Image Segmentation by Tree-Structured Markov Random Fields

Giovanni Poggi and Arturo R. P. Ragozini, *Student Member, IEEE*

***Abstract*— We propose a new algorithm, based on a tree-structured Markov random field (MRF) model, to carry out the unsupervised classification of images. It presents several appealing features: Due to the MRF model, it takes into account spatial dependencies, yet is computationally light because only binary MRF's are used and a progressive refinement of information takes place. Moreover, it is adaptive to the local characteristics of the image and provides useful side information about the segmentation process.**

***Index Terms*—Image classification, Markov random field.**

I. Introduction

IN UNSUPERVISED image classification, neither the number of classes present in the scene nor their statistical characteristics are known *a priori.* They must be determined as part of the classification procedure, usually by means of an iterative process where pixel classification and parameter estimation alternate until convergence. If simple pixel-by-pixel techniques are used, based on clustering in the feature space, this procedure has a reasonable computational complexity. Such techniques, however, completely neglect the spatial dependencies among neighboring pixels and usually provide results of unsatisfactory quality. Therefore, context-based techniques, such as those based on Markovian models or texture analysis, are often preferred, with the undesirable side effect of a huge increase in complexity.

Techniques based on Markov random field (MRF) models [1], in particular, provide high-quality classification but, in order to account for the interaction among neighboring pixels, converge very slowly to the desired solution. In the framework of unsupervised classification, things get only worse since, given the number of classes m, the convergence process must be repeated many times, in order to gradually upgrade the class parameter estimates. Then, the whole procedure is repeated for each m ranging in a suitable interval, in order to single out the optimal value of m, and the associated class parameters and image classification (e.g., [2]).

In this letter, we propose a tree-structured segmentation algorithm where an MRF model is adopted for each node of the tree. The input image, associated to the root of the tree, is segmented initially in just two classes, as described above; the two subimages composed of same-class pixels are then associated to the children nodes, and each of them is segmented again using a dedicated binary MRF. Such a recursive process then proceeds down the (unbalanced) tree until a suitable stopping condition is met.

This approach presents several appealing properties. First of all, the use of binary classifications only, together with the tree structure, guarantees a huge reduction of the processing time, while the MRF model still keeps into account the spatial dependencies exhibited in the image. As a matter of fact, the use of local binary MRF's even allows for a finer adaptation to the local characteristics of the image than usually happens with unstructured MRF's.

Moreover, to each node (class, subimage) of the tree meaningful pieces of information are associated, such as the gain related with the possible split of that node, the degree of homogeneity of that subimage, etc., which help gaining insight into the segmentation results [3]. This also suggests using this tree-structured segmentation tool interactively, with the analyst that, on the basis of such information and by visual inspection, decides each time whether to continue or stop the development of the tree and also which node to split. Finally, tree-structured segmentation can be an intermediate step for other algorithms, such as class-adaptive image compression [4].

In the following, we recall briefly the basic of MRF segmentation, present the mathematical definition of a tree-structured MRF, describe the proposed segmentation algorithm, and comment some experimental results.

II. Segmentation Based on MRF

Let $\mathcal{S} = \{s = (i, j)\}$ be a two-dimensional (2-D) lattice. Let a clique c be a subset of $\mathcal{S}$ in which each site is neighbor of all remaining sites (we will consider only cliques formed by couples of eight-connected sites), and let $\mathcal{C}$ be the set of all cliques. Now, let $\mathbf{X} = \{X(s) : \ s \in \mathcal{S}\}$ be an MRF defined on $\mathcal{S}$: Its probability function will be a Gibbs distribution, namely[1]

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

where Z is a normalizing constant. $V_c(\cdot)$ is the potential function defined on the cliques; in particular, we will take

$$V_c(\mathbf{x}) = \begin{cases} \beta, & \text{if } x(s) = x(s'),\ s,\ s' \in c \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where β is an edge penalty parameter.

[1]We use the short notation $p(x)$ to denote $p(X = x)$ or even, for continuous variates, $f_X(x)$. Likewise, we use $p(x|y)$ in place of $p(X = x|Y = y)$.

Manuscript received October 10, 1998. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. R. M. Mersereau.

The authors are with the Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni, Università Federico II di Napoli, 21-80125 Naples, Italy (e-mail: poggi@unina.it).

Publisher Item Identifier S 1070-9908(99)04966-4.

The realization $x(s)$ of $X(s)$ is the class assigned to pixel s, and the goal of the segmentation process is to determine $\mathbf{x}$, the realization of the whole field, based on the related observables $\mathbf{y}$. We assume that, given its class, each observable is conditionally independent of the others, and has a Gaussian distribution with class-dependent parameters.

Assuming to know the number of classes m, and the class parameters $\mu_i, \sigma_i^2, i = 1, \cdots, m$, as well as the edge penalty parameter β, one can resort to a maximum *a posteriori* probability (MAP) approach and estimate $\mathbf{x}$ as $\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ where the maximization will be carried out in some suitable suboptimal manner given the computational intractability of the problem.

If β and the class parameters are unknown they must be estimated jointly with $\hat{\mathbf{x}}$. Finally, if the number of classes is unknown, one can let m vary in a suitable range $(m_{\min}, m_{\max})$ with any reasonable prior $p(m)$, and maximize on m as well [2].

III. Tree-Structured Segmentation Based on MRF

Let us consider a binary tree T, which we identify with a set of positive integers t, called nodes, where $t = 1$ is the root of the tree. Each node t, except the root, has a single parent $q(t)$, and each internal node $t \in \overline{T}$ has two children, $l(t)$ and $r(t)$, with $q(l(t)) = q(r(t)) = t$. Nodes with no children $t \in \tilde{T}$ are called *terminal nodes* or *leaves*.

We associate to each node t in T the following items: a set of sites $\mathcal{S}_t \subseteq \mathcal{S}$, a field of observables $\mathbf{Y}_t = \{Y(s)\colon s \in \mathcal{S}_t\}$ with realization $\mathbf{y}_t$, a binary MRF $\mathbf{X}_t = \{X_t(s)\colon s \in \mathcal{S}_t\}$ with realization $\mathbf{x}_t$, where $x_t(s) \in \{l(t), r(t)\}$, and a set of parameters $\{\mu_t, \sigma_t^2, \beta_t\}$ where β_t is the edge penalty parameter associated with $\mathbf{X}_t$.

Now, let $\mathcal{S}_1 = \mathcal{S}$, the original set of sites, and, for each internal node $t \in \overline{T}$, let $\mathcal{S}_{l(t)} = \{s \in \mathcal{S}_t\colon x_t(s) = l(t)\}$, and $\mathcal{S}_{r(t)} = \{s \in \mathcal{S}_t\colon x_t(s) = r(t)\}$, which is to say that each set of sites $\mathcal{S}_t$ is obtained from the partition of the parent set of sites $\mathcal{S}_{q(t)}$ according to the realization $\mathbf{x}_{q(t)}$ of the associated MRF. Given this condition, we call tree-structured (TS) MRF the set of all $\mathbf{X}_t$'s in $\overline{T}$, namely $\mathbf{X}_T = \bigcup_{t\in\overline{T}} \mathbf{X}_t$ with realization $\mathbf{x}_T = \bigcup_{t\in\overline{T}} \mathbf{x}_t$ such that $p(\mathbf{x}_T) = p(\mathbf{x}_t, t \in \overline{T})$.

If we accept a TS-MRF model for our image, a simple recursive segmentation procedure can be implemented. Indeed, once a realization $\mathbf{x}_T$ of the TS-MRF is obtained, each site has a label attached to it corresponding to a leaf of the tree or, in other words, the image is partitioned in the $|\tilde{T}|$ classes $\mathcal{S}_t, t \in \tilde{T}$. For each site in class $\mathcal{S}_t$ we assume, as said before, that the observables $\mathbf{Y}_t$ are independent Gaussian with parameters μ_t, σ_t^2 and realization $\mathbf{y}_t$.

Our goal is to find both the structure of the segmentation tree and the realization of the TS-MRF defined over it that maximize the posterior probability of τ and $\mathbf{x}$ given $\mathbf{y}$

$$(\hat{\tau}, \hat{\mathbf{x}}_{\hat{\tau}}) = \arg\max_{\tau} \left[\arg\max_{\mathbf{x}_\tau} p(\tau, \mathbf{x}_\tau|\mathbf{y}) \right]$$

where the tree T is now a random variable and τ one of its realizations. We propose here a greedy algorithm based on a recursive test of hypothesis to grow the tree one node at a time, and to select the associated realization of the MRF.

Suppose we have reached the tentative tree-structure τ and associated $\mathbf{x}_\tau$; to decide whether to split a leaf of τ and which one, we compute the quantities

$$\begin{aligned} A(\tau', \mathbf{x}_{\tau'}) &= p(\tau', \mathbf{x}_{\tau'}|\tau, \mathbf{x}_\tau, \mathbf{y}) \\ &= \alpha p(\mathbf{y}|\tau', \mathbf{x}_{\tau'}, \tau, \mathbf{x}_\tau) p(\mathbf{x}_{\tau'}|\tau', \tau, \mathbf{x}_\tau) \end{aligned} \quad (4)$$

where τ' is either equal to τ, or is obtained by τ through the split of one of its leaves. The factor α, neglected in the following, accounts for the irrelevant denominator $p(\mathbf{y}|\tau, \mathbf{x}_\tau)p(\tau, \mathbf{x}_\tau)$ and for the conditional probabilities $p(\tau'|\tau, \mathbf{x}_\tau)$ which, lacking any information or reasonable hypotheses, are assumed all equal.

If $\tau' = \tau$, and hence $\mathbf{x}_{\tau'} = \mathbf{x}_\tau$, we obtain simply

$$A(\tau, \mathbf{x}_\tau) = p(\mathbf{y}|\tau, \mathbf{x}_\tau) = \prod_{t\in\tilde{\tau}} p(\mathbf{y}_t|t) \quad (5)$$

where $p(\cdot|t)$ indicates that for each $s \in \mathcal{S}_t$, $Y(s)$ is Gaussian with parameters μ_t and σ_t^2.

On the other hand, when τ' is obtained by τ through the split of one of its leaves, say $t' \in \tilde{\tau}$, it results by construction $\mathbf{X}_{\tau'} = \mathbf{X}_\tau \cup \mathbf{X}_{t'}$ and $\mathbf{x}_{\tau'} = \mathbf{x}_\tau \cup \mathbf{x}_{t'}$, hence $p(\mathbf{x}_{\tau'}|\tau', \tau, \mathbf{x}_\tau) = p(\mathbf{x}_{t'})$ and

$$\begin{aligned} A(\tau', \mathbf{x}_{\tau'}) &= p(\mathbf{y}|\tau', \mathbf{x}_{\tau'}) p(\mathbf{x}_{t'}) \\ &= \prod_{t\in\tilde{\tau}, t\neq t'} p(\mathbf{y}_t|t) p(\mathbf{y}_{t'}|\mathbf{x}_{t'}) p(\mathbf{x}_{t'}) \end{aligned} \quad (6)$$

where $p(\cdot|\mathbf{x}_t)$ indicates that for each $s \in \mathcal{S}_t$, $Y(s)$ is Gaussian with parameters that depend on the realization $x_t(s)$ namely, $\mu_{x_t(s)}$ and $\sigma^2_{x_t(s)}$.

Now it is clear that the likelihood ratio

$$\Lambda(t', \mathbf{x}_{t'}) \triangleq \frac{A(\tau', \mathbf{x}_{\tau'})}{A(\tau, \mathbf{x}_\tau)} = \frac{p(\mathbf{y}_{t'}|\mathbf{x}_{t'})}{p(\mathbf{y}_{t'}|t')} \cdot p(\mathbf{x}_{t'}) \quad (7)$$

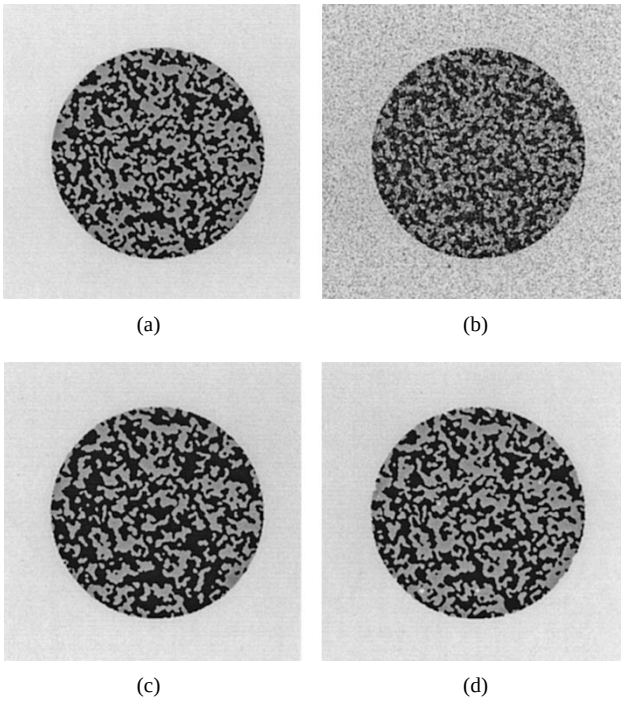
depends only on node t', and therefore can be easily maximized over $\mathbf{x}_{t'}$ alone. Let $\hat{\mathbf{x}}_t$ be the realization that maximizes $\Lambda(t, \mathbf{x}_t)$ and let us finally define the *split gain* associated to node $t \in \tilde{\tau}$ as $G(t) \triangleq \Lambda(t, \hat{\mathbf{x}}_t)$.

Now, the segmentation algorithm can be described unambiguously as follows.

1) Initialization. $\tau = \{1\}, \mathbf{X}_\tau = \emptyset$. $\mathcal{S}_1 = \mathcal{S}, \mathbf{Y}_1 = \mathbf{Y}, \mathbf{y}_1 = \mathbf{y}$. Estimate $\hat{\mu}_1, \hat{\sigma}_1^2$. Define $\mathbf{X}_1$ over $\mathcal{S}_1$ and find $\hat{\mathbf{x}}_1$ and $\hat{\beta}_1$. Evaluate $G(1)$.
2) Decision. $t' = \max_{t\in\tilde{\tau}} G(t)$. If $G(t') \leq 1$ go to step 4 else go to step 3.
3) Split. $\tau = \tau \cup l(t') \cup r(t')$, $\mathbf{X}_\tau = \mathbf{X}_\tau \cup \mathbf{X}_{t'}$, $\mathbf{x}_\tau = \mathbf{x}_\tau \cup \hat{\mathbf{x}}_{t'}$. $\forall t \in \{l(t'), r(t')\}$: single out $\mathcal{S}_t, \mathbf{Y}_t, \mathbf{y}_t$, estimate $\hat{\mu}_t, \hat{\sigma}_t^2$, define $\mathbf{X}_t$, find $\hat{\mathbf{x}}_t$ and $\hat{\beta}_t$, evaluate $G(t)$. Go to step 2.
4) Exit. Stop the algorithm with optimal tree $\hat{\tau} = \tau$ and optimal realization of the TS-MRF $\hat{\mathbf{x}}_{\hat{\tau}} = \mathbf{x}_\tau$.

IV. Experimental Results

To gain insight about the potentialities of the TS-MRF approach in image segmentation, we consider now a simple experiment. Fig. 1(a) shows a three-class synthetic image to



(a) (b)

(c) (d)

Fig. 1. Comparison of segmentation techniques. (a) Original image before adding white Gaussian noise. (b) Maximum likelihood segmentation, error rate 21.3%. (c) Segmentation based on a flat MRF model, error rate 4.2%. (d) Segmentation based on a TS-MRF model, error rate 2.3%.

which white Gaussian noise is added, so as to reach a signal-to-noise ratio of about 3 dB. Fig. 1(b) shows the result of maximum likelihood segmentation: the error rate is over 20%. Using a flat (unstructured) MRF model the result of Fig. 1(c) is obtained: the algorithm correctly detects the presence of three classes and the segmentation is much more accurate, with an error rate of about 4%. When a TS-MRF model is considered, the result of Fig. 1(d) is obtained: Again, the algorithm stops correctly at three classes, but the error rate is further reduced to about 2%. In particular, the inner circle is more accurately segmented and fine details are better preserved.

This behavior is readily explained by looking at the segmentation tree (Fig. 2) with all the significant parameters associated. The first split separates the inner circle from the rest of the image, and the two tightly intertwined inner classes are pooled together at this stage, even though their means are far

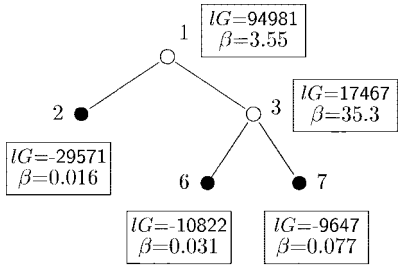


Fig. 2. Segmentation tree. For each node, the logarithm of the split gain $lG = \log G$ and the associated edge penalty β are reported. The root node 1 corresponds to the whole image, and is characterized by a large split gain which enables the separation of the outer region (node 2) from the inner circle (node 3). Node 2 is not splitted further because of its negative log-gain which indicates that the associated region should be considered homogeneous. The best binary MRF for this region has $\beta \simeq 0$ denoting the absence of any structure. Node 3, on the opposite, has large log-gain and large edge penalty, indicating that structures are present and should be segmented; note that the local edge penalty is much higher than that of node 1. Finally nodes 6 and 7 have parameters similar to those of node 2 and hence are considered terminal.

apart. This split is characterized by a large gain (the logarithm is reported) due to a large improvement in the data fitting, and a small increase in the model complexity ($\hat{\beta}_1 = 3.55$). A similar behavior is observed for node 3, which leads to the accurate separation of the two inner classes. Note that the edge penalty estimated *locally* for this node is quite different from $\hat{\beta}_1$: demonstrating a good adaptation to the spatial characteristics of this region, independent of the rest of the image. Node 2, corresponding to the outer region, and nodes 6 and 7, all share similar characteristics: negative log-gains, which prevent from further splits, and very small edge penalties which indicate the absence of significant spatial structures in those regions.

REFERENCES

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[2] J.-K. Fwu and P. M. Djuric, "Unsupervised vector image segmentation by a tree-structure ICM algorithm," *IEEE Trans. Med. Imag.*, vol. 15, pp. 871–880, Dec. 1996.

[3] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees.* Belmont, CA: Wadsworth, 1984.

[4] G. Gelli and G. Poggi, "Compression of multispectral images by spectral classification and transform coding," *IEEE Trans. Image Processing*, vol. 8, pp. 476–489, Apr. 1999.