SEGMENTATION OF REMOTE-SENSING IMAGES BY SUPERVISED TS-MRF

G. Poggi, G. Scarpa

Dip. di Ing. Elettronica e Telecomunicazioni
Università Federico II di Napoli
via Claudio, 21 – 80125 Napoli, Italy
{poggi,giscarpa}@unina.it

J. Zerubia

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles – BP 93
06902 Sophia Antipolis Cedex (France)
zerubia@sophia.inria.fr

ABSTRACT

In this work we specialize the recently proposed tree-structured MRF model to supervised segmentation of multispectral satellite images. This model allows a hierarchical representation of a 2-D field by means of a sequence of binary MRFs, each corresponding to a node in the tree. One can fit the intrinsic structure of the data to this tree-structured model, thereby defining a multi-parameter, flexible, MRF. Although a global MRF model is defined on the whole tree, optimization as well as estimation are carried out node by node, with a significant reduction in complexity. Experiments on a test SPOT image prove the superior performance of the algorithm w.r.t. other MRF-based or variational algorithms for supervised segmentation.

Key-words: Image classification, object-oriented segmentation, Markov random fields, hierarchical model.

1. INTRODUCTION

Segmentation is one of the most important low-level processing carried out on remote-sensing imagery, especially relevant for subsequent image classification and interpretation. In the statistical framework, the segmentation problem is approached by choosing suitable probabilistic models for both the data y and the unknown segmentation map x, and by estimating the map according to some useful statistical criteria. A popular choice is the *Maximum A Posteriori* (MAP) criterion, where x is selected as the map that maximizes the joint probability distribution $p(y|x)p(x)$. Since the likelihood term $p(y|x)$ can be easily modeled, the key problem in the MAP approach is the selection of a meaningful prior $p(x)$.

The Markov random field (MRF) model [1] is a relatively simple, yet effective, tool to encompass prior knowledge in the segmentation process, and in fact the interest on MRFs has been steadily growing in recent years. By modelling the segmentation map as a MRF, one assumes that each pixel depends statistically on the rest of the image only through a selected group of neighbors. This greatly simplifies the problem of assigning a meaningful prior, since only local characteristics of the image need be specified.

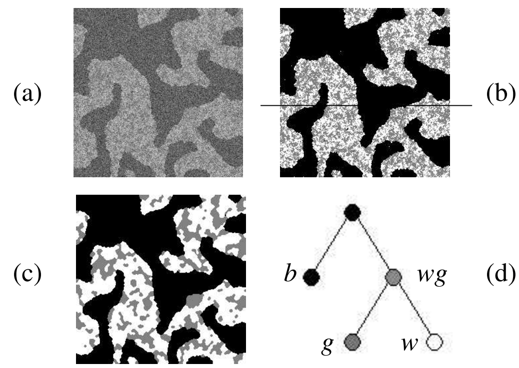


Fig. 1. Synthetic experiment: (a) noisy data; (b) hidden segmentation map; (c) segmentation by Ising model. (d) tree structure of the data: *b*, black class; *g*, grey class; *w*, white class; *wg*, white-grey merging class.

Many MRF models have been proposed [1] to describe the different features encountered in the images of interest. One of the most challenging (and largely unsolved) modelling tasks is to account for the spatial variations of image statistics. Let us consider for example the image in Fig.1, where the data (a) were generated by adding Gaussian noise to a synthetic segmentation map (b). As can be seen, following for example the traced line, the class random process is clearly non-stationary. Conventional models provide poor results in the presence of such "structured" data, like the segmentation map shown in (c), obtained with a second order Ising model. Here, one class is well detected while the other two are almost everywhere mixed because of an over-regularization. Even more sophisticated hierarchical MRFs (e.g., [2]) would not cope well with such data because their parameters do not depend on the classes. In fact, the correct model for this example is the binary tree of classes shown in Fig.1(d), where the "black" class is first separated from the other two based on a dedicated binary

0-7803-8554-3/04/$20.00 ©2004 IEEE.

MRF model, while the "white" and "gray" classes are further separated by means of a different binary MRF model.

This example, although unrealistic, makes clear the potential advantage of using such a structured model, called tree-structured MRF (TS-MRF) [3], since many real images present a similar hierarchical structure although not always so easy to detect. In fact, the TS-MRF model aims at describing the hidden structure of the data by a sequence of binary MRFs, each corresponding to a node in the tree. In such a model, parameters can be defined locally to each node, so as to correctly split the classes depending only on their statistical characteristics. In addition, thanks to its hierarchical structure, the TS-MRF model allows for recursive optimization procedures, which help reducing complexity.

This work addresses the problem of supervised image segmentation based on a TS-MRF model. We will use prior information to define a sensible tree structure, estimate all class parameters, and define a suitable likelihood term to couple with the TS-MRF prior. The algorithm performance will then be assessed by numerical experiments.

2. SUPERVISED TS-MRF METHOD

A random field X defined on a lattice $\mathcal{S}$ is said to be a MRF with respect to a given neighborhood system if the Markovian property holds for each site s. Quite often, in order to limit complexity, only the 4 or 8 closest pixels (system η^1 and η^2, respectively) are included in a pixel's neighborhood. Nonetheless, the MRF model proves quite powerful because most of the dependencies can be captured through local interactions.

The distribution of a positive MRF can be proved to have a Gibbs form [1], that is

$$p(x) = \frac{1}{Z}\exp[-U(x)] = \frac{1}{Z}\exp[-\sum_{c\in\mathcal{C}} V_c(x_c)] \qquad (1)$$

where the $V_c(\cdot)$ functions are called potentials, $U(x)$ denotes the energy, Z is a normalizing constant, and c indicates a clique of the image. Note that each potential V_c depends only on the values taken on the clique sites $x_c = \{x_s, s \in c\}$ and, therefore, accounts only for local interactions. As a consequence, local dependencies in X can be easily modeled by defining suitable potentials $V_c(\cdot)$.

In the following developments we will focus on the Ising model because of its simplicity, but the reader should be aware that the TS-MRF can be based on any binary model, no matter how complex. The Ising model can be defined both on η^1 and η^2 neighborhoods, and only two-site cliques are taken into account, on which a potential is defined as

$$V_c(x_c) = \begin{cases} \beta & \text{if } x_p \neq x_q, \quad p,q \in c \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

where $\beta > 0$ is the "edge-penalty" parameter.

In such a model, as well as other commonly used MRFs, the potential function of a two-site clique depends only on the presence or absence of an edge (a class transition), without regard to involved classes. As a consequence one parameter is sufficient to define the model. If we remove such a constraint and use a different potential function for each different pair of classes, many more parameters become necessary, $\frac{1}{2}(K-1)K$ where K is the number of classes. In this generalized Ising model, the clique potentials will become

$$V_c(x_c) = \begin{cases} \beta_{kh} & \text{if } x_p = k \neq x_q = h, \quad p,q \in c \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

where $\beta_{kh} = \beta_{hk} > 0$, $k \neq h$, is the edge-penalty parameter for a transition $h - k$.

Starting from this very general model, it is possible to reduce the number of parameters by taking into account some class properties. Looking at the example of Fig.1, it is easy to realize that the white (*w*) and gray (*g*) classes have the same relationship with the black class (*b*), that is, a *b*-*w* edge has the same statistical properties of a *b*-*g* edge. Therefore, only two independent parameters should be used instead of three. Such class properties can be easily represented by means of the hidden binary structure of Fig.1(d). In the example considered, the two relevant parameters are associated with the two internal nodes of the tree. The former, associated with the root, penalizes *b*-*w* and *b*-*g* edges (assumed to be equivalent), while the latter, associated with the other internal node, penalizes *w*-*g* edges. Therefore, at the root level, where *b* is split from *w* and *g*, it does not matter if a site is labeled as *w* or *g*, but only if it falls into the macro-class *wg*. In general, given a binary structure with K terminal nodes, the number of internal nodes, and hence the number of parameters, will be $K-1$, rather than $\frac{1}{2}(K-1)K$. Such a small number of parameters can be reliably estimated from the available data but, at the same time, allow one to accurately describe class-dependent nonstationarities in the image.

We now proceed to a more formal description of the model. Let us consider a binary tree T, identified by its nodes and their mutual relationships, and composed of internal nodes $\overline{T}$ and terminal nodes $\tilde{T}$. Also, let $\wedge$ be a binary operator which gives the nearest common ancestor of two nodes of T. We can define a tree-structured MRF through its clique potentials, still expressed by (3), but with the additional $\frac{1}{2}K(K-3)+1$ constraints that $\beta_{kh} = \beta_{pq} = \beta_t$ if $k \wedge h = p \wedge q = t$

$$V_c(x_c) = \begin{cases} \beta_{x_p \wedge x_q} & \text{if } x_p \neq x_q, \quad p,q \in c \\ 0 & \text{otherwise.} \end{cases} \qquad (4)$$

Then, the distribution of the TS-MRF becomes simply

$$p(x) = \frac{1}{Z}\exp[-\sum_{t\in\overline{T}} \beta_t \mathcal{N}_t] = \prod_{t\in\overline{T}} \frac{1}{Z_t}\exp[-\beta_t \mathcal{N}_t]. \qquad (5)$$

where $\mathcal{N}_t = \mathcal{N}_t(x)$ is the number of cliques with edge-penalty β_t.

The complexity of this model could still seem prohibitive for a practical implementation because of the dimensionality of the parameter space, dependent on the number of classes. However, thanks to its structural constraints, a recursive optimization procedure can be used which, although sub-optimal, involves only one edge-penalty at a time.

Once given a binary tree structure T which relates hierarchically the classes, the label field x can be completely expressed through the set of the binary fields x^t associated with the internal nodes of T. An example is shown in Fig.4, where the 8-class label field is obtained through a binary field, associated with the tree root, that separates class 8 from the others, another field, associated with the left child of the root, that separates class 7 form the remaining ones, and so on. Note that, but for the root, each field is defined on an irregular lattice resulting from the realization of the binary field associated to the parent node. Therefore, each binary field X^t depends only on its ancestor fields $X^{\omega(t)} = \{X^t\}_{t\in\omega(t)}$, where $\omega(t)$ denotes the set of ancestors of t. On the other hand, the number $\mathcal{N}_t$ of cliques with edge-penalty β_t depends only on x^t and, for the above considerations, on $x^{\omega(t)}$, while it is independent of other component binary fields, that is $\mathcal{N}_t = \mathcal{N}_t(x^t, x^{\omega(t)})$.

Hence, in order to find a MAP estimate of a segmentation with TS-MRF prior, we can maximize recursively the terms in (5), together with the corresponding likelihoods, starting from the root and descending the tree until all leaves are reached. Each term depends only on a single binary field once its ancestor fields are given, and can be maximized by itself with conventional techniques like simulated annealing, ICM, etc. Note, again, that in the step corresponding to node t, only the parameter β_t must be estimated, and that $\mathcal{N}_t$ is a sufficient statistic for β_t.

Of course, to carry out the maximization, we also need the likelihood term $p(y^t|x^t, x^{\omega(t)})$ for each node t, where Y^t is the set of data coupled with X^t. As usually done, let us assume the data to be conditionally independent, so we only need to define the terms $p(y^t_s|x^t_s, x^{\omega(t)}_s)$, where x^t_s is equal either to the left, $l(t)$, or right, $r(t)$, child of t. When a child is a terminal node, the corresponding likelihood is known *a priori* because we are in a supervised context, and in particular it is assumed here to be Gaussian with known parameters. However, the child could be itself an internal node, in which case no obvious likelihood is available. In this situation, we define the likelihood term as the best matching Gaussian among all its descendant leaves, namely

$$p(y^t_s|x^t_s, x^{\omega(t)}_s) = \max_{k\in\gamma(x^t_s)} p(y_s|x_s = k) \qquad (6)$$

where $\gamma(t)$ is the set of descendant leaves of t, and $x^t_s \in \{l(t), r(t)\}$. In other words, to decide if the current site should be assigned to the left or right node, the best two Gaussian distributions corresponding to "true" classes are considered, one being the most likely in $\gamma(l(t))$, the other in $\gamma(r(t))$. This way, the tree-structure involves only the prior MRF model while no structural constraints are transferred on the likelihood term $p(y|x)$.

Note that the best fitting Gaussian chosen at this point is only a temporary choice, taken to well fit the data during this intermediate split, but further splits, based on newly available contextual information, are free to change such a decision.

3. EXPERIMENTAL RESULTS

The supervised TS-MRF algorithm was tested on a SPOT satellite image provided by the Costel laboratory (University of Rennes 2). The image (see Fig.2) consists of three bands with different wavelengths in the visible spectrum and represents the Bay of Lannion in France in August 1997. A ground-truth, composed of a learning set for parameter estimation, and a test set for accuracy assessment, is also available.

Fig.3 shows the segmentation map obtained with the TS-MRF whose structure is drawn in Fig.4. Such a structure was chosen a priori based on the observation of the class statistics and their spatial relationships (on a preliminary segmentation). We compared the performance of the proposed method to that of other MRF based methods, the (non-structured) Ising and the H-MAP proposed in [2], as well as to the variational methods M1 and M2 proposed in [4]. To this end, we computed several global accuracy indicators derived from the confusion matrix: the global accuracy τ, the Kappa parameter κ, and the global accuracy τ^{norm} computed on the normalized confusion matrix. The results (see Tab.1) show that the TS-MRF method outperforms all reference techniques considered, for all indicators. This is true not only for global accuracy indicators but also in terms of class-wise accuracy as results from the confusion matrices, not reported here for brevity.

In particular, all methods exhibit poor results for meadows, due to a very high overlap between temporary and permanent meadows, but the TS-MRF classifier proves much better than the other ones. For example, the user's accuracies for temporary and permanent meadows are 48,9% and 26,8% respectively with TS-MRF, while they drop to 32,0% and 21,3% for the H-MRF, the best reference method, and a similar behaviour is observed for the producer's accuracy. This remarkable case makes clear the benefits of the proposed tree-structured model: these two classes are always kept together as a single macro-class during the initial splits, and only at the last node they are separated, based on locally estimated parameters, with higher accuracy.

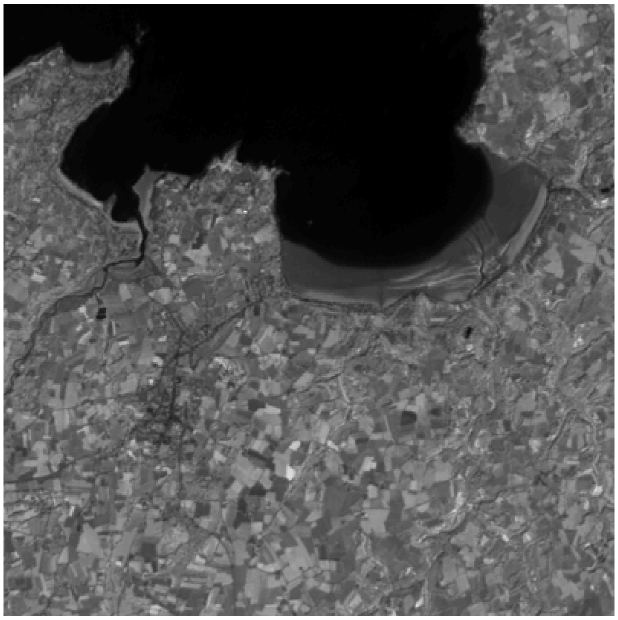


Fig. 2. SPOT image: channel XS3.

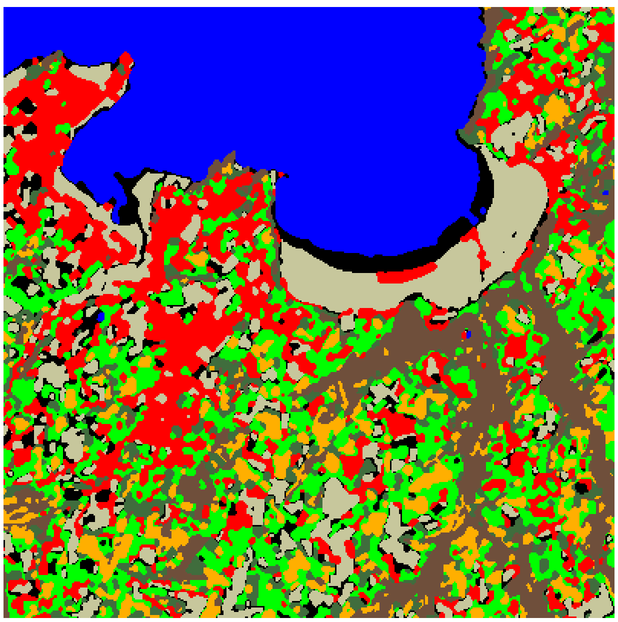
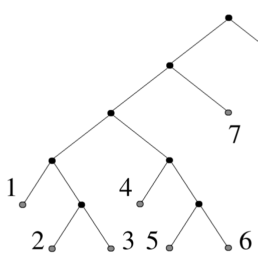


Fig. 3. TS-MRF segmentation map.



1-Corn
2-Temporary meadows
3-Permanent meadows
4-Urban Areas
5-Sand and bare soils
6-Vegetables
7-Forests and heath
8-Sea and water

Fig. 4. Tree structure of the model.

Methods	ISING	H-MRF	M1	M2	TS-MRF
τ	66,0%	70,5%	71,3%	70,0%	73,3%
κ	58,8%	64,2%	not av.	not av.	67,4%
τ^{norm}	45,3%	50,1%	not av.	not av.	52,8%

Table 1. Summary of the global accuracy indicators.

4. CONCLUSIONS AND FUTURE WORK

In this work we have presented a tree-structured MRF prior model useful for image segmentation and classification. Such a model allows to generalize any "blind" MRF model whose clique potentials depend only on the presence/absence of edges to a corresponding "non-blind" model where also the edge-colors (classes) are taken into account. The proposed model requires $M \times (K-1)$ parameters, with M the number of parameters in the basic blind model and K the number of classes. The imposed structural constraints between classes, however, allow for the recursive factorization of the joint distribution in $K-1$ terms, each one corresponding to an internal node of the tree and involving only a sub-set of parameters. As a consequence, fast recursive optimization and estimation procedures can be used.

In the supervised classification of a real remote-sensing image TS-MRF has shown fully satisfactory results. We point out, however, that the choice of the tree, which should fit the "intrinsic" structure of the data, is still an open problem, and certainly deserves further investigation.

5. ACKNOWLEDGMENTS

We would like to thank the GSTB, the COSTEL laboratory, University of Rennes II, the VISTA project of IRISA Rennes (F) for providing the SPOT data, the ground-truth and the results of comparative classifications. We also thank the Italian MIUR and the EU for their support through PRIN and IMAVIS projects, respectively.

6. REFERENCES

[1] S.Z.Li, *Markov random field modeling in image analysis*, Springer-Verlag, 2001.

[2] A. Chardin, P. Pérez, "Unsupervised Image Classification with a Hierarchical EM Algorithm," *Proc. ICCV'99*, Vol.2, pp.969-974, Kerkyra, Greece, Sept. 1999.

[3] C.D'Elia, G.Poggi, G.Scarpa, "A Tree-Structured Markov random field model for Bayesian image segmentation," *IEEE Transactions on Image Processing*, vol.12, no.10, pp.1259-1273, Oct. 2003.

[4] C.Samson, L.Blanc-Féraud, G.Aubert, J.Zerubia, "Two variational models for multispectral image classification," *Proc. EMMCVPR'01*, Sophia Antipolis, France, Sept.3-5, LNCS 0302-9743, Springer-Verlag, 2001.