Open access • Journal Article • DOI:10.1109/LGRS.2010.2047711

SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images

Yuliya Tarabalka, Mathieu Fauvel, Jocelyn Chanussot, Jon Atli Benediktsson

Institutions: Grenoble Institute of Technology, French Institute for Research in Computer Science and Automation, University of Iceland

Published on: 18 May 2010 - IEEE Geoscience and Remote Sensing Letters (IEEE)

Topics: Contextual image classification, Hyperspectral imaging, Support vector machine and Markov random field

SVM and MRF-Based Method for Accurate Classification of Hyperspectral Images

Yuliya Tarabalka, Mathieu Fauvel, Jocelyn Chanussot, Jon Atli Benediktsson

SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images

Yuliya Tarabalka, *Student Member, IEEE*, Mathieu Fauvel, Jocelyn Chanussot, *Senior Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—The high number of spectral bands acquired by hyperspectral sensors increases the capability to distinguish physical materials and objects, presenting new challenges to image analysis and classification. This letter presents a novel method for accurate spectral-spatial classification of hyperspectral images. The proposed technique consists of two steps. In the first step, a probabilistic support vector machine pixelwise classification of the hyperspectral image is applied. In the second step, spatial contextual information is used for refining the classification results obtained in the first step. This is achieved by means of a Markov random field regularization. Experimental results are presented for three hyperspectral airborne images and compared with those obtained by recently proposed advanced spectral-spatial classification techniques. The proposed method improves classification accuracies when compared to other classification approaches.**

***Index Terms*—Classification, hyperspectral images, Markov random field (MRF), support vector machine (SVM).**

I. Introduction

HYPERSPECTRAL imaging sensors measure the energy of the received light in tens or hundreds of narrow spectral bands in each spatial position in the image [1]. Thus, every pixel can be represented as a high-dimensional vector across the wavelength dimension, called the spectrum of the material in this pixel. Since different substances exhibit different spectral signatures, hyperspectral imagery is a well-suited technology for accurate image classification, which is an important task in many application domains (monitoring and management of the environment, precision agriculture, etc.).

Most classification methods process each pixel independently without considering the correlations between spatially adjacent pixels (so-called *pixelwise* classifiers) [2], [3]. In particular, support vector machines (SVMs) have shown good performances for classifying high-dimensional data when a limited number of training samples are available [3], [4]. Furthermore, spatial contextual information should help for an accurate scene interpretation. Therefore, it is very important to develop *spectral-spatial* classification techniques that are capable to consider spatial dependences between pixels [5]–[8].

In general, two categories of spectral-spatial classification methods can be distinguished. First, spatial contextual information is exploited in the classification stage. For instance, spectral and spatial information can be combined within a feature vector of each pixel, and then, a pixelwise classification technique can be applied to the obtained set of vectors [6], [9]. Another group of methods from this category first defines the objects within the image scene and then classifies each object [2], [5]. Second, spatial dependences are considered in the decision rule [10]. An example is a pixelwise classification followed by spatial regularization of the classification map.

Markov random fields (MRFs) are probabilistic models that are commonly used to integrate spatial context into image classification problems [7], [10], [11]. In the MRF framework, the *maximum a posteriori* (MAP) decision rule is typically formulated as the minimization of a suitable energy function [12]. An extensive literature is available on MRF-based image classification techniques. In particular, the research groups of Farag [7], Bruzzone [10], and Gong [11] have investigated the integration of the SVM technique within an MRF framework for accurate spectral-spatial classification of remote sensing images. All of them use SVMs to estimate class conditional probability density functions and MRFs to estimate context-based class priors. Farag *et al.* [7] have applied the mean field-based SVM regression algorithm for density estimation, with the purpose of hyperspectral image classification. Good classification results are reported, although no comparison with other advanced spectral-spatial classification techniques is published.

This letter presents a novel SVM- and MRF-based (SVMMRF) method for spectral-spatial classification of hyperspectral images. In the first step of the proposed method, a probabilistic SVM pixelwise classification of the hyperspectral image is applied. In the second step, spatial contextual information is used for refining the classification results obtained in the first step. This is achieved by means of the MRF regularization. An important difference from previously proposed methods [7], [10], [11] consists in defining and integrating the "fuzzy no-edge/edge" function into the spatial energy function involved in MRFs, aiming at preserving edges while performing spatial regularization.

Manuscript received January 25, 2010; revised March 16, 2010. Date of publication May 18, 2010; date of current version October 13, 2010. This work was supported in part by the Marie Curie Research Training Network "HYPER-I-NET."

Y. Tarabalka is with the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA Lab), Grenoble Institute of Technology, 38402 Grenoble, France, and also with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: yuliya.tarabalka@hyperinet.eu).

M. Fauvel is with the Modelling and Inference of Complex and Structured Stochastic Systems (MISTIS) Team, National Institute for Research in Computer Science and Control (INRIA), 38334 Saint Ismier, France (e-mail: Mathieu.fauvel@inrialpes.fr).

J. Chanussot is with GIPSA Lab, Grenoble Institute of Technology, 38402 Grenoble, France (e-mail: jocelyn.chanussot@gipsa-lab.grenoble-inp.fr).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: benedikt@hi.is).

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/LGRS.2010.2047711

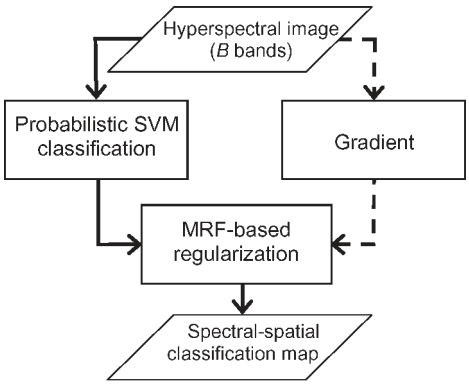

Fig. 1. Flowchart of the proposed SVMMRF classification scheme.

The second contribution of this letter consists in the experimental comparison of the presented approach with other recently proposed advanced spectral-spatial classification techniques. Experimental results are demonstrated on three hyperspectral airborne images recorded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and the Reflective Optics System Imaging Spectrometer (ROSIS).

The outline of this letter is as follows. In the next section, an SVMMRF classification scheme for hyperspectral images is presented. Experimental results are discussed in Section III. Finally, conclusions are drawn in Section IV.

II. SVMMRF Classification Scheme

The flowchart of the proposed SVMMRF classification method is shown in Fig. 1. At the input, a B-band hyperspectral image is given, which can be considered as a set of n pixel vectors $\mathbf{X} = \{\mathbf{x}_j \in \mathbb{R}^B, j = 1, 2, \ldots, n\}$. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$ be a set of information classes in the scene. Classification consists in assigning each pixel to one of the K classes of interest.

A. Probabilistic SVM Classification

The first step of the proposed procedure consists in performing a probabilistic SVM pixelwise classification of the hyperspectral image [4], [13]. Other probabilistic classifiers could be used. However, SVMs are extremely well suited to classify hyperspectral data [3]. The standard SVMs do not provide probability estimates for the individual classes. In order to get these estimates, pairwise coupling of binary probabilistic estimates is applied [13], [14].

B. Computation of the Gradient

Independent of the previous step, a one-band gradient of the hyperspectral image is computed, which is further used for defining the fuzzy no-edge/edge function. Approaches for defining a one-band gradient from the B-band image are analyzed in [15]. Here, we first compute horizontal, vertical, and two diagonal gradients (corresponding to the directions 0°, 90°, 45°, and 135°, respectively), using Sobel masks [16], where each of the gradients is computed as the sum of the gradients of every spectral channel. The resulting one-band gradient $\nabla(\mathbf{X}) = \{\rho_j \in \mathbb{R}, j = 1, 2, \ldots, n\}$ is found as the average of the four obtained directional gradients.

C. MRF-Based Regularization

In the final step, the regularization of the SVM classification map is performed, using the MAP-MRF framework. This framework is based on the interpixel class dependence assumption, which means that a pixel belonging to a class ω_i is likely to have neighboring pixels belonging to the same class. In our work, an eight-neighborhood is assumed (let $\mathbf{N}_i$ be the set of neighbors for a given pixel $\mathbf{x}_i$).

We adopt the Metropolis algorithm, based on stochastic relaxation and annealing, for computing the MAP estimate of the true classification map given the initial (pixelwise) classification map [17], [18]. The considered method is based on the Bayesian approach and aims at minimizing the global energy in the image, by iterative minimization of local energies (defined hereafter) associated with randomly chosen image sites, i.e., pixels.

Let $\mathbf{L} = \{L_j, j = 1, 2, \ldots, n\}$ be a generic set of information class labels for the image $\mathbf{X}$. We propose to compute the local energy of a given site associated with a pixel $\mathbf{x}_i$ as

$$U(\mathbf{x}_i) = U_{\text{spectral}}(\mathbf{x}_i) + U_{\text{spatial}}(\mathbf{x}_i) \tag{1}$$

where $U_{\text{spectral}}(\mathbf{x}_i)$ is the spectral energy function from the observed data and $U_{\text{spatial}}(\mathbf{x}_i)$ is the spatial energy term computed over the local neighborhood $\mathbf{N}_i$. We define the spectral energy term as

$$U_{\text{spectral}}(\mathbf{x}_i) = -\ln\{P(\mathbf{x}_i|L_i)\} \tag{2}$$

where $P(\mathbf{x}_i|L_i)$ is estimated by pairwise coupling of probability estimates from "one-versus-one" SVM outputs [11], [14].

For the spatial energy term, two different expressions are investigated. We first consider the standard spatial energy expression, used, for instance, in [10], which is computed as

$$U^{\text{NE}}_{\text{spatial}}(\mathbf{x}_i) = \sum_{\mathbf{x}_j \in \mathbf{N}_i} \beta(1 - \delta(L_i, L_j)) \tag{3}$$

where $\delta(\cdot, \cdot)$ is the Kronecker delta function ($\delta(a, b) = 1$ if $a = b$, and $\delta(a, b) = 0$ otherwise) and β is a parameter that controls the importance of the spatial versus spectral energy terms. The superscript "NE" means that no edge information is taken into account. The term $U^{\text{NE}}_{\text{spatial}}(\mathbf{x}_i)$ is proportional to the number of neighboring pixels of $\mathbf{x}_i$ assigned to one of the classes different from L_i. This spatial energy term is particularly suitable for the images with large spatial structures. However, if a small one-pixel object is present in the image, this model will favor assigning this pixel to the class of the surrounding objects.

In order to mitigate this drawback of the previous spatial term and to preserve small structures and edges in the classification map, we propose to integrate the edge information into the spatial energy function. The computation of an accurate edge map for hyperspectral images is a challenging task. For instance, it can be obtained by thresholding the gradient image $\{\rho_j \in \mathbb{R}, j = 1, 2, \ldots, n\}$. For this purpose, an appropriate threshold

must be chosen. Instead of computing the edge map, we propose to define the following "fuzzy no-edge/edge function":

$$\varepsilon(\mathbf{x}_j) = 1 - \frac{\rho_j}{\alpha + \rho_j} \tag{4}$$

where α is a parameter controlling the approximate edge threshold. From here, the following spatial energy function is proposed:

$$U^E_{\text{spatial}}(\mathbf{x}_i) = \sum_{\mathbf{x}_j \in \mathbf{N}_i} \beta\varepsilon(\mathbf{x}_j)(1 - \delta(L_i, L_j)). \tag{5}$$

The superscript "E" means that the edge information is taken into account. In the following, we thus refer to two different methods, namely, SVMMRF-NE and SVMMRF-E, when (3) and (5) are used for computing the spatial energy, respectively.

We briefly summarize the considered Metropolis algorithm for optimizing the energy function. In each iteration, an image site (i.e., a pixel $\mathbf{x}_i$) is randomly chosen. The local energy of the given site $U(\mathbf{x}_i)$ is computed by (1). Then, a new class label L_i^{new} is randomly selected for the site $\mathbf{x}_i$, and a new local energy $U^{\text{new}}(\mathbf{x}_i)$ is computed. If the variation of the energy $\Delta U = U^{\text{new}}(\mathbf{x}_i) - U(\mathbf{x}_i) < 0$, the new class label is assigned to $\mathbf{x}_i$: $L_i = L_i^{\text{new}}$. Otherwise, the new class assignment is accepted with the probability $p = \exp(-\Delta U/T)$. Here, T is a global control parameter called "temperature" [18]. The optimization begins at a high temperature, which is gradually lowered as the relaxation procedure proceeds. This procedure avoids converging to local *minima*.

III. Experimental Results and Discussion

We applied the proposed SVMMRF-NE and SVMMRF-E classification approaches to three hyperspectral airborne images described in the following:

1) The *Indian Pines* image is of a vegetation area that was recorded by the AVIRIS sensor. The image is of 145 by 145 pixels, with a spatial resolution of 20 m/pixel and 200 spectral channels. A three-band false color image and the reference data are shown in Fig. 2. Sixteen classes of interest are considered, which are detailed in Table II, with a number of training and test samples for each class. Training samples have been randomly chosen from the reference data.
2) The *Center of Pavia* image was recorded by the ROSIS sensor over the urban area of Pavia, Italy. It is of 900 by 300 pixels, with a spatial resolution of 1.3 m/pixel and 102 spectral channels. The reference data contain nine thematic classes and 56 070 labeled pixels. Thirty samples for each class were randomly chosen from the reference data as training samples.
3) The *University of Pavia* image is of an urban area, acquired by the ROSIS sensor. It is of 610 by 340 pixels, with 103 spectral channels. The reference data contain nine classes of interest. The training and test sets are composed of 3921 and 40 002 pixels, respectively.

More information about the images can be found in [8].

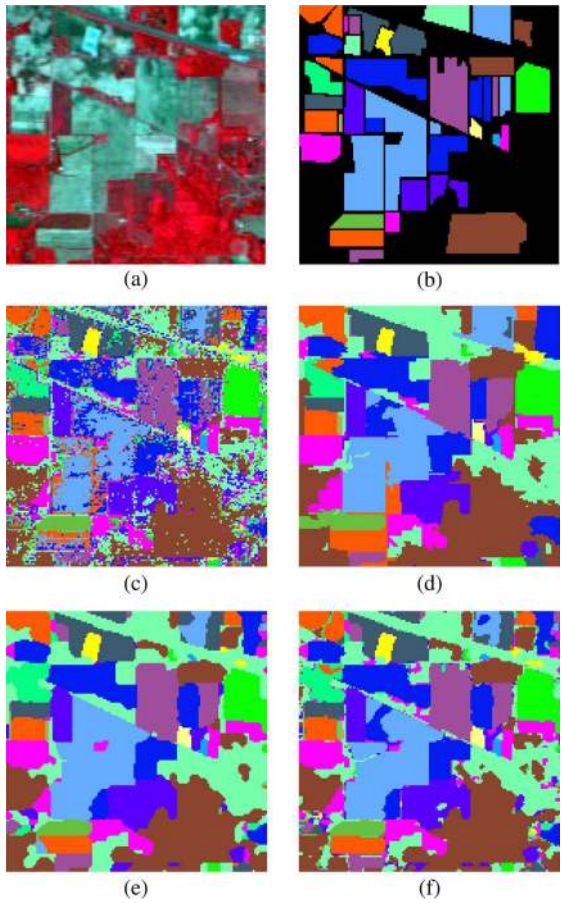

Fig. 2. *Indian Pines* image. (a) Three-band color composite. (b) Reference data. (c) SVM pixelwise classification map. (d) SVMMSF + MV classification map. (e) SVMMRF-NE classification map. (f) SVMMRF-E classification map.

In all experiments, the probabilistic one-versus-one SVM classification with the Gaussian radial basis function (RBF) kernel was applied. The optimal parameters C (parameter that controls the amount of penalty during the SVM optimization [4]) and γ (spread of the RBF kernel) were chosen by fivefold cross validation. The temperature T was varied during the Metropolis relaxation procedure [18]: The initial temperature was set to $T^1 = 2$ (a relatively low value of the initial temperature results in a faster execution of the algorithm). After every 10^6 (order of the number of pixels in an image) iterations, the temperature for the next iteration $(k+1)$ was recomputed as $T^{k+1} = 0.98T^k$. The optimal value of the parameter $\alpha = 30$ was experimentally derived (the same optimal value of α was obtained for the three considered data sets).

Furthermore, we have investigated the performances of the SVMMRF-NE and SVMMRF-E algorithms for different values of the context weight parameter β. Table I reports the SVMMRF-NE and SVMMRF-E overall (percentage of correctly classified pixels) and average (mean of the percentage of correctly classified pixels for each class) classification accuracies for the three considered data sets. It can be concluded that the optimal parameter is $\beta \in [1, 2]$ for the SVMMRF-NE approach and $\beta \in [2, 4]$ for the SVMMRF-E approach (for both methods, the corresponding overall accuracies are nonsignificantly different over the given range of values). Moreover, the

TABLE I
SVMMRF-NE AND SVMMRF-E CLASSIFICATION ACCURACIES FOR DIFFERENT VALUES OF THE PARAMETER β

Image	Accuracy, %	β for SVMMRF-NE					β for SVMMRF-E						
		0.5	1	2	3	5	0.5	1.0	2.0	3.0	4.0	5.0	10.0
Indian	Overall	90.42	92.05	92.31	92.53	**92.78**	84.99	88.19	**91.83**	91.59	90.64	90.03	90.23
Pines	Average	94.70	**95.83**	89.81	83.24	91.76	90.50	92.85	**95.69**	95.37	92.62	95.01	88.67
Center	Overall	97.26	**97.38**	97.26	97.31	97.27	97.11	97.46	97.56	**97.60**	97.50	97.44	97.33
of Pavia	Average	93.98	**94.13**	93.94	94.05	93.73	93.41	94.19	94.24	**94.59**	94.23	94.12	93.98
University	Overall	85.30	86.40	86.89	87.57	**86.93**	84.20	85.79	87.18	**87.63**	87.57	**87.63**	87.50
of Pavia	Average	91.64	91.72	**92.12**	91.90	91.75	91.03	92.22	92.97	**93.41**	93.27	93.28	93.39

TABLE II
NUMBER OF LABELED SAMPLES (NUMBER OF SAMPLES) AND CLASSIFICATION ACCURACIES IN PERCENTAGE FOR THE *INDIAN PINES* IMAGE

	No. of Samp.		SVM	ECHO	WH+MV	SVMMSF+MV	SVMMRF-NE	SVMMRF-E
	Train	Test						
Overall Accuracy	-	-	78.17	82.64	86.63	91.80	**92.05**	91.83
Average Accuracy	-	-	85.97	83.75	91.61	94.28	**95.83**	95.69
Kappa Coefficient (κ)	-	-	75.33	80.38	84.83	90.64	**90.93**	90.71
Corn-no till	50	1384	78.18	83.45	94.22	93.21	93.28	**98.48**
Corn-min till	50	784	69.64	75.13	78.06	**96.56**	83.93	90.82
Corn	50	184	91.85	92.39	88.59	95.65	**99.46**	98.37
Soybeans-no till	50	918	82.03	90.10	96.30	93.91	98.58	**98.91**
Soybeans-min till	50	2418	58.95	64.14	68.82	81.97	**82.09**	76.92
Soybeans-clean till	50	564	87.94	89.89	90.78	97.16	**97.70**	97.34
Alfalfa	15	39	74.36	48.72	94.87	94.87	**97.44**	**97.44**
Grass/pasture	50	447	92.17	94.18	95.08	94.63	**97.54**	**97.54**
Grass/trees	50	697	91.68	96.27	**97.99**	97.27	97.70	97.56
Grass/pasture-mowed	15	11	**100**	36.36	**100**	**100**	**100**	**100**
Hay-windrowed	50	439	97.72	97.72	99.54	**99.77**	99.54	99.54
Oats	15	5	**100**	**100**	**100**	**100**	**100**	**100**
Wheat	50	162	98.77	98.15	**99.38**	**99.38**	**99.38**	**99.38**
Woods	50	1244	93.01	94.21	97.11	**99.68**	98.39	99.04
Bldg-Grass-Tree-Drives	50	330	61.52	81.52	69.39	68.79	**88.18**	79.70
Stone-steel towers	50	45	97.78	97.78	95.56	95.56	**100**	**100**

methods are robust to the choice of β, and quite a wide range of values of β leads to high classification accuracies.[1]

Table II summarizes the global (overall average accuracies and kappa coefficient [8]) and class-specific classification accuracies for the *Indian Pines* image. In order to compare the performances of the proposed method with other recently proposed advanced classification techniques, we have included results of the pixelwise SVM classifiers, the well-known ECHO (Extraction and Classification of Homogeneous Object) spatial classifier [5], classification using majority vote within the adaptive neighborhoods defined by watershed segmentation (WH + MV) [19], as well as the results obtained using the construction of a minimum spanning forest from the probabilistic SVM-derived markers followed by majority voting within connected regions (SVMMSF + MV) [8]. Fig. 2 shows some of the corresponding classification maps. As can be seen from the table, all the spectral-spatial approaches yield higher classification accuracies when compared to the pixelwise method. The proposed SVMMRF-NE and SVMMRF-E techniques give the highest global and most of the best class-specific accuracies. Following the results of the McNemar's test, the SVMMRF-NE, SVMMRF-E, and SVMMSF + MV accuracies are not significantly different, using 5% level of significance. From Fig. 2, it can be seen that the corresponding three classification maps are comparable and contain more homogeneous regions, when compared to the SVM classification map. Since the considered image contains large spatial structures and reference data do not comprise region edges, the advantage of the SVMMRF-E method versus the SVMMRF-NE method is not obvious here.

[1]A similar study has shown robustness of the SVMMRF-E method to the choice of the parameter α.

TABLE III
GLOBAL CLASSIFICATION ACCURACIES IN PERCENTAGE FOR THE *CENTER OF PAVIA* IMAGE

	Overall Accuracy	Average Accuracy	Kappa Coefficient (κ)
SVM	95.64	90.60	92.71
ECHO	96.22	92.47	93.70
WH+MV	96.26	92.08	93.75
SVMMSF+MV	96.62	92.78	94.35
SVMMRF-NE	97.38	94.13	95.61
SVMMRF-E	**97.60**	**94.59**	**95.97**

Table III gives the global classification accuracies for the *Center of Pavia* data, where the same techniques are used for comparison. The proposed SVMMRF-E method yields the best classification accuracies. This image of an urban area contains small spatial structures, such as shadows and trees. Therefore, the inclusion of the edge information in the context-based regularization improves the classification performances.

Table IV reports the global classification accuracies for the *University of Pavia* image. For this data set, the SVMMSF + MV classifier gives the best accuracies, and the SVMMRF-E method outperforms the SVMMRF-NE technique in terms of accuracies. According to the results of the McNemar's test, all the corresponding classification maps are significantly different, using 5% level of significance. From these results, the following conclusions can be derived: 1) the advantage of the

TABLE IV
GLOBAL CLASSIFICATION ACCURACIES IN PERCENTAGE FOR THE *UNIVERSITY OF PAVIA* IMAGE

	Overall Accuracy	Average Accuracy	Kappa Coefficient (κ)
SVM	81.01	88.25	75.86
ECHO	87.58	92.16	83.90
WH+MV	85.42	91.31	81.30
SVMMSF+MV	**91.08**	**94.76**	**88.30**
SVMMRF-NE	86.89	92.12	83.14
SVMMRF-E	87.63	93.41	84.07

edge-based SVMMRF-E method for the classification of urban images is confirmed, and 2) the MRF-based regularization includes the spatial context information from only the closest neighborhoods (in our case, eight neighborhoods) when classifying an image. Therefore, the proposed method is efficient only in the case if there is no large misclassified region in the initial pixelwise classification map (this assumption often holds). If such a region exists, the MRF-based method cannot reconstitute its true class label. This happens in the case of classification of the *University of Pavia* image, where some relatively large regions remain misclassified. The SVMMSF + MV method works differently: If there is a suspicion that a pixel might be misclassified, this pixel remains unclassified, and the classification decision is further taken by the region growing step. This appears to be a robust procedure for classifying large regions with uncertain spectral properties.

Furthermore, the proposed SVMMRF-E method was applied and has shown the best classification (overall and average) accuracies at the Third HYPER-I-NET summer school on hyperspectral imaging student contest "Evaluation of an unknown hyperspectral data set and information extraction." The contest was organized by P. Gamba on September 10, 2009, in Pavia, Italy, where the students in teams were supposed to provide a classification map of the rural area acquired by the Compact Airborne Spectrographic Imager sensor.

When comparing the results of several classifiers, an important issue is the computational cost of each classifier. Here, we compare the computational times for classification of the *University of Pavia* image using different methods. We conducted experiments on an Intel Core 2 Duo 2.40-GHz processor with 3.5-GB RAM. The processing times in seconds were 3339 for the SVM method, 3353 for the WH + MV method, 3351 for the SVMMSF + MV method, 3444 for the SVMMRF-NE method, and 3450 for the SVMMRF-E method. None of the algorithms has been implemented in parallel (which would further speed up computational times). While the SVM classifier is a computationally demanding algorithm, other considered methods require at maximum 3% more time to be executed. In terms of duration, the proposed SVMMRF-NE method takes 93 s longer for the classification of the data than the SVMMSF + MV approach, and the SVMMRF-E method takes 6 s longer than the SVMMRF-NE method.

IV. CONCLUSION

A novel accurate SVMMRF method for spectral-spatial classification of hyperspectral images has been presented in this letter. The method consists in performing a probabilistic SVM pixelwise classification, followed by MRF-based regularization for incorporating spatial and edge information into classification. Experimental results have demonstrated that the proposed method yields accurate classification maps within a short time interval and is sufficiently robust for classifying different kinds of images.

ACKNOWLEDGMENT

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

REFERENCES

[1] C.-I Chang, *Hyperspectral Data Exploitation: Theory and Applications*. New York: Wiley-Interscience, 2007.
[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] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
[5] 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.
[6] 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. 11, pp. 3804–3814, Nov. 2008.
[7] 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.
[8] 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.
[9] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
[10] F. Bovolo and L. Bruzzone, "A context-sensitive technique based on support vector machines for image classification," in *Proc. PReMI*, 2005, pp. 260–265.
[11] 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. Environ.*, vol. 101, no. 2, pp. 167–180, Mar. 2006.
[12] A. Solberg, T. Taxt, and A. Jain, "A Markov random field model for classification of multisource satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 1, pp. 100–113, Jan. 1996.
[13] J. Platt, "Probabilities for support vector machines," in *Advances in Large Margin Classifiers*, A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds. Cambridge, MA: MIT Press, 2000, pp. 61–74.
[14] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multiclass classification by pairwise coupling," *J. Mach. Learn. Res.*, no. 5, pp. 975–1005, 2004.
[15] Y. Tarabalka, J. Chanussot, and J. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010.
[16] R. Gonzalez and R. Woods, *Digital Image Processing,* 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
[17] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, 1953.
[18] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.
[19] Y. Tarabalka, J. Chanussot, J. A. Benediktsson, J. Angulo, and M. Fauvel, "Segmentation and classification of hyperspectral data using watershed," in *Proc. IGARSS*, Boston, MA, 2008, pp. III-652–III-655.