CONTEXTUAL REMOTE-SENSING IMAGE CLASSIFICATION THROUGH SUPPORT VECTOR MACHINES, MARKOV RANDOM FIELDS AND GRAPH CUTS

Andrea De Giorgi, Gabriele Moser, Sebastiano B. Serpico

University of Genoa, Dept. of Electrical, Electronic, Telecommunications Engineering and Naval Architecture, via Opera Pia 11a, I-16145, Genoa, Italy, gabriele.moser@unige.it

ABSTRACT

The problem of remote-sensing image classification is addressed in this paper by proposing a novel contextual classification method that integrates support vector machines (SVMs), Markov random fields (MRFs), and graph cuts. The proposed approach is methodologically explained by the aim to combine the robustness to dimensionality issues and the generalization capability of SVMs, the effectiveness of Markov models in characterizing the spatial contextual information associated with an image, and the capability of graph cut techniques in tackling complex problems of global minimization in computationally acceptable times. In the proposed method, the MRF minimum-energy problem is formalized in terms of an appropriate SVM kernel expansion and addressed through graph cuts. Parameter estimation is automated through two specific algorithms, based on the Ho-Kashyap and Powell numerical procedures. Experiments are carried out with two data sets consisting of multichannel SAR and multispectral high-resolution images.

Index Terms— Support vector machines, Markov random fields, graph cuts.

1. INTRODUCTION

In remote sensing applications of land-use or land-cover mapping, such as forest inventory, urban planning, and resource management, accurate results can generally be achieved through supervised classification methods. A currently popular and effective approach is represented by Support Vector Machines, which obtain accurate results in multiple applications and exhibit remarkable properties of generalization capability [1]. However, a limitation to the use of SVMs for the classification of remote sensing images is due to their intrinsic non-contextual nature, which is especially critical with very high resolution (VHR) images, in which the correlation between neighboring pixels and the spatial-geometrical structure are particularly important. This kind of contextual information can be efficiently modeled and incorporated in the classification process through Markov Random Fields. An MRF is a model for the spatial-contextual information and allows a Bayesian classification problem to be formulated in terms of the minimization of a suitable "energy function" [2]. Several methods have been proposed to address this minimization problem and, among them, graph-cut (GC) approaches have lately been receiving considerable attention. GC methods formulate the energy minimization problem as a maximum flow problem on an appropriate graph associated with the image: in the case of binary classification, they allow converging to a global minimum in generally acceptable times; in the multiclass case, they are proven to reach a "strong local minimum," which is characterized by good optimality properties [3].

These comments explain the potential of integrating the SVM, MRF, and GC approaches in a unique framework with the aim of combining their benefits with respect to a remote sensing image classification problem. In this paper, first, a novel energy-function formulation is proposed that incorporates the contextual information modeled by an MRF into an SVM classifier. In this formulation, through a "Markovian kernel," a relationship between the Markovian minimum-energy rule and the SVM discriminant function is established in a suitable transformed space. Then, this formulation is combined with GC techniques to develop a novel contextual SVM-MRF-GC classification method. A previous classifier was proposed in [4] to combine the SVM and MRF approaches, performing energy minimization through the iterated conditional mode (ICM) algorithm that converges only to a local minimum and may exhibit critical dependence on the initialization. The method proposed here, which, unlike the previous ICM-based algorithm, requires a kernel-based reformulation of the global posterior energy, overcomes these limitations through the combination with GCs.

2. METHODOLOGY

2.1. Proposed formulation of the energy function

Let $\mathcal{I}$ be a regular pixel lattice associated with a remote-sensing image composed of d channels. Each pixel is char-

This work was supported by the Italian Ministry of Education, University, and Research under the "PRIN 2012 - Very high spatial and spectral resolution remote sensing: a novel integrated data analysis system" project. The support is gratefully acknowledged.

acterized by a feature vector $\boldsymbol{x}_i$ ($\boldsymbol{x}_i \in \mathbb{R}^d$) and a class label y_i. Let $\mathcal{L} \subset \mathcal{I}$ be the set of training pixels (i.e., y_i is *a-priori* known for $i \in \mathcal{L}$). We refer first to binary classification, so two possible labels $y_i = 1$ or $y_i = -1$ can be assigned to the i-th pixel according to its class membership. Hence, a continuous-valued random field $\mathcal{X} = \{\boldsymbol{x}_i\}_{i \in \mathcal{I}}$ of feature vectors and a binary-valued random field $\mathcal{Y} = \{y_i\}_{i \in \mathcal{I}}$ of class labels are defined on the pixel lattice $\mathcal{I}$. In order to consider $\mathcal{Y}$ as an MRF, a neighborhood system $\{\partial i\}_{i \in \mathcal{I}}$ is assumed to be defined on the lattice $\mathcal{I}$, where $\partial i \subset \mathcal{I}$ is the set of the neighbors of the i-th pixel [2].The label field $\mathcal{Y}$ is an MRF with respect to this neighborhood system if its joint probability mass function is strictly positive and if the Markovianity property holds [2]. Under this assumption, several energy minimization techniques, including the aforementioned ICM, can be expressed in terms of a local energy-difference function, i.e. [2, 4]:

$$\Delta U_i(\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) = U_i(-1|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) - U_i(1|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}). \quad (1)$$

where:

$$U_i(y_i|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) = g(\boldsymbol{x}_i|y_i) + \beta \mathcal{E}_i(y_i|\boldsymbol{y}_{\partial i}) \quad (2)$$

is composed of two contributions [3]. The first one is related to non-contextual pixelwise statistics. The second one is a prior energy function $\mathcal{E}_i$ that characterizes the adopted MRF model. The positive parameter β determines the influence of the spatial contribution on the overall energy function [2]. However, GC methods are not formalized in terms of only the local energy difference function. Therefore, to achieve a fully rigorous SVM-MRF-GC integration, a new explicit formulation is needed for the global posterior energy function. Specifically, this function can be written as:

$$\begin{aligned} U(\mathcal{Y}|\mathcal{X}) &= \sum_{i \in \mathcal{I}} U_i(y_i|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) = \\ &= \sum_{i \in \mathcal{I}} \frac{U_i(y_i|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) + U_i(-y_i|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i})}{2} + \\ &+ \sum_{i \in \mathcal{I}} \frac{U_i(y_i|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) - U_i(-y_i|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i})}{2}. \end{aligned} \quad (3)$$

Using $\Delta U_i(\boldsymbol{x}_i, \boldsymbol{y}_{\partial i})$, it is easily proven that:

$$U_i(-y_i|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) - U_i(y_i|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) = y_i \Delta U_i(\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}). \quad (4)$$

Similarly:

$$\begin{aligned} &U_i(y_i|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) + U_i(-y_i|\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) = \quad (5) \\ &= [g(\boldsymbol{x}_i|1) + g(\boldsymbol{x}_i|-1)] + \beta[\mathcal{E}_i(1|\boldsymbol{y}_{\partial i}) + \mathcal{E}_i(-1|\boldsymbol{y}_{\partial i})], \end{aligned}$$

where the first bracketed term is obviously independent of $\mathcal{Y}$. For many popular choices of the function $\mathcal{E}_i$, such as, for instance, the well known Potts model [2], also the second bracketed term is a constant (i.e., it is independent of $\boldsymbol{y}_{\partial i}$). Assuming that the adopted MRF model satisfies this property, i.e., $\mathcal{E}_i(1|\boldsymbol{y}_{\partial i}) + \mathcal{E}_i(-1|\boldsymbol{y}_{\partial i})$ is a constant, and plugging Eqs.(4) and (5) into (3), we obtain:

$$U(\mathcal{Y}|\mathcal{X}) = -\frac{1}{2} \sum_{i \in \mathcal{I}} y_i \Delta U_i(\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) + G(\mathcal{X}), \quad (6)$$

where $G(\mathcal{X})$ depends on the feature vectors but not on the labels and has no effect on the minimization result.

As proven in [4], ΔU_i can be expressed as an SVM-like discriminant function associated with a special kernel, named Markovian kernel, i.e.:

$$\Delta U_i(\boldsymbol{x}_i, \boldsymbol{y}_{\partial i}) = \sum_{j \in \mathcal{S}} \alpha_j y_j K_{\mathsf{MRF}}(\boldsymbol{x}_i, \varepsilon_i; \boldsymbol{x}_j, \varepsilon_j) + b, \quad (7)$$

where

$$K_{\mathsf{MRF}}(\boldsymbol{x}_i, \varepsilon_i; \boldsymbol{x}_j, \varepsilon_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j) + \beta \varepsilon_i \varepsilon_j \quad (8)$$

is the Markovian kernel and ε_i is an additional feature defined through the adopted MRF model:

$$\varepsilon_i = \mathcal{E}_i(-1|\boldsymbol{y}_{\partial i}) - \mathcal{E}_i(1|\boldsymbol{y}_{\partial i}). \quad (9)$$

The set $\mathcal{S} \subset \mathcal{L}$ of support vectors and the weight coefficients α_j ($j \in \mathcal{S}$) are obtained through the training of an SVM with kernel K_{MRF} (further details can be found in [4]). Plugging Eqs. (7) and (8) into (6), we obtain the following expression for the global posterior energy function (up to additive and multiplicative terms constant with respect to $\mathcal{Y}$):

$$\begin{aligned} \tilde{U}(\mathcal{Y}|\mathcal{X}) &= -\sum_{i \in \mathcal{I}} y_i \left[\sum_{j \in \mathcal{S}} \alpha_j y_j K_{\mathsf{MRF}}(\boldsymbol{x}_i, \varepsilon_i; \boldsymbol{x}_j, \varepsilon_j) + b \right] = \\ &= -\sum_{i \in \mathcal{I}} y_i \sum_{j \in \mathcal{S}} [\alpha_j y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) + b] - \beta \sum_{i \in \mathcal{I}} y_i \varepsilon_i \sum_{j \in \mathcal{S}} \alpha_j y_j \varepsilon_j. \end{aligned} \quad (10)$$

Hence, the global posterior energy function corresponding to the application of a Markovian classifier in the transformed space $\mathcal{F}$ associated with a kernel K can be expressed as a linear combination of two contributions related to either the pixelwise information conveyed by the feature vectors $\boldsymbol{x}_i$ or the contextual information characterized through the additional feature ε_i ($i \in \mathcal{I}$). The MRF minimum energy rule in the space $\mathcal{F}$ can be related to a kernel expansion, similar to a traditional SVM, by using an additional feature associated with the adopted MRF model and the Markovian kernel. This conclusion holds for an arbitrary choice of the kernel K, associated with an either finite- or infinite-dimensional space $\mathcal{F}$ [4].

2.2. The proposed classification method

The proposed contextual classifier, named MSVC-GC (Markovian Support Vector Classifier with Graph Cuts), is iterative and is based on the application of a GC approach to minimize the energy function in Eq. (10). Here, we refer to binary classification; the extension to the multiclass case is carried out through decomposition into multiple binary subproblems and application of the one-against-one approach, which generally favors high accuracy and minimizes sensitivity to unbalanced classes [4]. The method performs the following processing steps:

- Initialization phase:
  a) Initialize β (e.g., $\beta = 1$) and automatically optimize SVM regularization and kernel parameters, collected in a vector $\boldsymbol{\theta}$, through the Powell-span-bound algorithm that minimizes an upper bound (named "span bound") on the leave-one-out error rate [4].
  b) Train a non-contextual SVM with the kernel K and the estimate of $\boldsymbol{\theta}$ obtained in step a), and generate a preliminary classification map by running the trained SVM.
- Iterative phase: for a fixed number of iterations, perform the following steps:
  1. Compute ε_i for each $i \in \mathcal{I}$ on the basis of the current classification map (see Eq. (9)).
  2. Estimate β by applying the automatic algorithm proposed in [5] and [4] and based on the Ho-Kashyap numerical procedure.
  3. Train an SVM with the Markovian kernel K_{MRF}, the contextual feature ε_i updated in step 1, and the estimates obtained for $\boldsymbol{\theta}$ and β in steps a) and 2, respectively.
  4. Compute $\tilde{U}(\mathcal{Y}|\mathcal{X})$ on the basis of the support vectors $\mathcal{S}$ and the related weights α_j ($j \in \mathcal{S}$) obtained in step 3 (see Eq. (10)).
  5. Minimize $\tilde{U}(\mathcal{Y}|\mathcal{X})$ with respect to $\mathcal{Y}$ through the GC "swap move" algorithm: it is a GC method to solve min-cut/max-flow problems with multiple binary graph cuts and determines a cycle of iterations for each pair of labels such that, through a sequence of optimal substitutions, a strong local minimum (i.e., a local minimum such that there is no better local label configuration) is reached. More details can be found in [3].

As mentioned in Sec. 1, convergence is ensured for the application of GC in step 5 [3]. However, convergence is not analytically guaranteed for the proposed iterative algorithm (although a good convergent behavior was remarked in all experiments). Thus, the iterative phase is run for a fixed number of iterations and the output classification map is the one with the minimum value of the global posterior energy within the maps generated in all iterations.

3. EXPERIMENTAL RESULTS

The experimental validation of MSVC-GC was carried out with two different data sets. The first data set, named "Itaipu", was acquired over Itaipu (Brazil/Paraguay border) and was composed of a 3-channel 4-m resolution IKONOS image (1999×1501). It presented seven classes, including vegetated and urban/built-up land covers (Table 1). This image is a challenging classification problem because it exhibits both spatially homogeneous classes (e.g.,"water") and very heterogeneous classes with regular structures and textures (e.g., "urban"). The second data set (named "Pavia") was composed of 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, affected by speckle and including two main classes, i.e., "dry soil" and "wet soil". Training and test fields were chosen to be spatially disjoint to minimize the correlation between training and test samples and the resulting possible bias in accuracy assessment. Additional details on "Itaipu" and "Pavia," can be found in [4], [5], respectively.

The experimental setup was as follows: (i) a Gaussian radial basis function kernel [1] was used for K; (ii) the Potts model was used for $\mathcal{E}_i$; (iii) a second order neighborhood system was used; (iv) the number of iterations for MSVC-GC was 20. Test areas were located inside image regions associated with the classes, without test samples at the spatial borders between the classes. This is common practice in remote sensing to prevent including mixed pixels in ground-truth data. The results of MSVC-GC were compared with those given by the following: (i) the classifier proposed in [4] that integrates the SVM and MRF approaches and uses ICM to solve the energy minimization problem (MSVC-ICM); (ii) a traditional non-contextual SVM; (iii) a classical MRF-based classifier, in which the Potts model was used and the pixelwise class posterior probabilities were computed according to a Gaussian model (MRF-Gauss) for "Itaipu" and by the "k-nearest neighbor" algorithm (MRF-k-NN) for "Pavia" [2]; (iv) a previous approach to SVM-MRF combination (MRF-SVM-Post), consisting of an MRF-based classifier, in which the Potts model was used and the pixelwise class posterior probabilities were approximated by the method described in [6]. Details on the optimization of the parameters of these techniques can be found in [4].

MSVC-GC obtained very high classification accuracies on the test sets for all three data sets. The values of the overall accuracy (OA) were around 99% for "Itaipu," and 96% for "Pavia." (see Table 1).A visual analysis of the related classification maps makes it possible to also appreciate the behaviors of the considered methods in spatially critical image areas. In particular, a remarkable improvement of MSVC-GC over MSVC-ICM is pointed out by a visual analysis of the discrimination of the "urban" class (see Fig. 1). In a VHR image, this class is spatially heterogeneous and generally consists of

Table 1: Experimental results: training- and test-sample sizes; classification accuracies (on the test set) of MSVC-GC, MSVC-ICM, a non-contextual SVM, MRF-Gauss or MRF-k-NN, and MRF-SVM-Post.

data set	class	training samples	test samples	MSVC-GC	MSVC-ICM	SVM	MRF-Gauss / MRF-k-NN	MRF-SVM-Post
Itaipu	urban	1891	18982	98.38%	97.47%	58.27%	78.02%	84.76%
	herbaceous	1098	5546	100%	99.77%	92.28%	96.48%	99.40%
	shrub and brush	1891	48919	99.99%	100%	98.76%	99.81%	99.93%
	forest land	189	674	91.54%	100%	91.39%	99.41%	100%
	barren land	1148	3855	66.28%	72.37%	68.33%	75.88%	72.27%
	built-up land	1891	5267	99.77%	96.28%	93.45%	98.82%	99.85%
	water	1891	91546	99.99%	99.97%	99.95%	99.94%	99.95%
	overall accuracy			99.03%	98.98%	93.92%	96.85%	97.66%
	average accuracy			93.68%	95.12%	86.06%	92.62%	93.74%
Pavia	dry soil	6205	5082	96.64%	95.43%	92.84%	94.79%	94.06%
	wet soil	1346	1927	94.71%	98.24%	91.13%	96.11%	95.23%
	overall accuracy			96.11%	96.20%	92.37%	95.15%	94.38%
	average accuracy			95.68%	96.84%	91.99%	95.45%	94.65%

a collection of subclasses, which are associated with different objects, structures, or materials and often make it difficult to correctly discriminate "urban" as a unique land cover.

4. CONCLUSION

A novel contextual classification method has been proposed by integrating the SVM, MRF, and GC approaches. A formulation of the global posterior energy function corresponding to the application of an MRF-based classifier in the transformed space associated with a kernel function is developed. This formulation holds with respect to an arbitrary kernel and for a large family of spatial MRF models. A GC-based classifier is formalized in the resulting SVM-MRF integrated framework and the related parameters are automatically optimized through numerical techniques based on the Powell and Ho-Kashyap algorithms. The experimental results pointed out that the proposed method could achieve high classification accuracies on the test samples, thus suggesting the effectiveness of the proposed SVM-MRF-GC integrated approach, which outperforms previous classifiers based on either the SVM or the MRF approach. As compared to a previous integrated SVM-MRF classifier based on the ICM energy minimization approach, improvements were noted in terms of spatial behavior and especially of discrimination of urban areas in VHR imagery. These results may be interpreted as due to the integration of GCs which allow reaching a global or strong local minimum of the energy function associated with the considered MRF spatial model. Possible future extension could be the integration with multiscale or multiresolution MRFs [7].

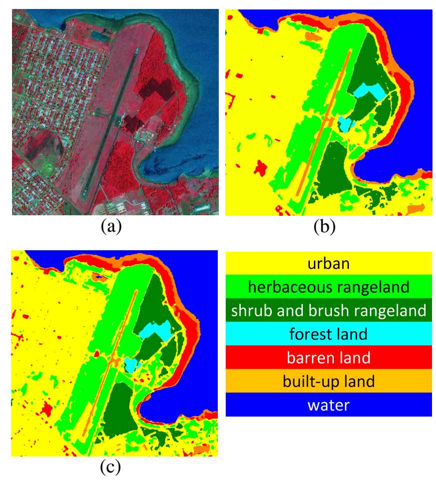

Fig. 1: "Itaipu" data set. Details of: (a) RGB false color composition; (b) map generated by MSVC-GC; (c) map generated by MSVC-ICM.

5. REFERENCES

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

[2] Z. Kato and J. Zerubia, "Markov random fields in image segmentation," *Foundations and Trends in Signal Processing*, vol. 5, pp. 1–155, 2011.

[3] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 11, pp. 1222–1239, 2001.

[4] G. Moser and S. B. Serpico, "Combining suppport vector machines and markov random fields in an integrated framework for contextual image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, pp. 2734–2752, 2013.

[5] 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.

[6] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, 2004.

[7] G. Moser, E. Angiati, and S. B. Serpico, "Multiscale unsupervised change detection on optical images by Markov random fields and wavelets," *IEEE Geosci. Remote Sensing Letters*, vol. 8, pp. 725–729, 2011.