A Context-Sensitive Bayesian Technique for the Partially Supervised Classification of Multitemporal Images

R. Cossu, *Member, IEEE*, S. Chaudhuri, and L. Bruzzone, *Senior Member, IEEE*

***Abstract*—An advanced context-sensitive classification technique that exploits a temporal series of remote sensing images for a regular updating of land-cover maps is proposed. This technique extends the use of spatio-contextual information to the framework of partially supervised approaches (that are capable of addressing the updating problem under the realistic, though critical, constraint that no ground-truth information is available for some of the images to be classified). The proposed classifier is based on an iterative partially supervised algorithm that jointly estimates the class-conditional densities and the prior model for the class labels on the image to be classified by taking into account spatio-contextual information. Experimental results point out that the proposed technique is effective and that it significantly outperforms the context-insensitive partially supervised approaches presented in the literature.**

***Index Terms*—Contextual classification, expectation–maximization (EM) algorithm, Markov random fields (MRFs), partially supervised classification, partially supervised updating of land-cover maps.**

I. Introduction

REMOTE sensing images regularly acquired by spaceborne sensors on a specific area of interest can be analyzed with automatic classification techniques to derive updated land-cover maps of the studied site. At the operating level, to obtain a high accuracy, supervised classification algorithms are usually adopted. For classifier training, these algorithms require that ground-truth information be available. Unfortunately, in many real cases, it is not possible to rely on training data for all the images needed to ensure an updating of land-cover maps as frequently as required by the applications [1]–[5]. For this reason, some of the remotely sensed images acquired in the investigated area cannot be used to periodically update land-cover maps. Therefore, the process of temporal updating of land-cover maps results in a complex and challenging problem.

In the literature, different classification approaches based on partially supervised techniques capable of addressing the above-described problem have been proposed [1]–[5]. These approaches generate an accurate land-cover map by processing a new image even when the related training set is not available. In greater detail, they initialize the classifier parameters on an old image of the investigated area (for which ground-truth is available) and then update the parameter values according to specific unsupervised strategies. However, all partially supervised methods proposed so far give a final classification map in a pixelwise manner, assuming that the pixels in the image are independent of one another. On the contrary, in general the nonimpulsive spatial autocorrelation function of remote sensing images does not justify this assumption. In other words, in many remote sensing images, objects on the ground are much bigger than the ground instantaneous field of view of the sensor; thus, adjacent pixels are more likely to belong to the same class. This is particularly true for last-generation satellite sensors (e.g., Ikonos and Quickbird), which have a very high geometrical resolution. For these reasons, it is important to extend the partially supervised framework to context-sensitive methods. This is the specific aim of the present work.

In this letter, we propose a novel partially supervised classifier capable of considering the spatio-contextual information included in the neighborhood of each pixel in the classification process. To this end, an approach based on a specific procedure that properly integrates Markov random fields (MRFs) and the expectation–maximization (EM) algorithm is defined. MRFs are used because such an approach has revealed effective in the exploitation of spatio-contextual information [6]–[11].

The experimental analysis was carried out on multitemporal images acquired by the Thematic Mapper sensor of the Landsat-5 satellite on the Island of Sardinia, Italy. The results obtained confirm the effectiveness of the proposed partially supervised context-sensitive approach.

II. Proposed Partially Supervised Context-Sensitive Classifier

A. Problem Formulation

Let $\mathbf{X}_1 = \{x^1_{1,1}, x^1_{1,2}, \ldots, x^1_{J,K}\}$ and $\mathbf{X}_2 = \{x^2_{1,1}, x^2_{2,1}, \ldots, x^2_{J,K}\}$ denote two multispectral images made up of $J \times K$ pixels and acquired in the area under analysis at times t_1 and t_2, respectively. Let $\boldsymbol{x}^i_{j,k}$ be the $1 \times d$ feature vector (i.e., a row vector) associated with the pixel at position (j,k) of image $\mathbf{X}_i$, i.e., $x^i_{j,k}$, where d is the size of the input feature space. Let us assume that the same set $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$ of M land-cover classes

Manuscript received February 9, 2005; revised April 20, 2005. This work was carried out under the framework of the India-Trento Program for Advanced Research (ITPAR).

R. Cossu and L. Bruzzone are with the Department of Information and Communication Technology, University of Trento, I-38050 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

S. Chaudhuri is with the Department of Electrical Engineering, Indian Institute of Technology—Bombay, 400076 Mumbai, India (e-mail: sc@ee.iitb.ac.in).

Digital Object Identifier 10.1109/LGRS.2005.851541

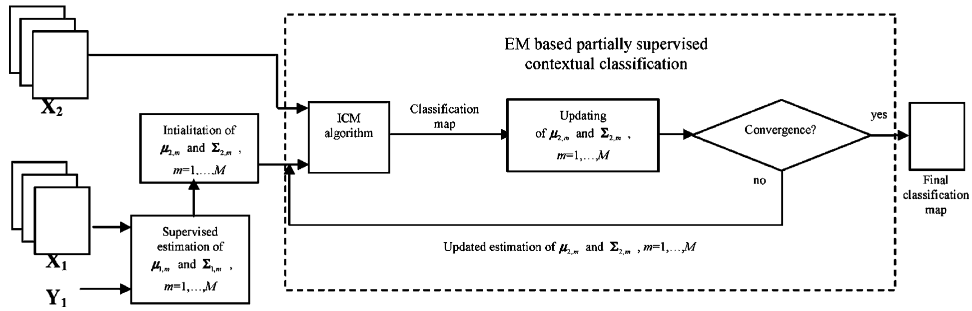


Fig. 1. Scheme of the proposed partially supervised contextual classifier. At each iteration of the EM algorithm a complete application of the ICM procedure is required.

characterizes the considered geographical area at both t_1 and t_2. Let $l^i_{j,k} \in \Omega$ be the class label of pixel $x^i_{j,k}$, and let $\mathbf{L}^i_l = \{l^i_{j,k},\ 1 \le j \le J,\ 1 \le k \le K\}$ be one particular set of class labels for the entire image, i.e., one possible classification map. The set $\mathbf{C} = \{\mathbf{L}^i_l, 1 \le l \le L\}$ (with $L = M^{J\times K}$) is made up of all the possible sets of labels in the image $\mathbf{X}_i$. Finally, let us assume that a reliable training set $\mathbf{Y}_1$ is available at t_1, while a training set is not available at t_2. The aim of this work is to develop an advanced classification approach capable of generating an accurate land-cover map related to $\mathbf{X}_2$ (i.e., the image for which ground-truth information is not available), by exploiting the information in the training set $\mathbf{Y}_1$, the signal associated with the image $\mathbf{X}_2$, and the spatio-contextual information in $\mathbf{X}_2$.

B. Proposed Technique

By taking into account the spatio-contextual information, the Bayes rule for minimum error [12] can be rewritten as

$$\begin{aligned}\mathbf{L}^i_V &= \underset{\mathbf{L}^i_l \in \mathbf{C}}{\arg\max}\left\{P_i\left(\mathbf{L}^i_l|\mathbf{X}_i\right)\right\}\\ &= \underset{\mathbf{L}^i_l \in \mathbf{C}}{\arg\max}\left\{P_i\left(\mathbf{L}^i_l\right) p_i\left(\mathbf{X}_i|\mathbf{L}^i_l\right)\right\}\end{aligned} \tag{1}$$

where $\mathbf{L}^i_V$ is the estimated classification map for image $\mathbf{X}_i$, $P_i(\mathbf{L}^i_l)$ is the prior model for the class labels, and $p_i(\mathbf{X}_i|\mathbf{L}^i_l)$ is the joint density function of the pixel values in image $\mathbf{X}_i$ given the set of labels $\mathbf{L}^i_l$. Maximization of (1) requires the estimation of $P_i(\mathbf{L}^i_l)$ and $p_i(\mathbf{X}_i|\mathbf{L}^i_l)$, which are very complex tasks.

To simplify the problem, two standard assumptions are made: 1) we assume the conditional independence of the joint density function with respect to the pixels in the image; and 2) we model the spatio-contextual information in a local spatial neighborhood [11]. Under the aforementioned assumptions, the estimated class label for the pixel $x^i_{j,k}$ (given the estimates of the class labels for its neighboring pixels in the image) is

$$\begin{aligned}l^i_{j,k} = \omega_v \quad \text{if} \quad \omega_v = &\underset{\omega_m \in \Omega}{\arg\max}\left\{p_i\left(\boldsymbol{x}^i_{j,k}|l^i_{j,k} = \omega_m\right)\right.\\ &\left.\times P_i\left(l^i_{j,k} = \omega_m|\left\{l^i_{g,h};(g,h) \in \mathbf{N}(j,k)\right\}\right)\right\}\end{aligned} \tag{2}$$

where $\mathbf{N}(j,k)$ is the set of neighboring pixels of $x^i_{j,k}$ (a first- or second-order neighborhood system is usually adopted) [6], [7]. It is worth noting that the spatio-contextual information in this formulation is included in the prior model of classes. In terms of the Markovian approach [6], it is possible to prove (using the equivalence of MRF to the Gibbs random field) that under the aforementioned assumptions, this procedure is equivalent to minimizing the following energy function [6], [7]:

$$\begin{aligned}U\left(\mathbf{X}_i, \mathbf{L}^i_l\right) = &-\sum_{x^i_{j,k} \in \mathbf{X}_i} \ln p_i\left(\boldsymbol{x}^i_{j,k}|l^i_{j,k}\right)\\ &+ \sum_{x^i_{j,k} \in \mathbf{X}_i} U_{\text{context}}\left(l^i_{j,k}|\left\{l^i_{g,h};(g,h) \in \mathbf{N}(j,k)\right\}\right)\end{aligned} \tag{3}$$

where the second term models the spatial context by taking into account the prior model of the class labels $\mathbf{L}^i_l$. For the image segmentation problem [9]–[11], a homogeneous and isotropic M-Level MRF model is usually adopted.

In a completely supervised framework, the minimization of (3) can easily be carried out in two different steps, resulting in a standard context-sensitive Bayesian classifier. In the first step, the available training set is used to estimate the class-conditional density functions [the noncontextual energy term in (3)]. Once these estimates are obtained, the class label for each pixel in the image can then be initialized by minimizing the noncontextual energy function in (3). In the second step, an optimization procedure is used [e.g., simulated annealing, iterated conditional mode (ICM) algorithm] [7], [11], [13], [14] in order to minimize the total energy function (3) and to derive the final estimates of the class labels.

However, in the partially supervised framework, minimization of (3) results in very a complex and challenging task, as we lack a reliable training set to estimate the conditional density functions of $\mathbf{X}_2$. We, therefore, propose to estimate jointly the density functions and the contextual prior model for class labels according to a specific partially supervised method. To this end, an iterative technique based on the expectation–maximization (EM) algorithm [15], [16] integrated with the ICM optimization procedure is used.

The proposed partially supervised approach is based on the observation that the statistical distribution of the pixel values in $\mathbf{X}_2$ can be described by a mixed-density distribution where the component densities are the conditional density functions of the classes and the mixing parameters depend on the prior model of class labels. Accordingly, the estimate of the class-conditional density functions is a mixture density estimation problem that can be solved by exploiting the iterative EM algorithm. In detail, the log-likelihood that must be maximized by the EM algorithm is given by

$$\begin{aligned}\Theta &= \log \prod_{x_{j,k}^2 \in \mathbf{X}_2} \hat{p}_2\left(\boldsymbol{x}_{j,k}^2\right) \\ &= \sum_{x_{j,k}^2 \in \mathbf{X}_2} \log \sum_{\omega_m \in \Omega} P_2\left(l_{j,k}^2 = \omega_m \mid \left\{l_{g,h}^2; (g,h) \in \mathbf{N}(j,k)\right\}\right) \\ &\quad \times p_2\left(\boldsymbol{x}_{j,k}^2 \mid \omega_m\right). \end{aligned} \tag{4}$$

It can be proved that the required iterative updating equations for the mean vector $\boldsymbol{\mu}_{2,m}$ and the covariance matrix $\boldsymbol{\Sigma}_{2,m}$ describing the density function of the class $\omega_m \in \Omega, (m = 1, \ldots, M)$, which are assumed to be Gaussian, are the following:

$$\boldsymbol{u}_{2,m}^{t+1} = \frac{\sum_{x_{j,k}^2 \in \mathbf{X}_2} P_2^t\left(l_{j,k}^2 = \omega_m \mid \boldsymbol{x}_{j,k}^2\right) \cdot \boldsymbol{x}_{j,k}^2}{\sum_{x_{j,k}^2 \in \mathbf{X}_2} P_2^t\left(l_{j,k}^2 = \omega_m \mid \boldsymbol{x}_{j,k}^2\right)} \tag{5}$$

and

$$\sum_{2,m}^{t+1} = \frac{\sum_{x_{j,k}^2 \in \mathbf{X}_2} P_2^t\left(l_{j,k}^2 = \omega_m \mid \boldsymbol{x}_{j,k}^2\right)\left(\boldsymbol{x}_{j,k}^2 - \boldsymbol{\mu}_{2,m}^{t+1}\right)^T\left(\boldsymbol{x}_{j,k}^2 - \boldsymbol{\mu}_{2,m}^{t+1}\right)}{\sum_{x_{j,k}^2 \in \mathbf{X}_2} P_2^t\left(l_{j,k}^2 = \omega_m \mid \boldsymbol{x}_{j,k}^2\right)} \tag{6}$$

where the superscripts t and $t+1$ refer to the values of the parameters at the current and next iterations, respectively; T refers to the vector transpose operation, and

$$\begin{aligned} &P_2^t\left(l_{j,k}^2 = \omega_m \mid \boldsymbol{x}_{j,k}^2\right) \\ &= \frac{p_2^t\left(\boldsymbol{x}_{j,k}^2 \mid \omega_m\right) P_2^t\left(l_{j,k}^2 = \omega_m \mid \left\{l_{g,h}^2; (g,h) \in \mathbf{N}(i,j)\right\}\right)}{\sum_{\omega_c \in \Omega} p_2^t\left(\boldsymbol{x}_{j,k}^2 \mid \omega_c\right) P_2^t\left(l_{j,k}^2 = \omega_c \mid \left\{l_{g,h}^2; (g,h) \in \mathbf{N}(i,j)\right\}\right)}. \end{aligned} \tag{7}$$

Accordingly, the prior model of class labels needs to be estimated at each iteration of the EM algorithm so that (7) can be computed (and consequently the mean and covariance matrix updated). In this context, it is worth noting that based on both the definition of $U_{\text{context}}(.)$ and (2), we can write [7]–[9]

$$\begin{aligned} &P_2^t\left(l_{j,k}^2 = \omega_m \mid \left\{l_{g,h}^2, (g,h) \in \mathbf{N}(j,k)\right\}\right) \\ &\quad = \frac{1}{z} \exp\left\{-\beta \sum_{(g,h) \in \mathbf{N}(i,j)} \left[1 - \delta\left(l_{j,k}^2, l_{g,h}^2\right)\right]\right\} \end{aligned} \tag{8}$$

where z is a normalizing constant, β is a model parameter that tunes the influence of the spatio-contextual information on the classification process (it can be estimated from the image or determined empirically [10]), and $\delta(.)$ is the Kronecker delta function, defined as

$$\delta\left(l_{j,k}^2, l_{g,h}^2\right) = \begin{cases} 1, & \text{if } l_{j,k}^2 = l_{g,h}^2 \\ 0, & \text{if } l_{j,k}^2 \neq, l_{g,h}^2. \end{cases} \tag{9}$$

In practice, an ICM is nested in the EM, so that at each iteration of the EM algorithm a complete ICM cycle is performed and an (intermediate) context-sensitive classification map is obtained (see Fig. 1). Such a map is necessary to compute the weighting factors $P_2(l_{j,k}^2 = \omega_m \mid \boldsymbol{x}_{j,k}^2)$ to be used in the updating equations of the EM algorithm.

Summarizing, the proposed contextual partially supervised classifier consists of the following iterative algorithm.

1. Initialize the mean vector and the covariance matrix of each class with the supervised estimates obtained for image $\mathbf{X}_1$ using the training set $\mathbf{Y}_1$;
2. Given the current estimates of the class-conditional densities, apply the ICM algorithm to classify image $\mathbf{X}_2$;
3. For each pixel in $\mathbf{X}_2$ and for each class $\omega_m \in \Omega$, given the current estimates, compute the contextual prior model for class labels $P_2^t(l_{j,k}^2 = \omega_m \mid \{l_{g,h}^2; (g,h) \in \mathbf{N}(i,j)\})$ and the class conditional density $p_2(\boldsymbol{x}_{j,k}^2 \mid \omega_m)$;
4. Update the mean and covariance matrix of each class by means of (5) and (6);
5. Go to step 2 until convergence is reached, i.e., the difference between the parameter values estimated at two iterations is smaller than a given threshold $\varepsilon > 0$.

TABLE I
NUMBER OF PATTERNS IN THE TRAINING AND TEST SETS USED IN THE EXPERIMENTAL ANALYSIS. A TRAINING SET IS AVAILABLE FOR THE SEPTEMBER 1995 IMAGE ONLY. A TEST SET IS AVAILABLE FOR BOTH IMAGES. ALL CLASSES ARE REPRESENTED BY THE SAME NUMBER OF PATTERNS IN THE TWO TEST SETS

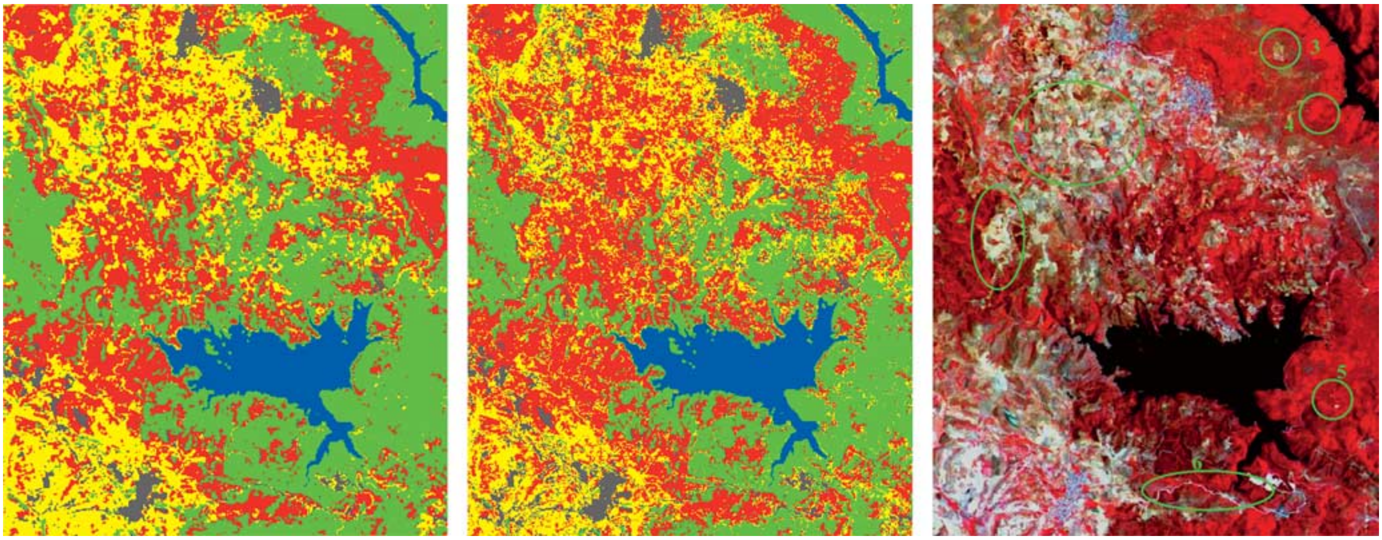
Land-cover class	Number of patterns	
	September 1995 Training set	September 1995 and July 1996 Test sets
Pasture	554	589
Forest	304	274
Urban area	408	418
Water body	804	551
Vineyard	179	117
Overall	2249	1949

TABLE II
OVERALL CLASSIFICATION ACCURACIES EXHIBITED BY THE CONSIDERED CLASSIFIERS (TRAINED IN A SUPERVISED WAY ON THE SEPTEMBER 1995 IMAGE) BEFORE APPLICATION OF THE PARTIALLY SUPERVISED TECHNIQUE

Classification technique	Overall classification accuracy (%)	
	Test set (September 1995)	Test set (July 1996)
PBB	90.97	50.43
CSB	91.43	49.72

TABLE III
CLASSIFICATION ACCURACIES EXHIBITED BY THE PROPOSED PS-CSB CLASSIFIER ON THE JULY 1996 TEST SET AFTER RETRAINING. FOR COMPARISON THE RESULTS OBTAINED BY THE PS-PBB CLASSIFIER PROPOSED IN [1] ARE ALSO GIVEN

Land-cover class	Classification accuracy (%) (July 1996 test set)	
	PS-CSB	PS-PBB
Pasture	94.23	94.06
Forest	96.72	87.22
Urban area	96.17	93.06
Water body	100.00	100.00
Vineyard	76.07	64.10
Overall	95.54	92.76



(a) (b) (c)

Fig. 2. Classification map of the image acquired in July 1996 obtained by (a) the proposed PS-CSB classifier and (b) the PS-PBB classifier presented in [1] (red $\rightarrow$ pasture; green $\rightarrow$ forest; gray $\rightarrow$ urban area; blue $\rightarrow$ water body; yellow $\rightarrow$ vineyard). c) False-color composition of the July 1996 image (channels 4, 3, and 1 are considered). Numbered areas of interest are discussed in the text.

III. Experimental Results

In order to assess the effectiveness of the proposed approach, a number of experiments were carried out on a dataset made up of two multispectral images acquired by the Thematic Mapper multispectral sensor of the Landsat-5 satellite. The selected test site was a section (412×493 pixels) of a scene including Lake Mulargia on the Island of Sardinia, Italy. The two images used in the experiments were acquired in September 1995 (t_1) and July 1996 (t_2). The available ground-truth was used to derive a training set for the t_1 image and a test set for both the t_1 and the t_2 images. Five land-cover classes (i.e., urban area, forest, pasture, water body, and vineyard), which characterized the test site at the above-mentioned dates, were considered. A detailed description of the training and test sets of both images is given in Table I. It is worth noting that we assume that a training set for the t_2 image is not available. The assumption of Gaussian distributions was made for the density functions of the classes in all experiments (this is a reasonable assumption, as we considered Thematic Mapper images).

For comparison, both a pixel-based Bayesian (PBB) classifier (which exploits the Bayes rule for the minimum error) and a context-sensitive Bayesian (CSB) classifier (both based on the standard supervised approach) were trained on the $\mathbf{X}_1$ image using the available training set $\mathbf{Y}_1$. As regards the CSB classifier, a first order neighborhood system was adopted. The parameter β was automatically computed by the minimum perturbation strategy proposed in [10] and set equal to 0.94. After the supervised training on the $\mathbf{X}_1$ image, the effectiveness of the classifiers was evaluated on the test sets related to both images (see Table II). On the one hand, the classifiers provided high overall classification accuracies for the September 1995 test set, while on the other hand, as expected, they showed very low accuracy on the July 1996 test set, because the two images have significantly different properties. In particular, the overall classification accuracies provided by the PBB and CSB algorithms for the July test set were 50.43% and 49.72%, respectively, which are not acceptable results.

At this point, the values of the parameters of the CSB classifier were updated on the t_2 image (July 1996) using the proposed partially supervised technique. The processes converged in 19 iterations. The overall and class-by-class accuracies exhibited by the proposed partially supervised context-sensitive Bayesian (PS-CSB) classifier after the retraining phase are given in Table III. For the sake of comparison, the results obtained after convergence (23 iterations) by the partially supervised pixel-based Bayesian (PS-PBB) approach proposed in [1] are also reported.

By a comparison of Tables II and III, it can be seen that the classification accuracies provided by the considered contextual partially supervised classifier for the July 1996 test set are much higher than those shown by the classifier trained on the September 1995 image (i.e., 95.54% versus 49.72%). In greater detail, the retrained classifier showed sharply higher accuracies on all land-cover classes. Most interestingly, the proposed PS-CSB classifier notably outperforms the PS-PBB classifier proposed in [1]. The most significant improvement can be observed for the two minority classes (i.e., vineyard and forest). Here accuracy increases by about 12% and 10%, respectively.

Fig. 2(a) and (b) shows the maps obtained by the PS-CSB and PS-PBB classifiers, respectively, after convergence. From a qualitative analysis of these maps, we can observe that with the proposed PS-CSB classifier one can obtain a more regularized map where punctual noise is strongly reduced. For further insight into the behavior of the proposed technique let us consider the false-color composition of channels 4, 3, and 1 related to the July 1996 image [see Fig. 2(c)]. We have highlighted some areas of interest (AOI) on which we shall mainly focus our qualitative analysis. The first difference that is observed between the two maps is in the relative occurrence of the vineyard class. A visual inspection of AOIs 1 and 2 seems to confirm the presence of vineyards detected by the PS-CSB classifier. The same seems to occur in AOI 3. AOIs 4 and 5 clearly show the ability of the proposed contextual approach to identify and correctly classify noisy pixels (it is unlikely that pixels in the middle of the forest should belong to the vineyard class as given by the PS-PBB). Finally, AOI 6 shows an example where the PS-PBB classifier seems to outperform the PS-CSB slightly. In particular, some roads in the image are not correctly classified by the proposed approach, because a limited spatial resolution does not favor a context-based approach.

IV. Discussion and Conclusion

In this letter, a novel context-sensitive approach to the partially supervised classification of multitemporal remote sensing images has been proposed. This approach extends the use of the spatio-contextual information, modeled with an MRF approach, to the partially supervised classification framework. In particular, the proposed technique integrates MRF methodology with the EM algorithm in the context of an optimizing procedure based on the ICM method.

It is worth noting that the addressed partially supervised framework is extremely important also with respect to operational applications of remote sensing. This is motivated by the fact that it is often not possible to update land-cover maps regularly according to the standard supervised classification of a temporal series of images, because ground-truth information for some images is completely lacking in real problems. In these cases, partially supervised approaches represent a reliable and effective alternative to completely unsupervised techniques.

Though extensive experiments on other datasets are necessary for a final validation of the method, our results on the considered dataset are very good. Therefore, the resulting classifier is very promising for the analysis of medium geometrical resolution images (e.g., Thematic Mapper images) and even more promising in the analysis of very high geometrical resolution images acquired by last generation satellite sensors (e.g., Ikonos and Quickbird), which are characterized by a strong interpixel correlation.

References

[1] L. Bruzzone and D. Fernández Prieto, "Unsupervised retraining of a maximum-likelihood classifier for the analysis of multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 456–460, Feb. 2001.

[2] L. Bruzzone, R. Cossu, and G. Vernazza, "Combining parametric and nonparametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images," *Inf. Fusion*, vol. 3, pp. 289–297, 2002.

[3] L. Bruzzone and D. Fernández Prieto, "A partially unsupervised cascade classifier for the analysis of multitemporal remoting images," *Pattern Recognit. Lett.*, vol. 23, no. 9, pp. 1063–1071, 2002.

[4] L. Bruzzone and R. Cossu, "A multiple cascade-classifier system for a robust a partially unsupervised updating of land-cover maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 1984–1996, Sep. 2002.

[5] L. Bruzzone and R. Cossu, "Advanced classification techniques: Partially supervised approaches," in *Frontiers of Remote Sensing Information*, C. H. Chen, Ed, Singapore: World Scientific, 2003, ch. 12, pp. 285–314.

[6] S. Z. Li, *Markov Random Field Modeling in Computer Vision*. Berlin, Germany: Springer-Verlag, 1995.

[7] R. Chellapa and A. K. Jain, Eds., *Markov Random Fields: Theory and Applications*. New York: Academic, 1993.

[8] A. H. S. Solberg, T. Taxt, and A. K. 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.

[9] P. C. Smits and S. G. Dellepiane, "Synthetic aperture radar image segmentation by a detail preserving Markov random field approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 844–857, Jul. 1997.

[10] F. Melgani and S. B. Serpico, "A Markov random field approach to spatio-temporal contextual image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2478–2487, Nov. 2003.

[11] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.

[12] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974.

[13] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images," *Pattern Recognit.*, vol. 22, no. 6, pp. 747–762, 1989.

[14] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc. B*, vol. 48, pp. 259–302, 1986.

[15] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, vol. 39, pp. 1–38, 1977.

[16] T. K. Moon, "The expectation-maximization algorithm," *Signal Process. Mag.*, vol. 13, pp. 47–60, 1996.