Supervised Segmentation of Remote Sensing Images Based on a Tree-Structured MRF Model

Giovanni Poggi, Giuseppe Scarpa, and Josiane B. Zerubia, *Fellow, IEEE*

***Abstract*—Most remote sensing images exhibit a clear hierarchical structure which can be taken into account by defining a suitable model for the unknown segmentation map. To this end, one can resort to the tree-structured Markov random field (MRF) model, which describes a K-ary field by means of a sequence of binary MRFs, each one corresponding to a node in the tree. Here we propose to use the tree-structured MRF model for supervised segmentation. The prior knowledge on the number of classes and their statistical features allows us to generalize the model so that the binary MRFs associated with the nodes can be adapted freely, together with their local parameters, to better fit the data. In addition, it allows us to define a suitable likelihood term to be coupled with the TS–MRF prior so as to obtain a precise global model of the image. Given the complete model, a recursive supervised segmentation algorithm is easily defined. Experiments on a test SPOT image prove the superior performance of the proposed algorithm with respect to other comparable MRF-based or variational algorithms.**

***Index Terms*—Hierarchical fields, image classification, image segmentation, Markov random fields (MRFs), regression trees, structured images.**

I. 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. Segmentation techniques based on clustering in the observation space, like minimum distance, maximum likelihood, and the like [1], are well understood and exhibit a low complexity, but usually do not guarantee a satisfactory performance. To improve accuracy, one has to take into account prior information about the observed image, especially its spatial structure and the dependencies between neighboring pixels.

In the statistical framework, this is done by defining a suitable joint probabilistic model $p(x, y) = p(x)p(y|x)$ for the unknown segmentation map x and the data y and by estimating the map according to some useful statistical criteria. A popular choice is the maximum *a posteriori* probability (MAP) criterion, where x is selected as the map that maximizes the joint probability distribution. The key point becomes the selection of suitable models for the conditional data likelihood $p(y|x)$ and the prior $p(x)$.

Manuscript received May 25, 2004; revised April 7, 2005.

G. Poggi is with the Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni, Università Federico II di Napoli, 80125 Napoli, Italy (e-mail: poggi@unina.it).

G. Scarpa is with the Pattern Recognition Department, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 14131 Prague, Czech Republic.

J. B. Zerubia is with the Unité de Recherche de Sophia Antipolis, l'Institut National de Recherche en Informatique et en Automatique, 06902 Sophia Antipolis Cedex, France.

Digital Object Identifier 10.1109/TGRS.2005.852163

Although data modeling is an important problem and the object of intense research, e.g., see [2]–[4], here we focus on the prior term and will therefore follow the common simplifying assumption that the data be conditionally independent given the class, and Gaussian distributed, leaving for future work the refinement of this hypothesis. As for the prior $p(x)$, it should be meaningful and present a limited complexity for analytical and numerical treatment. The Markov random field (MRF) model [5] meets all these requirements because it is a relatively simple, yet effective, tool to encompass prior knowledge in the segmentation process. By modeling 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 through the definition of local potential functions whose integration gives the global energy of a Gibbs distribution [5].

Many MRF models have been proposed (see [6] and [7]) to describe the different features encountered in the images of interest, often with remarkable results in the applications. Very few models, however, try to capture the local variations of image statistics, which represent one of the most relevant phenomena to account for. Instead, it is implicitly assumed that interregion dependencies can be governed by a single set of parameters, to be estimated based on data from the whole image. However, images are highly structured sources of information, and this oversimplification can lead to major errors.

Consider for example the image in Fig. 1, where the data (a) were generated by adding Gaussian noise to a synthetic segmentation map (b). This image exhibits different statistical behaviors in different areas, with labels changing rapidly in some regions and much more slowly in others. A segmentation based on conventional nonadaptive models of the data provides very poor results, such as the segmentation map shown in (c), obtained with a second-order Potts model, where the two fine-grain classes are badly mixed because of an overregularization. The problem, here, is that the model parameters are estimated on the whole image irrespective of the different statistics encountered in different regions. Spatially adaptive models (e.g., [8]) could certainly improve upon this result but they are quite cumbersome and, more important, do not address the real problem, which is that a *class-adaptive* model is needed here. In particular, to fit well this image, a more complex model is required, such as the hierarchy of classes shown in Fig. 1(d): here a first set of parameters (just one parameter in the example) describes the relationship between the "black" class and the other two, which are indistinguishable at this level, while a second set of parameters describes the relationship between the "white" and "gray" classes.

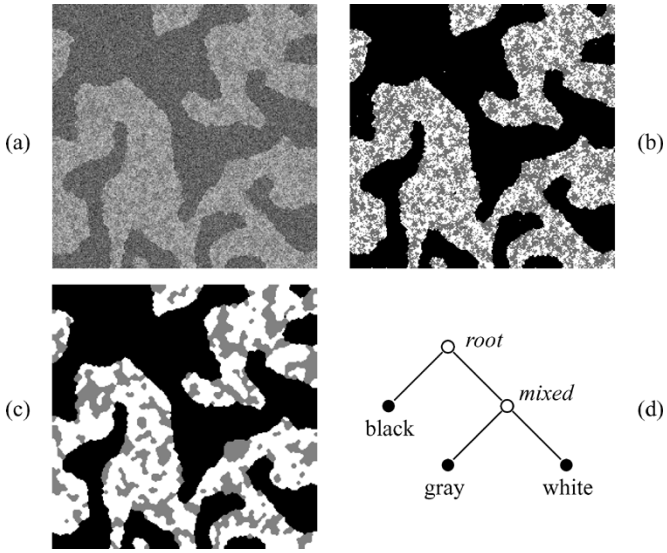

Fig. 1. Synthetic experiment. (a) Noisy data. (b) Hidden segmentation map. (c) Segmentation by Potts model. (d) Tree structure of the data: *b*, black class; *g*, gray class; *w*, white class; *wg*, white-gray merging class.

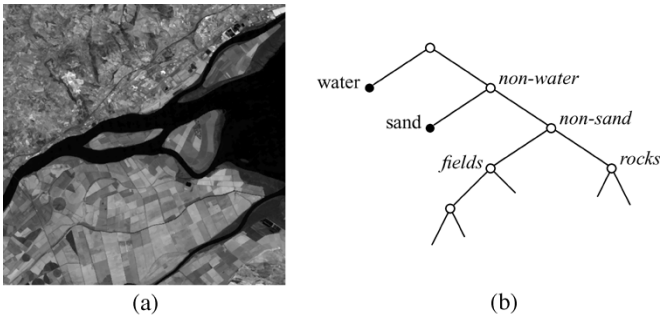

Fig. 2. (a) Landsat image with (b) corresponding tree structure.

This toy example has been built ad hoc to highlight the potential advantage of such a structured model, but many remote sensing images present a similar hierarchical structure although not always so easy to detect. The image of Fig. 2(a), for example, also could, and probably should, be described by a hierarchical model. A possible tree structure is shown in Fig. 2(b): the "water basins" have very smooth and long contours, no matter what other regions they border, so they can be singled out immediately. Similar considerations apply for the "sand/river banks" class. The remaining classes could then be grouped in "fields" and "rocks" macroclasses, and these could be further refined in other layers of the hierarchy.

Based on these observations, we have proposed in [9], and later refined in [10], a hierarchical MRF model called tree-structured MRF (TS–MRF). Such a model aims at describing the hidden structure of the data by a sequence of binary MRFs, each corresponding to a node in the associated tree, with all parameters defined locally to each node. In [9] and [10], the development of the TS–MRF model was strictly functional to its application to unsupervised segmentation. The resulting segmentation algorithm, thanks to its recursive nature and to the underlying structured model, was both accurate and efficient, even for large, complex images. However, because of the unsupervised framework, some of the modeling power of the hierarchical approach was sacrificed to the need of addressing a cluster validation problem by estimating the underlying tree structure itself, together with the problem of estimating the class-conditional data distributions. The lack of any prior knowledge led us to consider only elementary modeling tools as building blocks of the TS–MRF, and forbid a thorough statistical characterization of the problem.

In this work, we consider the supervised framework, where both the number of classes present in the image and their statistical characteristics are supposed to be known. This frees us from the need to "build" the tree of classes from scratch and allows us to focus more in depth on the modeling task. In particular, we have two major goals. First, we want to investigate the theoretical aspects of the method so as to explore its description capability as well as the limitations due to the superimposed structure and the recursive optimization procedure. In addition, we aim at generalizing the method by defining the TS–MRF "class," i.e., a whole set of MRF models that can be derived from well-established basic MRFs by introducing suitable structural constraints, and that present all the advantages discussed throughout this work. In such a class, also models that integrate different basic MRFs are included, so that data with complex structures can be handled effectively. For example, in the presence of a road network, a specific model [11] aimed at preserving fine structures can be used to describe the roads whereas more conventional models treat the other regions.

Another remarkable difference between unsupervised and supervised TS–MRF-based algorithms concerns the likelihood model of the data. In the former case, in fact, an approximate likelihood model had to be used at each node because it was not known, at segmentation-time, how many and which classes belonged to each subtree. On the contrary, the supervised algorithm is always fully aware of which classes are involved in each node segmentation and can therefore resort to a rigorous compound model of the data and carry out a global recursive optimization based on a conditioned best-matching component selection. Note that such a likelihood modeling can be extended to an unsupervised context as well, by iterating the optimization over the tree once a first, maybe simple, segmentation is carried out and an initial tree structure is built.

In next section, we will describe the TS–MRF model, its generalization, and its recursive properties that lead to a fast and adaptive segmentation algorithm. In Section III, the algorithm will be specialized to a simple case (all Ising models, which are binary Potts models), and its performance will be assessed by numerical experiments and compared with those of competing techniques. Finally, Section IV will draw conclusions.

II. Supervised Segmentation Based on the TS–MRF Model

A random field X defined on a lattice $\mathcal{S}$ is said to be a MRF with respect to a given neighborhood system η if the Markovian property holds for each site s, i.e.,

$$p(x_s|x - x_s) = p\left(x_s|x_{\eta(s)}\right) \tag{1}$$

where $\eta(s)$ indicates the neighbors of s. The distribution of a positive MRF can be proven [6] to have a Gibbs form

$$p(x) = \frac{1}{Z}\exp\left[-U(x)\right] = \frac{1}{Z}\exp\left[-\sum_{c\in\mathcal{C}} V_c(x_c)\right] \tag{2}$$

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, namely, a set of neighboring sites. 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, depends only on 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 Potts model because of its simplicity, but the reader should be aware that the TS–MRF can be based on a wide variety of binary[1] models, no matter how complex. The Potts 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

$$V_c(x_c) = \begin{cases} \beta, & \text{if } x_p \neq x_q, \quad p, q \in c \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

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), but not on the individual classes involved. As a consequence, only one parameter is needed to define the model no matter how many classes are present. If we remove such a constraint, and use a different potential function for each different pair of classes, many more parameters become necessary, $(1/2)(K-1)K$ where K is the number of classes. In this generalized Potts model, the clique potentials will become

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

where $\beta_{hk} = \beta_{kh} > 0$ 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*), i.e., 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. This kind of class properties can be easily represented by means of the hidden tree 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. In general, given a binary tree structure with K terminal nodes, the number of internal nodes, and hence the number of parameters, will be $K-1$, rather than $(1/2)(K-1)K$. Such a small number of parameters can be reliably estimated from the available data and, at the same time, allow one to accurately describe class-dependent behaviors 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 Λ. 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 (4), but with the additional constraints that $\beta_{hk} = \beta^t$ whenever $h \wedge k = t$

$$V_c(x_c) = \begin{cases} \beta^{h \wedge k}, & \text{if } x_p = h \neq k = x_q, \quad p, q \in c \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Then, the distribution of the TS–MRF becomes simply

$$p(x) = \frac{1}{Z} \exp\left[-\sum_{t \in \overline{T}} \beta^t N^t \right] = \prod_{t \in \overline{T}} \frac{1}{Z_t} \exp[-\beta^t N^t] \tag{6}$$

where $N^t = N^t(x)$ is the number of cliques with edge-penalty β^t.

The complexity of this prior model, with its $K-1$ different edge-penalties, could still seem high with respect to the reference Potts model, which is defined by just one edge-penalty parameter. However, thanks to its structural constraints, a simple recursive optimization procedure can be used which 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. In Fig. 2(b), for example, the seven-class label field is obtained through a binary field, associated with the tree root, that separates class 1 (water) from the others, another field, associated with the right child of the root, that separates class 2 (sand) from the remaining ones, and so on. Note that, but for the root, each field is defined on the irregular lattice resulting from the realization of the binary field associated with the parent node. Therefore, each binary field X^t depends only on its ancestor fields $X^{\omega(t)} = \{X^t\}_{t \in \omega(t)}$, which define the support of the field, where $\omega(t)$ denotes the set of ancestors of t. On the other hand, the number 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, i.e., $N^t = N^t(x^t, x^{\omega(t)})$.

A. Recursive Properties

It is easy to prove [12] that, for each node t in the tree, given X^t and $X^{\omega(t)}$, the fields which lie on the left subtree stemming from t are independent from the fields which lie on the right subtree. As a consequence, we can also write

$$p(x) = \prod_{t \in \overline{T}} p\left(x^t | x^{\omega(t)}\right). \tag{7}$$

Also, for each internal node t, given the ancestor fields $X^{\omega(t)}$, the field built on the subtree rooted in t, is itself a TS–MRF. As a consequence, given $X^{\omega(t)}$, the terminal binary fields X^t (associated with terminal splits) are Ising MRFs, i.e.,

$$p\left(x^t | x^{\omega(t)}\right) = \frac{1}{Z\left(x^{\omega(t)}\right)} \exp[-\beta^t N^t]. \tag{8}$$

Note, however, that this property does not hold for nonterminal binary fields, because, in this case, the partition function

[1]As a matter of fact, the restriction to *binary* models is itself a useful but not strictly necessary simplification.

Z is itself a function of x^t. In other words, not all the terms in (7) are Ising distributions, as one could believe for the similarity between (6) and (7). Nonetheless, in order to find a MAP estimate of a segmentation with TS–MRF prior probability, one can recursively maximize the terms in (6), together with the likelihood parts, starting from the root and descending the tree until all leaves are reached. Each term depends only on a binary field X^t once its ancestor fields $x^{\omega(t)}$ are given, and it does have an Ising form. As a consequence, each one can be maximized, just like with an ordinary Ising MRF, by using simulated annealing [5], ICM [7], etc. Note, again, that in the step corresponding to node t, only the parameter β^t must be estimated, and that N^t is a sufficient statistic for β^t. Therefore, when the prior parameters are unknown, they can be estimated by means of the EM algorithm [13] or the more general ICE [14], following again a recursive schedule.

B. Posterior Model and Optimization

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, we assume the data to be conditionally independent, so that we only need to define the terms $p(y_s^t|x_s^t, x_s^{\omega(t)})$, where x_s^t 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 descendent leaves, namely

$$p\left(y_s^t|x_s^t, x_s^{\omega(t)}\right) = \max_{k\in\gamma(x_s^t)} p(y_s|x_s = k) \tag{9}$$

where $\gamma(t)$ is the set of descendent leaves of t, and $x_s^t \in \{l(t), r(t)\}$. 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 among the left descendents $\gamma(l(t))$, and the other among the right descendents $\gamma(r(t))$. 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.

In order to clarify the whole algorithm we summarize it below. Since we are focusing on the supervised segmentation, several information are assumed to be known *a priori*: the number of classes K, the parameter set $\{\mu_k, \Sigma_k\}_{k=1,\ldots,K}$ defining the Gaussian likelihood components and, finally, a K-leaves binary-tree structure T whose $K-1$ internal nodes are indexed by $\overline{T} = \{1, \ldots, K-1\}$, with any ordering such that $t > n$ if $n \in \omega(t)$. The segmentation $\widetilde{x} = \{\widetilde{x}^t\}_{t\in\overline{T}}$ of data y, together with the estimation of the prior parameter set $\beta = \{\beta^t\}_{t\in\overline{T}}$, are then performed by the following general algorithm.

Algorithm 1 (Recursive optimization over TS-MRF)

1. *initialization*: $t = 1$, $\mathcal{S}^t = \mathcal{S}$;
2. *fitting*: $\forall s \in \mathcal{S}^t$ *compute* $p(y_s^t|x_s^t = l(t), \widetilde{x}_s^{\omega(t)})$ *and* $p(y_s^t|x_s^t = r(t), \widetilde{x}_s^{\omega(t)})$;
3. *optimization*: *find the estimates of* x^t *and* β^t *as*

$$(\widetilde{x}^t, \widetilde{\beta}^t) = \arg\max_{(x^t,\beta^t)} p\left(x^t|\widetilde{x}^{\omega(t)}, \beta^t\right) p\left(y_t|x_t, \widetilde{x}^{\omega(t)}\right). \tag{10}$$

4. *segmentation*: $\mathcal{S}^{l(t)} = \{s \in \mathcal{S}^t : \widetilde{x}_s^t = l(t)\}$, $\mathcal{S}^{r(t)} = \{s \in \mathcal{S}^t : \widetilde{x}_s^t = r(t)\}$,
   if $t < K-1$ *then* $t = t+1$ *and goto step 2. else exit.*

In particular, concerning step 3, we adopted the following partial optimal solution that iteratively alternates the estimates of x^t and β^t.

1. *set* $\beta_{(0)}^t = 0$, *compute* $\widetilde{x}_{(0)}^t$ *by* (10), *set* $k = 1$;
2. *compute the* MPL *estimate* $\widetilde{\beta}_{(k)}^t$ *on the current realization* $\widetilde{x}_{(k-1)}^t$;
3. *compute* $\widetilde{x}_{(k)}^t$ *defined in* (10) *by* ICM, *with the last* $\widetilde{\beta}_{(k)}^t$ *estimated;*
4. *if* $\widetilde{x}_{(k)}^t \approx \widetilde{x}_{(k-1)}^t$ *then set* $\widetilde{x}^t = \widetilde{x}_{(k)}^t$ *and exit, else set* $k = k+1$ *and goto 2.*

C. Model Generalization

To derive a TS–MRF model from the binary Potts model, we started from the K-ary Potts model, generalized it by introducing edge-dependent parameters $\{\beta_{kh}\}$, and finally introduced some structural constraints in the parameter space. More generally, it is possible to build a TS–MRF by integrating several reference MRF models, not necessarily of the same type nor just binary. To this end, let $T = \{\overline{T}, \Lambda\}$ be a generic tree with K leaves $k \in \Lambda$ and M internal nodes $t \in \overline{T}$, each with K_t offsprings and associated with a K_t-ary MRF model defined by

$$p_t(x^t, \mathcal{S}^t, \theta^t) = \frac{1}{Z_t(\mathcal{S}^t, \theta^t)} \exp\left[-U_t(x^t, \mathcal{S}^t, \theta^t)\right], \qquad t \in \overline{T} \tag{11}$$

where $\mathcal{S}^t$ is the support lattice for the field x^t associated with node t, and θ^t is the corresponding parameter set.

Once fixed the tree structure, the number of leaves will be then

$$K = 1 + \sum_{t\in\overline{T}} (K_t - 1) \tag{12}$$

while a K-class field x on a lattice $\mathcal{S}$ can be represented through the set of field components $\{x^t\}_{t\in\overline{T}}$, each defined recursively on a lattice $\mathcal{S}^t \subseteq \mathcal{S}$ shaped by the ancestor field components $x^{\omega(t)} = \{x^k\}_{k\in\omega(t)}$. Based on these notations and constraints, a general TS–MRF will take the form

$$\begin{aligned} p(x, \theta) &= \frac{1}{Z(\theta)} \exp\left[-U(x, \theta)\right] \\ &= \frac{1}{Z(\theta)} \exp\left[-\sum_{t\in\overline{T}} U_t\left(x^t, x^{\omega(t)}, \theta^t\right)\right] \end{aligned} \tag{13}$$

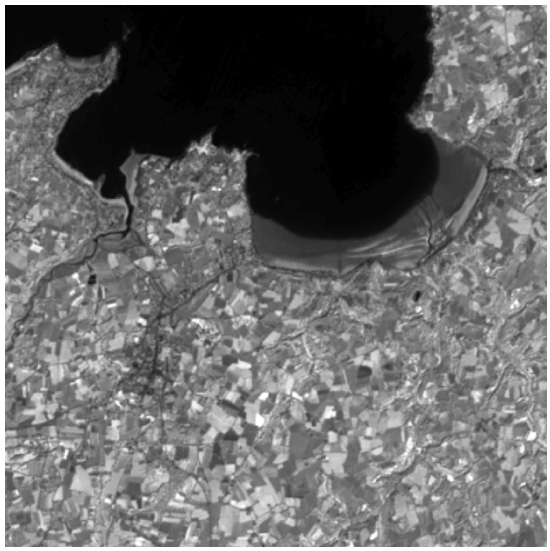

Fig. 3. SPOT image of Lannion Bay (400 × 400). August 15, 1997.

with $\theta \equiv (\theta^1, \theta^2, \ldots, \theta^M)$ and, for the above considerations, $x^{\omega(t)}$ in place of $\mathcal{S}^t$.

The TS–MRF model is therefore a very general and flexible class of models, allowing one to use complex component models with high-order cliques (e.g., the Chien model [11]), for a specific subset of classes, and simpler models for the remaining classes. This could be useful, for example, when some classes are highly correlated and very hard to tell apart. In such a case only an high-order interaction MRF model can provide good results, but such a model would be computationally intractable when defined on the whole image and for all the classes into it. On the contrary, we can use such model only for a terminal split of an ad hoc TS–MRF (the split of such highly correlated classes). By doing so, the recursive optimization of the TS–MRF will have a reduced computational burden on almost all the tree except for the limited region where the complex model is used.

III. Experimental Results

To assess the performance of the proposed TS–MRF-based supervised segmentation algorithm, we experimented with a Systeme Pour l'Observation de la Terra (SPOT) satellite image, composed of three 1480 × 1024-pixel bands in the visible spectrum, a 400 × 400-pixel section of which (band 3) is shown in Fig. 3. The image represents the Bay of Lannion in France in August 1997 and was used by the geographers of the Costel laboratory (University of Rennes II, France) in the context of remote sensing researches. The goal of their study was to determine the land cover of the area, and to this end they built a list of eight classification categories: sea and water, sand and bare soil, urban areas, forests and heath, temporary meadows, permanent meadows, vegetables, and corn. Thanks to both photointerpretation and visits on-site, they selected samples of all eight categories in the multispectral SPOT image. Some of them (learning set) were used to estimate the mean spectral response and the interband covariance matrix of each category, while the remaining samples (test set) were kept to assess the accuracy of various classifications techniques. Both the training and the test sets corresponding to the image of Fig. 3 are shown in Fig. 4 (top and bottom respectively).

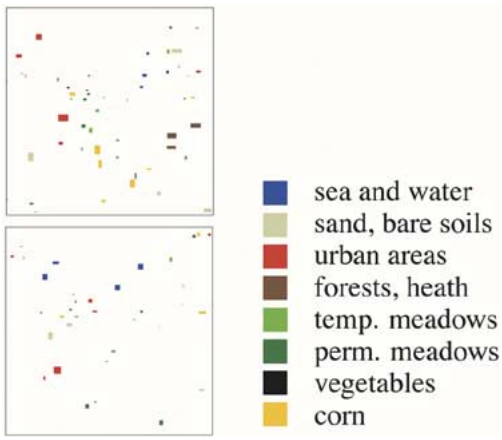

Fig. 4. Ground-truth, (top) training and (bottom) test set.

A. Accuracy Assessment Method

The accuracy of both the proposed and some reference segmentation methods has been assessed by analyzing the confusion matrix for the test set. Recall that the entry on ith row and jth column of this matrix, say a_{ij}, is the number of sample pixels from jth class that have been classified as belonging to the ith class.

Various indicators are derived from this matrix. First, two error measures can be computed for each class: the user's accuracy of class i is defined as a_{ii}/a_{i+}, where a_{i+} is the ith row marginal (sum of row entries); conversely, the producer's accuracy of this class is defined as a_{ii}/a_{+i}, where a_{+i} is the ith column marginal.

Beside these two class-based parameters, three global quality indicators are also computed. The overall accuracy of the method, defined as $\tau = \sum_i a_{ii}/N$, is the percentage of sample pixels that are correctly classified. Another common indicator is the so-called Kappa parameter, defined as $\kappa = (N\sum_i a_{ii} - \sum_i a_{i+}a_{+i})/(N^2 - \sum_i a_{i+}a_{+i})$, which discounts successes obtained by chance and is therefore more conservative (it can be also negative). Finally, in order to give the same weight to all classes' contributions to the accuracy, irrespective of the number of samples in each one, the confusion matrix can be normalized with the iterative proportional fitting algorithm [15], so that all column and row marginals sum up to unity. The overall accuracy τ^{norm} computed on such a modified matrix is called normalized accuracy.

B. Reference Techniques

Several researchers working on segmentation have used the SPOT data described above and therefore we can benefit from their numerical results to make a performance comparison [16]–[19]. We collected results from two different experiments, one performed on the whole image, and another performed on the 400 × 400-pixel section of Fig. 3.

The reference methods are as follows:

- MD: minimum distance;
- ML: maximum likelihood;
- DA: discriminant analysis;
- ICM: Potts MRF optimized by ICM;
- H-MAP (H-MPM): hierarchical MRF model optimized by MAP (MPM) criteria [16], [17];
- M1X and M2X: two variational methods [18], [19].

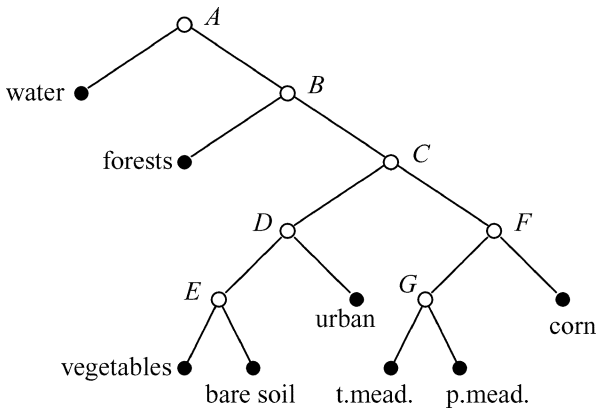

Fig. 5. Tree structure of the model manually set.

MD, ML, and DA are well known spectral segmentation techniques, while ICM is a conventional MRF-based contextual segmentation algorithm. Such techniques were used as reference in Chardin's Ph.D. thesis [16], where a multiresolution hierarchical MRF model was proposed providing two implementation: H-MAP and H-MPM. Later on, two variational methods M1X and M2X were proposed by Samson *et al.* [19] and compared again with ML, ICM, and H-MAP experimenting on the small SPOT image.

C. Implementation of the Proposed Algorithm

All the reference algorithms cited above, as well as the TS–MRF-based algorithm proposed here, are supervised, meaning that the number of classes is known in advance and the associated parameters are estimated on the learning set. For the proposed algorithm, however, we also have to select one of the possible binary trees whose leaves will correspond to the known classes. This is an important choice, and as remarked in the introduction, the selected tree should reflect the intrinsic structure of the data, with the classes grouped together, hierarchically, according to their spectral features and mutual spatial relationships. In the case of unsupervised TS–MRF segmentation, this problem is solved by testing the "split-gain" parameter associated with all nodes of the tree [10], and similar ideas could be probably used here; but we leave this delicate topic for future investigation and resort instead to a simple inspection of the data. The tree structure selected for the SPOT image of Fig. 3 is shown in Fig. 5.

We remark, also, that all the reference algorithms treat the three spectral bands of the SPOT image as uncorrelated, namely, with a diagonal covariance matrix for each class. This is a rough approximation, and our proposed algorithm makes use of the full class covariance matrices. However, to establish a fair comparison with the reference techniques, a version based on diagonal covariance matrices was also developed. From now on, these two algorithms will be referred to as TS/C and TS/U, respectively.

The classwise means and standard deviations for each of the three bands employed in the TS/U algorithm, and recovered from Chardin's thesis [16], are gathered in Table I.

The description of the TS algorithm is completed by specifying that all binary segmentations were carried out by means of an estimation–maximization procedure similar to that proposed in [20] but with ICM in place of simulated annealing, while the edge-penalty parameters were estimated by the maximum-pseudolikelihood technique proposed in [21].

TABLE I
CLASSWISE STATISTICS: MEAN μ AND STANDARD DEVIATION σ FOR EACH BAND

	water	forests	vegetab	b.soil	urban	t.mead	p.mead	corn
	76,2	67,1	87,2	103,3	91,0	80,8	78,4	74,0
μ	30,4	30,4	51,6	71,4	50,9	42,2	39,3	35,4
	21,7	108,8	111,4	98,1	93,9	130,0	142,6	147,8
	11,1	3,0	6,2	10,4	8,2	4,1	4,3	2,7
σ	4,6	2,5	9,4	10,1	6,6	5,0	4,8	2,5
	13,9	16,7	31,7	10,4	15,6	19,5	22,7	10,6

D. Analysis of the Results

The first experiment discussed here was carried out on the whole SPOT image. Global accuracy indicators, reported in Table II for all proposed and most reference techniques, show some significant regular patterns.

First of all, contextual techniques regularly outperform noncontextual ones, as was to be expected, with the only contrasting figure of the τ^{norm} indicator for MD, which will be justified later on. Second, the TS/C technique, which exploits the actual covariance matrices of the data, exhibits by far the best performance, but this was also easily predictable since the other techniques neglect important information. Last, and by our point of view most interesting, TS/U outperforms all other techniques working with diagonal covariances. The advantage is just marginal with respect to H-MPM (but for τ^{norm}) but consistent. Moreover, the most useful comparison is with the ICM technique, which can be considered as the unstructured counterpart of our TS/U: in this case, all indicators show a significant improvement, strengthening the case of the hierarchical approach.

These synthetic data are certainly revealing, but global accuracy indicators hide deep differences of behavior between the classifiers, so we now proceed to a detailed class-based examination of results. In Tables III–VI, we report the confusion matrices for ML, ICM, TS/U, and TS/C, with the associated classwise producer's and user's accuracies.

For a few classes, namely water, bare soil, and forests, all techniques provide very high accuracy, with the TS/C on top, so we will not pay attention to them anymore. The vegetable class is also a clear-cut situation: the training and test data are so scarce that a reliable classification is nearly impossible, and probably this class should not have been considered at all. The only technique that provides a nonzero user accuracy is the MD, which is why its *normalized* global accuracy is higher than expected.

Let us now turn to more interesting cases, starting with the urban areas. For this class, ML accuracies are rather low, about 66% and 60%, and ICM improves only marginally these figures up to 70% and 66%, respectively. TS/U segmentation, instead, performs dramatically better, bringing the accuracies to about 77% and 89%, respectively. The most remarkable phenomenon, here, is the decrease of the number of urban areas pixels that are classified as temporary meadows from 96 (20%) to 17 (4%). The residual errors are due to confusion of urban areas and bare soil classes, which indeed have similar spectral responses.

TABLE II
SUMMARY OF THE GLOBAL ACCURACY INDICATORS FOR THE CLASSIFICATIONS PERFORMED ON THE WHOLE IMAGE

Methods	DA	MD	ML	ICM	H-MAP	H-MPM	TS/U	TS/C
τ	78,8%	75,9%	79,3%	81,1%	81,6%	83,5%	84,1%	86,5%
κ	73,8%	70,5%	74,3%	76,5%	77,1%	79,5%	80,2%	83,2%
τ^{norm}	55,7%	60,7%	55,2%	55,3%	56,3%	52,7%	57,9%	60,8%

TABLE III
CONFUSION MATRICES FOR ML CLASSIFICATION PERFORMED ON THE WHOLE IMAGE

ML	water	forests	vegetab	b.soil	urban	t.mead	p.mead	corn	Total	User's acc
water	**528**	3							531	99,4%
forests	11	**1520**				2		26	1559	97,5%
vegetab	16	4	**0**	53	30	12	9		124	0%
b.soil				**1316**	49				1365	96,4%
urban				109	**283**	21	12	2	427	66,3%
t.mead		2	2	9	97	**138**	110	10	368	37,5%
p.mead		7	3	3	9	119	**65**	36	242	26,9%
corn		55		1		98	198	**432**	784	55,1%
Total	555	1591	5	1491	468	390	394	506	5400	
Prod.'s acc	95,1%	95,5%	0%	88,3%	60,5%	35,4%	16,5%	85,4%		

TABLE IV
CONFUSION MATRICES FOR ICM CLASSIFICATION PERFORMED ON THE WHOLE IMAGE

ICM	water	forests	vegetab	b.soil	urban	t.mead	p.mead	corn	Total	User's acc
water	**527**	2							529	99,6%
forests	12	**1534**						26	1572	97,6%
vegetab	16	5	**0**	27	10	11	3		72	0%
b.soil				**1368**	48				1416	96,6%
urban		1		88	**310**	19	17	4	439	70,6%
t.mead			5	4	96	**141**	109	5	360	39,2%
p.mead		4		3	4	117	**61**	31	220	27,7%
corn		45		1		102	204	**440**	792	55,6%
Total	555	1591	5	1491	468	390	394	506	5400	
Prod.'s acc	94,9%	96,4%	0%	91,8%	66,2%	36,2%	15,5%	86,9%		

TABLE V
CONFUSION MATRICES FOR TS/U CLASSIFICATION PERFORMED ON THE WHOLE IMAGE

TS/U	water	forests	vegetab	b.soil	urban	t.mead	p.mead	corn	Total	User's acc
water	**527**	1							528	99,8%
forests	13	**1518**						27	1558	97,4%
vegetab	15	11	**0**	44	17	19			106	0%
b.soil				**1343**	18			1	1362	98,6%
urban		1		94	**416**	7	17	2	537	77,5%
t.mead		1	5	6	17	**221**	117	10	377	58,6%
p.mead		11		4		66	**63**	30	174	36,2%
corn		48				77	197	**436**	758	57,5%
Total	555	1591	5	1491	468	390	394	506	5400	
Prod.'s acc	94,9%	95,4%	0%	90,1%	88,9%	56,7%	16,0%	86,2%		

TABLE VI
CONFUSION MATRICES FOR TS/C CLASSIFICATION PERFORMED ON THE WHOLE IMAGE

TS/C	water	forests	vegetab	b.soil	urban	t.mead	p.mead	corn	Total	User's acc
water	**543**	3							546	99,4%
forests	12	**1539**						26	1577	97,6%
vegetab			**0**	42	15	28	25	1	111	0%
b.soil				**1375**	10				1385	99,3%
urban		12		61	**443**	29	15	6	566	78,3%
t.mead		10	5	10		**259**	90	39	413	62,7%
p.mead		2		2		7	**123**	69	203	60,6%
corn		25		1		67	141	**365**	599	60,9%
Total	555	1591	5	1491	468	390	394	506	5400	
Prod.'s acc	97,8%	96,7%	0%	92,2%	94,7%	66,4%	31,2%	72,1%		

TABLE VII
OVERALL ACCURACY τ OF EACH SINGLE BINARY SEGMENTATION FOR ML, ICM, TS/U, AND TS/C ALGORITHMS

Node	A	B	C	D	E	F	G
ML	99,4	98,0	94,4	89,8	96,1	71,6	47,0
ICM	99,4	98,3	94,8	92,1	98,1	71,7	47,2
TS/U	99,5	98,0	97,6	93,3	96,8	74,2	60,8
TS/C	99,7	98,4	96,2	95,6	97,0	72,8	79,7

TABLE VIII
TS–MRF EDGE-PENALTY PARAMETERS β^t ESTIMATED FOR TS/U AND TS/C

Node	A	B	C	D	E	F	G
TS/U	1,78	0,94	1,24	2,98	0,71	1,19	3,00
TS/C	1,54	0,99	1,59	0,97	0,86	1,31	3,00

Something similar happens to the two meadows classes. Their accuracy indicators are always quite low, even when the correct covariance matrices are used, because of their significant reciprocal spectral overlap, and their further overlap with the corn class. Nonetheless, the use of the TS approach raises significantly the accuracy, especially for the temporal meadows (from 39% to 59% and from 36% to 56%) but also, to a minor extent, for the permanent meadows (from 28% to 36% for the user's accuracy). As happened with the urban class, the reason for such improvement is the decrease in temporary meadows pixels that are classified as permanent meadows. In both these cases, there has been a significant improvement in telling apart closely related classes, and this has been achieved by using a dedicated model for their segmentation. In fact, both the urban areas/bare soil and the meadows couples have been kept together during early splits and segmented only well down the tree, as shown in Fig. 5.

To further clarify this point, in Table VII we report the overall accuracy τ of each single binary segmentation of the TS/U and TS/C algorithms. The first row, for example, refers to the split (node A in Fig. 5) between the water class and all the remaining classes, for which TS/U and TS/C exhibit an overall accuracy of 99.5% and 99.7%, respectively, as results from the confusion matrices of Tables V and VI. For the ML and ICM techniques we also report the same figures, which are not related to a real binary split but nonetheless measure the correct separation between the same sets of classes, and allow us to better compare the relative accuracies of the techniques and to point out the major differences. For the first two splits (either real or virtual), concerning water and forests, all techniques perform equally well. Instead, going to the split between urban areas and bare soil on one side, and corn and meadows on the other side (node C), TS/U turns out to be much more accurate (2.4% of errors) than ICM (5.2%), and a similar behavior is observed for all the remaining nodes, especially for the temporary/permanent meadows split (node G).

The first two columns of Table VII allow us to make a further observation. In fact, they show that the water and forests classes, which account for almost half the test set, are so well characterized that all techniques provide almost perfect results on them. So, if we cleansed results from their contribution, the improvement of TS/U with respect to the other contextual techniques in terms of global indicators would grow even larger.

In Table VIII, we report the edge-penalty parameters $\{\beta^t\}_{t\in\overline{T}}$ for TS/U and TS/C estimated during the recursive optimization procedure. It can be noticed that, in a few cases, such parameters are clearly overestimated, as is the case of β^D for TS/U and β^G for both TS/U and TS/C. Such a behavior is likely due to the MPL estimator, which has been shown [11] to have a tendency to overestimate when working on small datasets, as is the case for splits that operate on small terminal regions. Nevertheless, such inaccurate estimates do not affect the global performance of the TS algorithms since, as shown in the previous section, the proposed recursive optimization decouples the local prior at each node from those of the descendents. Furthermore, even with such a bad estimate of the local parameter, the corresponding classification accuracy still outperforms (even abundantly in the case of node G) that of the reference techniques (see Table VII). This is again due to the recursive segmentation which keeps together two highly correlated classes, temporary and permanent meadows, until they are finally split with local field components.

TABLE IX
SUMMARY OF THE GLOBAL ACCURACY INDICATORS FOR THE CLASSIFICATIONS PERFORMED ON THE SMALL IMAGE

Methods	ML	ICM	H-MAP	M1X	M2X	TS/U	TS/C
τ	65,1%	66,0%	70,5%	71,3%	70,0%	73,3%	75,8%
κ	57,9%	58,8%	64,2%	–	–	67,5%	70,2%
τ^{norm}	51,5%	45,3%	50,1%	–	–	52,8%	54,5%

We carried out a second experiment on the 400 $\times$ 400 pixel section of the whole image, in order to compare performance also with the variational methods proposed in [18] which were tested on this image. Table IX reports the results of TS/U, TS/C, and some of the other reference techniques, in terms of the three global indicators, together with those of M1X and M2X, drawn from [18] and [19], available only in terms of global accuracy. Again, TS/U, and of course TS/C, outperform all other techniques, and all the considerations made above still apply.

To conclude this section, we report in Fig. 6 the classification maps provided by ML, ICM, TS/U, and TS/U+ (a modified TS/U algorithm to be discussed later) for the 400 $\times$ 400 pixel section of Fig. 3. As was to be expected, besides being less accurate, the ML map is very fragmented because of noise and mixed cells, and it is difficult to use for further processing. On the contrary, all contextual techniques provide maps that can be easily interpreted, with an increasing degree of homogeneity (and accuracy) going from ICM to TS/U+.

It is worth spending some words on the TS/U+ technique, because in this case we take advantage, although to a very limited extent, of the modeling power of the TS–MRF approach. In fact, by inspection of the TS/U map of Fig. 6(c), we noticed the presence of a large number of isolated points or thin lines classified as "vegetables" (black) at the boundary of regions of other classes, especially the "bare soil" (gray). To counter this phenomenon, due the presence of mixed boundary cells, we adopted, only in nodes D and E, a modified Potts model [22] which discourages boundary fragmentation. As a result, this problem is not present anymore in the TS/U+ map of Fig. 6(d). Such an improvement does not correspond to a significant increase in the global accuracy indicators (less than 0.3%) but this is likely due to the absence of boundary cells in the test set, which is composed of homogeneous, unconnected, rectangular regions.

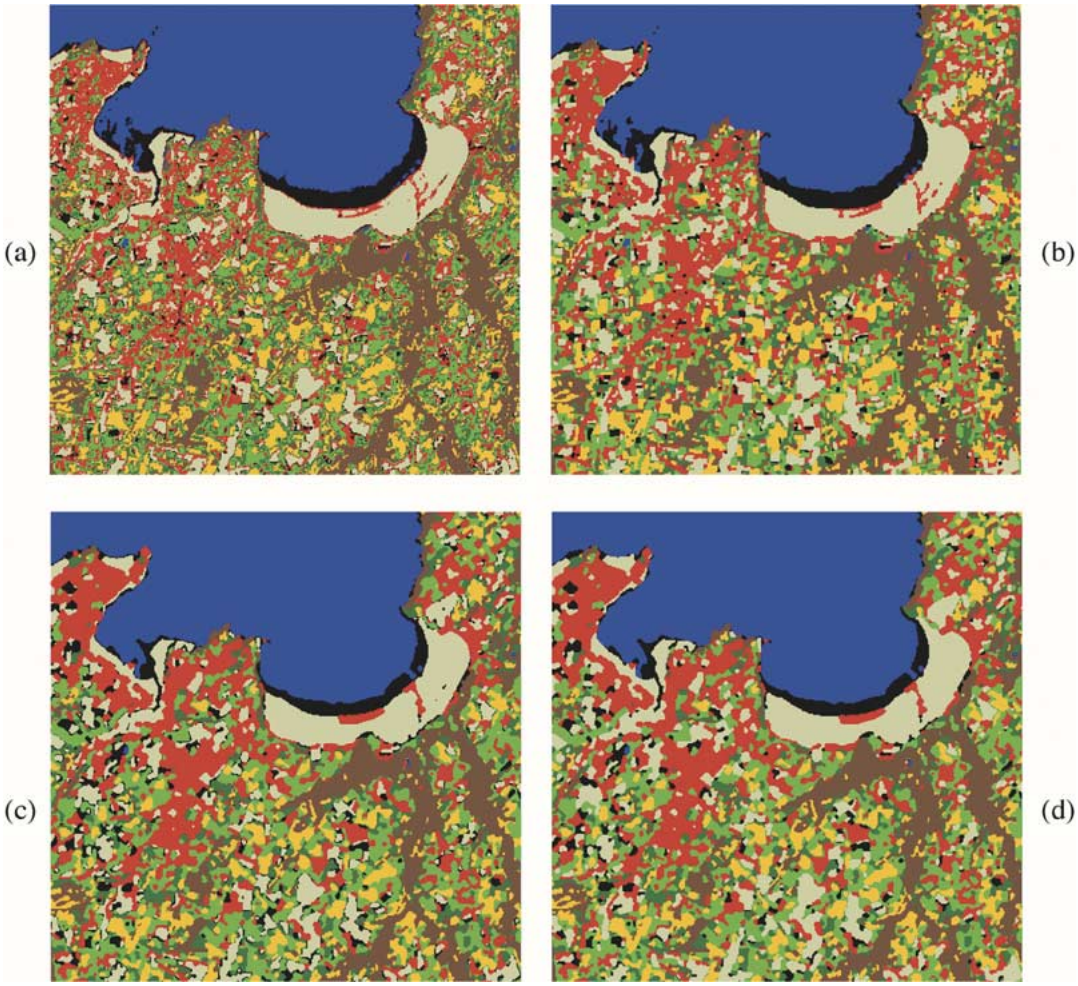

Fig. 6. Classifications of the SPOT image by (a) ML, (b) ICM, (c) TS/U, and (d) TS/U+.

E. Tree Structure Building

The TS–MRF class is quite general, and to define the exact TS–MRF to be used in a given application a specific tree structure must be selected together with all the component MRF models at the nodes. The final goal is to devise a fully automatic technique which, given the available information, identifies the best tree structure and chooses the most suitable component MRFs in a large bank of models. Of course, this is a very complex task, and is in fact one of the main topics to be addressed in future research. In this work, we were only interested in proving the potential of the TS–MRF approach in a supervised context and hence we limited ourselves to a simple case, where only binary Potts models are used, and the tree structure is selected manually. Nonetheless, to provide some insight on the problem of selecting the tree structure (but not the node models) we consider here an automatic procedure and compare results with those of the former ad hoc solution.

The proposed procedure builds the tree from the bottom, by associating the elementary classes with the leaves of the tree, and then iteratively merging couples of nodes. As pointed out before, with the TS–MRF approach we can separate highly correlated classes, difficult to split, by isolating them from the others and associating them to sibling nodes, so that a dedicated MRF can take charge of the task. Our conjecture, therefore, is that classes which are similar in terms of spectral response, or have a high degree of spatial adjacency, or have some other similar features, should be kept as "close" as possible in the tree. In the bottom-up perspective, this means that such classes should be merged first. The key point of the procedure is therefore a measure that quantifies class similarity and hence the goodness of a merge.

To this end, we resort to a probabilistic measure named *merging gain*. This measure was originally introduced in [10], in an unsupervised context, to decide whether two classes had to remain isolated or be merged. It is defined as

$$M^t = \frac{p(y^t|\theta^t)}{p(y^t|\widetilde{x}^t)p(\widetilde{x}^t)} \tag{14}$$

where t is the candidate father of the two nodes under test (say t' and t''), and $\widetilde{x}^t$ is the realization of the MRF at node t which segments the region associated with t returning those associated with t' and t''. The merging gain is large when the classes are spectrally close because the single-node likelihood $p(y^t|\theta^t)$ is very close to likelihood $p(y^t|\widetilde{x}^t)$ computed after segmentation for the two separated classes. Moreover, it is large when the two regions under test have an active boundary because the realization of the father MRF $\widetilde{x}^t$ is not very likely. Of course, to compute the merging gain, a preclassification of the image is

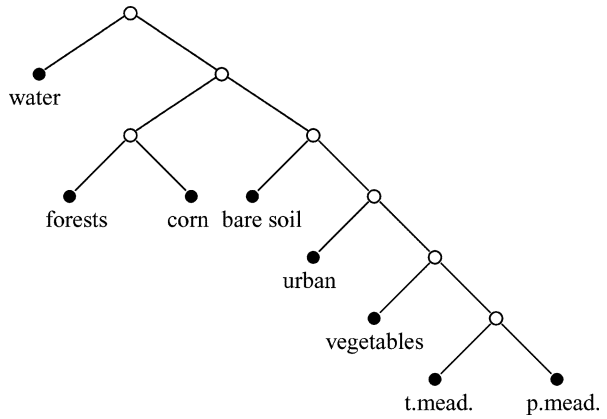

Fig. 7. Tree structure of the model automatically set.

needed, and here we used a simple maximum-likelihood classifier to this end.[2]

We used the proposed bottom-up algorithm to build an alternative tree structure for the experiment considered thus far. The resulting tree is shown in Fig. 7. With respect to the ad hoc tree, we notice some similarities (the water class is immediately isolated from the others, the meadows classes are kept together till the end) but also many differences. We used this new tree with the TS/C technique and compared the results with those of the same technique with the manually selected tree, obtaining minor and inconsistent differences: the overall accuracy τ goes down from 86.5% to 85.4%, and the κ parameter from 83.2% to 81.9%, while the normalized accuracy τ^{norm}, on the contrary, grows from 60.8% to 61.5%.

This algorithm of course is only a first approach to the problem, and results are not easily interpreted; but it appears that a simple procedure like this one can already produce results that are quite close to those obtained through a careful supervision.

IV. Conclusion and Future Developments

We have proposed a new technique for the supervised segmentation of remote sensing images, which provides a superior performance with respect to a wide range of reference techniques. This is not really surprising because the proposed algorithm is based on a class of hierarchical models, the TS–MRFs, which have many more degrees of freedom than conventional models: their sensible use cannot but improve segmentation accuracy. On the other hand, it is worth underlining that although this model is much richer than conventional flat models the segmentation algorithm has a limited computational complexity, because all segmentations are binary.

In this work, we focused essentially on the generalization of the theoretical model and on the definition of a suitable likelihood term for the optimization. The few experiments carried out had the only goal of demonstrating the method potential by comparison with a similar contextual technique based on an unstructured model. The most promising feature of the model, however, is probably its high adaptability, which we have barely touched here. Given the wide variety of structures that appear in typical images it is very likely that the use of ad hoc MRF models at each node can improve performance dramatically. Another interesting problem, that we hope to investigate in future research, is the selection of the tree structure given the classes. In fact, it would be highly desirable to have an automatic procedure that finds the best (or at least a good) tree structure given maybe a simple initial ML classification.

[2]For further details on the merging gain, the reader is referred to [10].

Acknowledgment

The authors would like to thank the French Space Agency (CNES) for providing the SPOT data through the ISIS programme and the GSTB, the COSTEL laboratory, University of Rennes II, the VISTA project of IRISA Rennes (France) for providing the ground-truth and the results of comparative classifications. The authors also thank the Italian MIUR for its support through the PRIN program, and the EU for its support through the IMAVIS project (5th framework).

References

[1] R. O. Duda, D. G. Stork, and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 2000.

[2] B. S. Manjunath, G. M. Haley, and W. Y. Ma, "Multiband techniques for texture classification and segmentation," in *Handbook of Image and Video Processing*, A. Bovik, Ed. Orlando, FL: Academic, 2000, pp. 367–381.

[3] D. E. Melas and S. P. Wilson, "Double Markov random fields and Bayesian image segmentation," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 357–365, Feb. 2002.

[4] S. Derrode and W. Pieczynski, "Signal and image segmentation using pairwise Markov chains," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2477–2489, Sep. 2004.

[5] 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.

[6] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 2nd ed. Berlin, Germany: Springer-Verlag, 2001.

[7] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*, 2nd ed. Berlin, Germany: Springer-Verlag, 2003.

[8] 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.

[9] G. Poggi and A. R. P. Ragozini, "Image segmentation by tree-structured Markov random field," *IEEE Signal Process. Lett.*, no. 7, pp. 155–157, Jul. 1999.

[10] C. D'Elia, G. Poggi, and G. Scarpa, "A tree-structured Markov random field model for Bayesian image segmentation," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1259–1273, Oct. 2003.

[11] X. Descombes, R. D. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood," *IEEE Trans. Image Process.*, vol. 8, no. 7, pp. 954–963, Jul. 1999.

[12] G. Scarpa, G. Poggi, and J. Zerubia, "A binary tree-structured MRF model for multispectral satellite image segmentation,", Sophia Antipolis, France, INRIA Res. Rep. 5062, Dec. 2003. [Online]. Available: http://www-sop.inria.fr/ariana/BIBLIO/Year/2003.html.

[13] A. P. Dempster, N. M. Larid, and D. B. Rubin, "Mixtures densities, maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[14] F. Salzenstein and W. Pieczynski, "Parameter estimation in hidden fuzzy Markov random fields and image segmentation," *Graph. Models Image Process.*, vol. 59, no. 4, pp. 205–220, 1997.

[15] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.*, vol. 37, no. 1, pp. 35–36, 1991.

[16] A. Chardin, "Modèles énergétiques hiérarchiques pour la résolution des problèmes inverses en analyze d'images—Application à la télédétection," Ph.D. thesis, Univ. Rennes, Rennes, France, Jan. 2000.

[17] L. Hubert-Moy, A. Cotonnec, L. Le Du, A. Chardin, and P. Pérez, "A comparison of parametric classification procedures of remotely sensed data applied on different landscape units," *Remote Sens. Environ.*, vol. 75, no. 2, pp. 174–187, Feb. 2001.

[18] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia, "Deux modèles variationnels pour la classification d'images multispectrales," *Trait. Signal*, vol. 18, no. 5–6, pp. 345–367, 2001.

[19] ——, "Two variational models for multispectral image classification," in *Proc. EMMCVPR*. Berlin, Germany: Springer-Verlag, Sep. 3–5, 2001, Lecture Notes in Computer Science.

[20] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random field using simulated annealing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 8, pp. 799–813, Aug. 1989.

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

[22] C. D'Elia, G. Poggi, and G. Scarpa, "An adaptive MRF model for boundary-preserving segmentation of multispectral images," in *Proc. 11th Eur. Signal Processing Conf.*, Toulouse, France, Sep. 2002.

Giovanni Poggi was born in Naples, Italy, on February 17, 1963. He received the Dr.Eng. degree in electronic engineering from the University Federico II of Naples, Naples, Italy, in 1988.

He has been with the Department of Electronic and Telecommunication Engineering, University Federico II since 1990, first as a Researcher, until 1998, then as an Associate Professor and, since November 2002, as a Full Professor of telecommunications. His current research activity is in statistical signal processing and, in particular, compression, segmentation, and filtering of images and video, with special attention to remote sensing applications.

Dr. Poggi serves as Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING.

Giuseppe Scarpa was born in Pompei, Italy, on January 5, 1975. He received the Dr.Eng. degree in telecommunication engineering and the Ph.D. degree in electronic and telecommunication engineering from the University of Naples, Naples, Italy, in 2001 and 2005, respectively.

In 2003, he was awarded a Marie Curie scholarship and was a Visiting Student at the INRIA institute for six months. In 2004, he was awarded a post-doc grant from ERCIM, and since April 2005 he has been a Research Fellow in the Pattern Recognition Department, Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic, Prague. His research interests concern the segmentation and classification of images and the theory of Markov random fields.

Josiane B. Zerubia (S'78–M'81–SM'99–F'03) received the M.Sc. degree from the Department of Electrical Engineering at Ecole Nationale Supérieure d'Ingénieurs Electriciens de Grenoble, Grenoble, France, in 1981, and the Dr.Eng. degree, the Ph.D. degree, and the "Habilitation," all from the University of Nice, Sophia-Antipolis, France, in 1986, 1988, and 1994, respectively.

She is a Permanent Research Scientist at l'Institut National de Recherche en Informatique et en Automatique, (INRIA), Sophia Antipolis, since 1989. She has been Director of Research since July 1995. She was Head of a remote sensing laboratory (PASTIS, INRIA Sophia-Antipolis) from mid-1995 to 1997. Since January 1998, she has been in charge of a new research group working on remote sensing (ARIANA, INRIA-CNRS-University of Nice). She has been an Adjunct Professor at Sup'Aero (ENSAE), Toulouse, France, since 1999. Before, she was with the Signal and Image Processing Institute, University of Southern California (USC), Los Angeles as a Postdoctoral Researcher. She also worked as a Researcher for the LASSY (University of Nice and CNRS) from 1984 to 1988 and in the Research Laboratory of Hewlett Packard in France and in Palo Alto, CA, from 1982 to 1984. Her current research interest is image processing using probabilistic models or variational methods. She also works on parameter estimation and optimization techniques. She has been a member of the editorial board of the French Society for Photogrammetry and Remote Sensing (SFPT) since 1998 and of the *International Journal of Computer Vision* since 2004.

Dr. Zerubia has been part of the IEEE BISP Technical Committee (SP Society) since 2005. She was Co-Chair of two workshops on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'01, Sophia Antipolis, France, and EMMCVPR'03, Lisbon, Portugal), Co-Chair of a workshop on Image Processing and Related Mathematical Fields (IPRM'02, Moscow, Russia), Chair of a workshop on Photogrammetry and Remote Sensing for Urban Areas, Marne La Vallee, France, 2003. She was part of the IEEE IMDSP Technical Committee (SP Society) from 1997 to 2003, Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING from 1998 to 2002, Guest Coeditor of a special issue of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE in 2003, member-at-large of the Board of Governors of IEEE SP Society from 2002 to 2004. She has been area editor of IEEE TRANSACTIONS ON IMAGE PROCESSING since 2003.