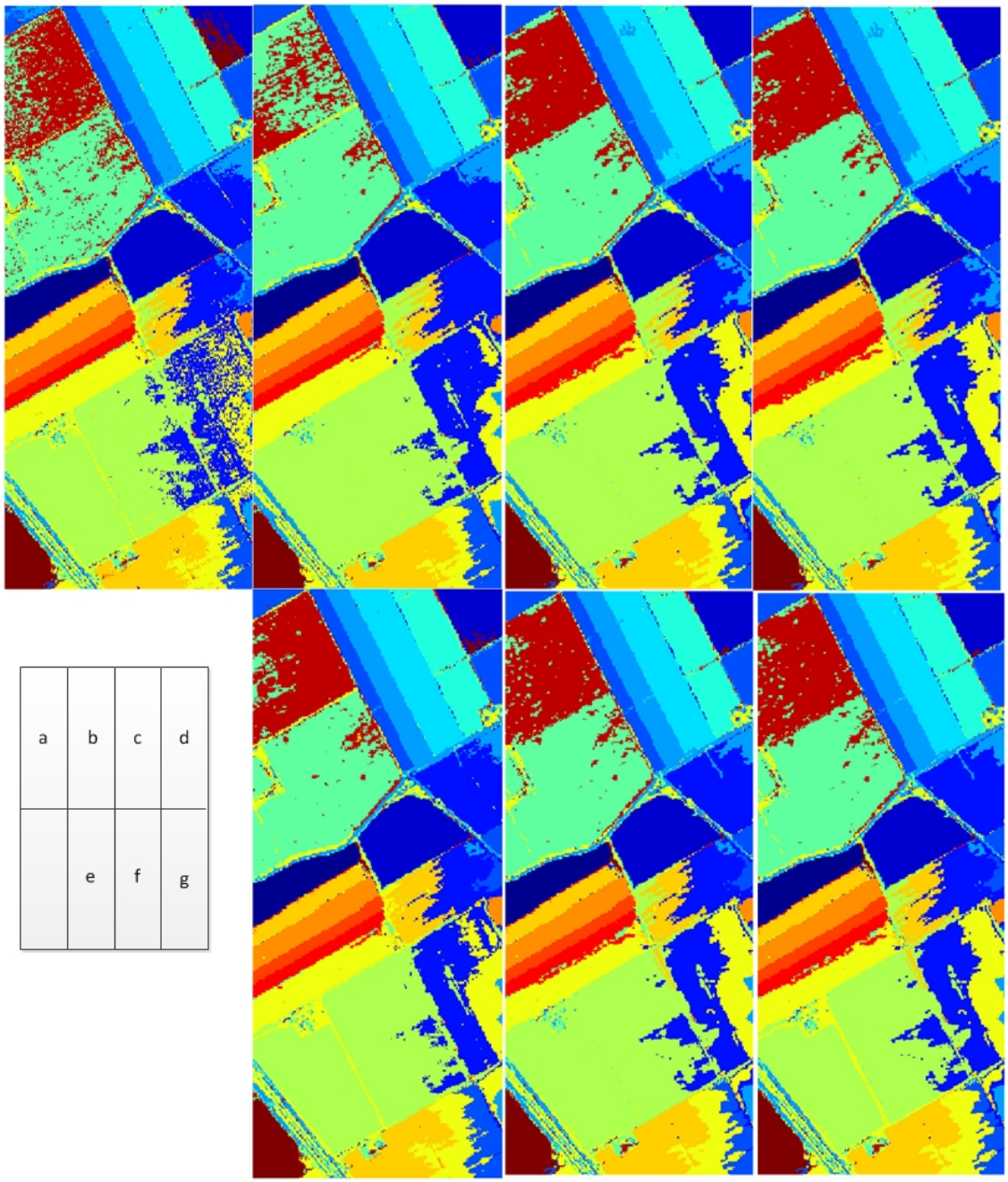

Fig. 7. Classification maps of different methods for Salinas: a) SVM, b) KmeansSVM-16, c) HMRFSVM-NE-16, d) HMRFSVM-E-16, e) KmeansSVM-20, f) HMRFSVM-NE-20 and, g) HMRFSVM-E-20.

energy term. With reference to [13], different values of β can considerably change the result of the classification and that poses a problem for this approach. In contrast, the method proposed in this paper is fully automatic, i.e., there is no need to initialize the parameters in order to achieve good results.

2) HMRFSVM-E vs. SVMMSF+MV: SVMMSF+MV was proposed in [30]. In this method, the original data set is initially classified by using a probabilistic pixel-wise classification technique. The output of this step provides both a classification map and a probability map. The outputs of the first step helps one to select the most reliably classified pixels. For providing a map of *markers* the classification and probability maps are considered to provide a Connected Components (CCs) labeling of the classification map. Then, for each CC, the region is compared to a threshold, M, in order to define whether the region is considered as being large or small. The M parameter is initialized by considering the resolution of the image along with typical sizes of the objects of interest. If the region is considered as small, the marker is the same with pixels of CC with probabilities more than S percent. The S parameter is set by considering the probability of the presence of small structures in the image (which also depends on the image resolution and the classes of interests). If the region is considered as large, the marker is P (defining

TABLE III
SALINAS: THE NUMBER OF TRAINING AND TEST SAMPLES; CLASSIFICATION ACCURACIES OF TEST SAMPLES IN PERCENTAGE FOR SVM, KMEANSSVM-16, HMRFSVM-NE-16, HMRFSVM-E-16, KMEANSSVM-20, HMRFSVM-NE-20 AND HMRFSVM-E-20.

Class		No. of Samples		SVM	KmeansSVM		HMRFSVM-NE		HMRFSVM-E	
No.	Name	Training	Test		16	20	16	20	16	20
1	Brocoli_green_weeds_1	252	1757	99.5	**100**	**100**	**100**	**100**	**100**	**100**
2	Brocoli_green_weeds_2	474	3252	**100**	**100**	**100**	**100**	**100**	**100**	**100**
3	Fallow	239	1737	99.3	**99.7**	99.5	**99.7**	**99.7**	**99.7**	**99.7**
4	Fallow_rough_plow	169	1225	99.2	**99.8**	**99.8**	**99.8**	**99.8**	**99.8**	**99.8**
5	Fallow_smooth	342	2336	**99.4**	98.5	98.8	85.3	98.9	86.8	99.1
6	Stubble	516	3443	**99.9**	99.7	**99.9**	98.5	99.6	98.6	99.8
7	Celery	442	3137	99.4	99.4	99.4	99.4	99.4	99.4	**99.5**
8	Grapes_untrained	1395	9876	88.4	95.8	94.3	95.6	95.0	**96.2**	95.1
9	Soil_vinyard_develop	775	5428	**99.9**	**99.9**	**99.9**	**99.9**	99.7	**99.9**	99.6
10	Corn_senesced_green_weeds	407	2871	**97.2**	96.7	95.7	90.8	90.8	90.6	91.9
11	Lettuce_romaine_4wk	141	927	98.7	**99.0**	98.9	**99.0**	**99.0**	**99.0**	**99.0**
12	Lettuce_romaine_5wk	232	1695	99.8	99.8	99.8	**100**	99.8	**100**	**100**
13	Lettuce_romaine_6wk	124	792	**99.4**	98.9	98.8	98.6	98.8	98.8	99.0
14	Lettuce_romaine_7wk	121	949	95.4	95.4	95.8	**97.2**	96.0	97.1	96.3
15	Vinyard_untrained	906	6362	76.9	43.2	90.0	**92.9**	92.6	92.8	92.8
16	Vinyard_vertical_trellis	231	1576	98.9	98.9	98.9	**99.3**	99.0	99.1	99.2
	Total Overall Accuracy	–	–	94.02	91.01	96.93	96.57	97.10	96.76	**97.24**
	Kappa Coefficient	–	–	0.9334	0.8993	0.9658	0.9618	0.9677	0.9639	**0.9692**

TABLE IV
SALINAS: THE RESULT OF MCNEMAR'S TEST TO VALIDATE WHETHER THE DIFFERENCE BETWEEN CLASSIFICATION ACCURACIES OF THE PROPOSED METHOD WITH BOTH PREDEFINED 16 AND 20 CLUSTERS IS SIGNIFICANTLY DIFFERENT FROM OTHER METHODS.

Salinas	M
HMRFSVM-E_16 vs. SVM	24.04
HMRFSVM-E_16 vs. KmeansSVM_16	41.74
HMRFSVM-E_16 vs. HMRFSVM-NE_16	5.41
HMRFSVM-E_20 vs. SVM	30.75
HMRFSVM-E_20 vs. KmeansSVM_20	3.83
HMRFSVM-E_20 vs. HMRFSVM-NE_20	3.66

TABLE V
INDIAN PINES: COMPARISON WITH THE-STATE-OF-THE-ART. THE METHODS WITH HIGHER ACCURACIES THAN THE PROPOSED APPROACH ARE SHOWN IN BOLD FACE

Method		Overall Accuracy	Kappa Coefficient
HMRFSVM-E		90.50	0.892
WH+MV [30]		89.63	0.848
EM+MV [27]		83.60	0.848
SVMMRF-E [13]		**91.83**	**0.907**
SVMMSF+MV [30]		**91.80**	**0.906**
MC-MSF [27]		86.66	0.848
MSSC-MSF [27]		**92.3**	**0.911**
M-HSEG^r [31]	SAM	77.53	0.744
S_{wght} = 0.0	Inf	76.63	0.734
M-HSEG^p [31]	SAM	81.59	0.791
S_{wght} = 0.0	Inf	81.16	0.786
M-HSEG^{op} [31]	SAM	89.23	0.877
S_{wght} = 0.0	Inf	89.00	0.874
M-HSEG^{op} [31]	SAM	88.72	0.871
S_{wght} = 0.2	Inf	89.01	0.874

the percentage of pixels within the large region to be used as markers) percent of its pixels with the highest probabilities. The output of this step is a Map of Markers. Furthermore, the result of the previous step leads to the construction of a minimum spanning forest. Finally, majority voting within the connected components provides the final segmentation and classification map. From the above description, it can be observed that the method is not automatic. In addition, in order to apply this method successfully, a comprehensive knowledge regarding the different structures of the input data is needed.

3) HMRFSVM-E vs. MSSC-MSF: The MSSC-MSF was introduced in [27]. In this method, the input image is at first classified by a pixel-wise SVM. Second, The input image is segmented with Watershed Segmentation and the result combined with an SVM using majority voting (MV). Third, the input data is segmented by EM and combined with SVM through MV. Then, the input data set is segmented with Recursive divide-and-conquer approximation of HSEG (RHSEG) and combined with SVM by using MV. Furthermore, the output of the three steps are used for marker selection. The output of this step is then used for the construction of a minimum spanning forest. Based on the above work flow, it is easy to see that MSSC-MSF is quite complicated and can be become computationally very demanding without parallel processing.

IV. CONCLUSION

In this paper a fully automated framework which takes into account both spectral and spatial information is introduced for classification of hyperspectral images. In the framework, SVM is used for the extraction of spectral information. In parallel, HMRF-EM is used for the extraction of spatial information. In the final step, those results are combined by using majority voting. The efficiency of the proposed method is tested in both situations with and without considering the gradient step. The proposed method is evaluated on two data sets (Indian Pines and Salinas). In both cases the new approach outperforms other studied methods. The classification of the proposed method works better than SVM in terms of accuracies and improves the results of overall accuracy by almost 8 and 3.2 percent for Indian Pines and Salinas, respectively. It should be noted that the concept of HMRF is used for the first time in the field of

remote sensing in this paper and the efficiency of that for the segmentation of hyperspectral images is demonstrated. Finally, it is shown in the paper that the method performs well in terms of accuracies compared with the state of the art. In addition, the proposed approach is fully automatic and user-friendly in contrast to most of the methods.

V. ACKNOWLEDGMENT

The authors would like to thank Prof. Landgrebe from Purdue University for providing the Indian pines data set. This research was supported in part by the Icelandic Research Fund for Graduate Students.

REFERENCES

[1] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.

[2] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT, no. 14, pp. 55 – 63, 1968.

[3] S. Tadjudin and D. Landgrebe, "Classification of high dimensional data with limited training samples," in *Tech. Rep., School of Electrical and Computer Engineering, Purdue University*, 1998.

[4] V. N. Vapnic, *Statistical Learning Theory*. Hoboken, NJ: Wiley, 1998.

[5] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis.* Hoboken, NJ: Wiley, 2009.

[6] P. Mantero, G. Moser, , and S. B. Serpico, "Partially supervised classification of remote sensing images using SVM-based probability density estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 559–570, 2005.

[7] G. Moser and S. B. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. PP, no. 99, pp. 1–19, 2012, DOI: 10.1109/TGRS.2012.2211882.

[8] 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, 2001.

[9] 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, 2005.

[10] H. Derin and P. A. Kelly, "Discrete-index markov-type random processes," *Proceedings of the IEEE*, vol. 77, no. 10, pp. 1485–1510, 1989.

[11] G. Moser, S. B. Serpico, and J. A. Benediktsson, "Land-cover mapping by Markov modeling of spatial-contextual information in very-high-resolution remote sensing images," *Proceedings of the IEEE*, vol. 101, pp. 631 – 651, March 2013.

[12] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov Random Field model and the Expectation-Maximization algorithm," *IEEE Trans. On Medical Imaging*, vol. 20, no. 1, pp. 45 – 57, 2001.

[13] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, p. 736 740, 2010.

[14] F. Bovolo and L. Bruzzone, "A context-sensitive technique based on support vector machines for image classification," *Proc. PReMI*, pp. 260–265, 2005.

[15] D. Liu, M. Kelly, and P. Gong, "A spatial-temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery," *Remote Sens. Environment*, vol. 101, no. 10, p. 167180, 2006.

[16] M. Khodadadzadeh, R. Rajabi, and H. Ghassemian, "Combination of region-based and pixel-based hyperspectral image classification using erosion technique and MRF model," *Electrical Engineering (ICEE), 2010 18th Iranian Conference on*, pp. 294 –299, May 2010.

[17] G. Zhang and X. Jia, "Simplified conditional random fields with class boundary constraint for spectral-spatial based remote sensing image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 5, 2012.

[18] B. Tso and R. C. Olsen, "Combining spectral and spatial information into hidden markov models for unsupervised image classification," *International Journal of Remote Sensing*, vol. 26, no. 10, pp. 2113–2133, 2005.

[19] G. Hazel, "Multivariate gaussian mrf for multispectral scene segmentation and anomaly detection," *IEEE Trans. on Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1199–1211, May 2000.

[20] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-6, no. 6, pp. 721 –741, nov. 1984.

[21] A. P. Dempster, N. M. Laird, and D. B. Bubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Roy. Stat. Soc*, vol. 39, no. 1, p. 138, 1977.

[22] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

[23] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontiniuties," *IEEE. Trans. Patt. Anal. Machine Intell.*, vol. 14, no. 3, pp. 367 – 383, 1992.

[24] V. N. Vapnik, *Statistical learning theory.* New York: Wiley, 1998.

[25] B. Scholkopf and A. J. Smola, *Learning with Kernels.* MIT Press, 2002.

[26] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classication of hyperspectral remote-sensing data over urban areas," *EURASIP Journal on Advances in Signal Processing*, pp. 1–14, 2009.

[27] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. on Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, 2010.

[28] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *PERS*, vol. 70, no. 5, pp. 627–633, 2004.

[29] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques,," *IEEE Trans. on Geos. and Remote Sens.*, vol. 47, no. 5, pp. 2973–2987, 2009.

[30] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, pp. 1267–1279, 2010.

[31] Y. Tarabalka, J. C. Tilton, J. A. Benediktsson, and J. Chanussot, "A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,*, vol. 5, no. 1, pp. 262–272, 2012.

Pedram Ghamisi graduated with a B.Sc. degree in Civil (Survey) Engineering from the Tehran South Campus of Azad University . Then, he obtained the M.Sc. degree in Remote Sensing at K.N.Toosi University of Technology in 2012. He received the Best Researcher Award for M.Sc. students in K. N. Toosi University of Technology in the academic year 2010-2011. He is currently a Ph.D. student in Electrical and Computer Engineering at the University of Iceland. His research interests are in remote sensing and image analysis with the current focus on spectral and spatial techniques for hyperspectral image classification. He serves as a reviewer for a number of journals including IEEE Trans. Image Processing, IEEE JSTARS and IEEE GRSL.

Jón Atli Benediktsson Jn Atli Benediktsson received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, 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 Pro Rector for Academic Affairs and Professor of Electrical and Computer Engineering at 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 published extensively in those fields. Prof. Benediktsson was the 2011-2012 President of the IEEE Geoscience and and Remote Sensing Society (GRSS) and has been on the GRSS AdCom since 2000. He was Editor of the IEEE Transactions on Geoscience and Remote Sensing (TGRS) from 2003 to 2008 and has served as Associate Editor of TGRS since 1999, the IEEE Geoscience and Remote Sensing Letters since 2003 and IEEE Access since 2013. He is on the International Editorial Board of the International Journal of Image and Data Fusion and was the Chairman of the Steering Committee of IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (J-STARS) 2007-2010. Prof. Benediktsson is a co-founder of the biomedical start up company Oxymap (www.oxymap.com). He is a Fellow of the IEEE and a Fellow of SPIE. He received the Stevan J. Kristof Award from Purdue University in 1991 as outstanding graduate student in remote sensing. In 1997, Dr. Benediktsson was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award, in 2000, he was granted the IEEE Third Millennium Medal, in 2004, he was a co-recipient of the University of Iceland's Technology Innovation Award, in 2006 he received the yearly research award from the Engineering Research Institute of the University of Iceland, and in 2007, he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society. He is co-recipient of the 2012 IEEE Transactions on Geoscience and Remote Sensing Paper Award. He is a member of the Association of Chartered Engineers in Iceland (VFI), Societas Scinetiarum Islandica and Tau Beta Pi.

Magnus Orn Ulfarsson received the B.S. and the M.S. degrees from the University of Iceland in 2002, and the Ph.D. degree from the University of Michigan in 2007. He joined University of Iceland in 2007 where he is currently an associate professor. His research interest include statistical signal processing, image processing, remote sensing, genomics and functional magnetic resonance imaging.