CONTEXTUAL REMOTE-SENSING IMAGE CLASSIFICATION BY SUPPORT VECTOR MACHINES AND MARKOV RANDOM FIELDS

Gabriele Moser and Sebastiano B. Serpico

University of Genoa, Dept. of Biophysical and Electronic Eng. (DIBE),
Via Opera Pia 11a, I-16145, Genoa (Italy), e-mail: sebastiano.serpico@unige.it

1. EXTENDED ABSTRACT

Supervised classification methods play a central role in the framework of environmental information extraction from remote-sensing images, for instance, in such applications as land-use or land-cover mapping, forest inventory, or urban-area monitoring. In this context, kernel-based support vector machine (SVM) methods [1] have recently been receiving a very strong attention [2], thanks to both their good analytical properties in terms of generalization capability and robustness to dimensionality issues [1] and to the accurate experimental results they obtained in several remote-sensing image classification problems (see, e.g., [3, 4]). However, SVM-based methods have been developed by focusing on the case of independent and identically distributed (i.i.d.) samples and class labels [1]; in terms of image classification, this strategy results in an intrinsically noncontextual approach and represents a significant limitation since it discards the information associated to the correlation among neighboring pixels in an image. This drawback is especially relevant when classifying high-resolution images, for which spatial information plays a primary role.

On the other hand, a general and powerful approach to the integration of spatial-contextual information in image analysis is represented by Markov random fields (MRFs), which are a general family of stochastic image models that allow both spatial-contextual information and further possible information sources to be integrated into Bayesian image-analysis schemes, through the minimization of suitable “energy functions” [5]. Successful applications of MRF-based modeling has been used in remote-sensing for classification, change-detection, denoising, and feature extraction (an overview can be found in [6]). However, an integration of the SVM and MRF frameworks is not trivial, due to the intrinsically nonbayesian nature of SVM.

In this paper, a novel approach is proposed that combines the SVM and MRF strategies to classification. This is accomplished by developing a novel reformulation of the Markovian minimum-energy classification rule, which

Table 1. "Itaipu" data set: classification accuracies of the proposed method, of a classical contextual MRF-based classifier, and of a noncontextual SVM-based classifier (see abstract text for details) over the test set.

CLASS:	urban	herbaceous	shrub-brush rangeland	forest	barren land	built-up	water	overall accuracy
proposed method	0.9747	0.9977	1	1	0.7237	0.9628	0.9997	0.9898
MRF	0.7802	0.9648	0.9981	0.9941	0.7588	0.9882	0.9994	0.9685
SVM	0.5827	0.9228	0.9876	0.9139	0.6833	0.9345	0.9995	0.9392

is proved to be equivalent to the application of an SVM in a suitable nonlinearly transformed feature space. This integrated SVM-MRF formulation allows the robustness of SVMs to overfitting and dimensionality issues and the contextual modeling capability of MRFs to be combined in order to jointly exploit both the spectral and spatial information associated with a remote-sensing data set.

Specifically, let us focus first on a binary classification problem. Given the mapping of the original feature space to a higher-dimensional inner-product space implicitly defined by a given kernel function [1], let us assume, as a working hypothesis, that the probability density functions of the transformed feature vector, conditioned to the two classes, can be modeled as Gaussian distributions with equal covariance matrices. As discussed in [7], this apparently restrictive assumption is actually pretty mild, thanks to the flexibility of the nonlinear mapping that may be implicitly represented by a kernel function and to the possibly very high (or even infinite) dimensionality of the resulting transformed space. Let us further assume that the random field of the class labels associated to the image pixels is an MRF [8]. Under these assumptions, we prove in the full paper that the Markovian minimum-energy rule is analytically equivalent to the application of a linear discriminant function in a further nonlinearly transformed feature space, which can be implicitly represented by a suitable kernel function as well. Specifically, this MRF-based kernel is expressed as a linear combination of two contributions, related to the pointwise information associated to the feature vector of each pixel and to the contextual information, respectively, and the MRF-based decision rule turns out to be equivalent to the application of an SVM-based classifier with this novel kernel.

This formalization leads to a kernel-based integration of the SVM and MRF approaches to image classification and is exploited here by proposing a specific novel formulation of the iterated conditional mode (ICM) algorithm for MRF-based classification. ICM is an iterative energy-minimization algorithm, which is known to converge at least to a local minimum of the energy and usually represents a good tradeoff between classification accuracy and computational burden [8, 5]. Accordingly, the proposed method is iterative, it is initialized with a preliminary classification result (generated, for instance, by a classical noncontextual SVM-based classifier), and, at each iteration, it updates the classification map by performing an SVM-based classification with the proposed MRF-based kernel. Generalization to multiclass classification is obtained by a one-against-one strategy [1].

Table 2. "Pavia" data set: classification accuracies of the proposed method, of a classical contextual MRF-based classifier, and of a noncontextual SVM-based classifier (see abstract text for details) over the test set.

CLASS:	barren land	wetland	overall accuracy
proposed method	0.9543	0.9824	0.9620
MRF	0.9479	0.9611	0.9515
SVM	0.9284	0.9113	0.9237

Furthermore, the proposed method is also endowed with specific algorithms aimed at automatically optimizing the values of the related internal parameters, i.e., regularization and kernel parameters included in the SVM formulation [1, 9], as well as further weight parameters, tuning the reciprocal role of the contextual and noncontextual contributions in the proposed kernel. In order to automatically identify optimal values for the former set of parameters, the approach proposed in [10] in the framework of SVM-based regression is extended here to classification. This approach is based on the numerical minimization of the span-bound functional, which is a (tight) analytical upper bound on the leave-one-out error [9] and has been found to be often strongly correlated also with test-set hold-out errors [9, 10]. As the span bound is a nondifferentiable function of the SVM regularization and kernel parameters, the Powell method is used to address the related numerical minimization [11, 10]. Moreover, with regard to the weight parameters, the method in [12], which is based on the application of the Ho-Kashyap numerical procedure to a suitable set of linear inequalities, is applied to the proposed SVM-MRF formulation.

The proposed method was applied with a Gaussian kernel and experimentally validated with two data sets represented by a multispectral IKONOS image (1999×1501 pixels) acquired over Itaipu (Brazil/Paraguay border) and including seven main classes (Table 1), and to a multipolarization and multifrequency SIR-C/XSAR image (700×280 pixels) acquired over Pavia (Italy), consisting of one X-band channel and three C-band channels with different polarizations, and including two main classes (Table 2). The results were compared with those given by: (i) a noncontextual SVM-based classifier (whose regularization and kernel parameters were automatically optimized by the above-mentioned method based on Powell minimization of the span bound); (ii) a contextual MRF-based classifier, applied by computing estimates of the class posterior probabilities [5] according to a Gaussian model (which is usually accepted for the class-conditional statistics in multispectral images) in the case of the "Itaipu" data set and through the "k-nearest neighbor" algorithm in the case of the "Pavia" data set. Tables 1 and 2 summarize the resulting classification accuracies with respect to test maps.

The developed algorithm obtained very accurate classification results with overall accuracies (OAs) around 99% and 96% for "Itaipu" and "Pavia," respectively, which suggests a strong effectiveness of the proposed approach to contextual remote-sensing image classification. The two methods used for comparison provided quite accurate results as well, even though the proposed technique allowed 1% and 2% increases in OA with respect to the classical MRF-based classifier for "Itaipu" and "Pavia," respectively, and 5% and 4% increases in OA with respect to the noncontextual SVM-based classifier for "Itaipu" and "Pavia," respectively. It is worth noting that such results were obtained with both high-resolution optical multispectral data and medium-resolution multichannel SAR data, which suggests the effectiveness of the proposed technique in combining both the flexibility of the SVM-based approach to nonparametric classification and the capability, typical of MRF-based methods, to exploit spatial-contextual information for image-classification purposes. Further details about both methodological/analytical aspects of the technique and its experimental validation (also including a visual analysis of the related classification maps) will be included in the full paper.

2. REFERENCES

[1] V. N. Vapnik, *Statistical learning theory*, Wiley Interscience, 1998.

[2] G. Camps-Valls and L. Bruzzone (eds.), *Kernel methods for remote sensing data analysis*, John Wiley & Sons., 2009.

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

[4] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, pp. 725–749, 2002.

[5] 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 Sensing*, vol. 34, no. 1, pp. 100–113, 1996.

[6] G. Moser, F. Melgani, and S. B. Serpico, *Advances in unsupervised change detection*, in "Frontiers of Remote Sensing Information Processing," editor: C. H. Chen, World Scientific Publishing, 2003.

[7] M. M. Dundar and D. A. Landrebe, "Toward an optimal supervised classifier for the analysis of hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, no. 1, pp. 271–277, 2004.

[8] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Appl. Stat.*, vol. 16, pp. 131–163, 1989.

[9] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Mach. Learn.*, vol. 46, pp. 131–159, 2002.

[10] G. Moser and S. B. Serpico, "Automatic parameter optimization for support vector regression for land and sea surface temperature estimation from remote-sensing data," *IEEE Trans. Geosci. Remote Sensing*, vol. 47, pp. 909–921, 2009.

[11] W. H. Press, S. A. Teukolsky, W. T. Wetterling, and B. P. Flannery, *Numerical recipes in C*, Cambridge University Press, Cambridge (UK), 2002.

[12] S. B. Serpico and G. Moser, "Weight parameter optimization by the Ho-Kashyap algorithm in MRF models for supervised image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, pp. 3695–3705, 2006.