

Semi-iterative Inferences with Hierarchical Energy-based Models for Image Analysis

Annabelle Chardin and Patrick Pérez

IRISA/INRIA

Campus de Beaulieu, F-35042 Rennes cedex, France.

tel: (+33) 2.99.84.72.73

fax: (+33) 2.99.84.71.71

{achardin,perez}@irisa.fr

Abstract. This paper deals with hierarchical Markov Random Field models. We propose to introduce new hierarchical models based on a hybrid structure which combines a spatial grid of a reduced size at the coarsest level with sub-trees appended below it, down to the finest level. These models circumvent the algorithmic drawbacks of grid-based models (computational load and/or great dependance on the initialization) and the modeling drawbacks of tree-based approaches (cumbersome and somehow artificial structure). The hybrid structure leads to algorithms that mix a non-iterative inference on sub-trees with an iterative deterministic inference at the top of the structure. Experiments on synthetic images demonstrate the gains provided in terms of both computational efficiency and quality of results. Then experiments on real satellite spot images illustrate the ability of hybrid models to perform efficiently the multispectral image analysis.

1 Background: Hierarchical Energy-based Models

Many inverse problems from image analysis can be managed by designing an *energy* function $U(x, y)$ which captures the interaction between a large number of unknown variables $x = (x_i)_i$ to be estimated, and the observed variables –the measurements or data–, $y = (y_j)_j$. The manipulation of this function is made tractable by its usual decomposition as a sum of *local* terms involving just a few variables at a time. This kind of problem is encountered in Markov random field-based approaches as well as in partial differential equation (PDE)-based approaches, where, in the first stage, the energy depends on a continuous function x . Within the framework of Markov random fields, x and y are random vectors and we have the following relation between the joint distribution and the energy function: $P(x, y) \propto \exp\{-U(x, y)\}$. As far as the inference of x is concerned, the Bayesian estimation theory provides two standard estimators: the Maximum A Posteriori (MAP) estimator which corresponds to the global minimizer of the energy function ($\hat{x} = \arg \max_x P(x|y) = \arg \min_x U(x, y)$) and the Modes of Posterior Marginals (MPM) estimator ($\forall i, \hat{x}_i = \arg \max_{x_i} P(x_i|y)$)

which requires the computation of marginals by summation over huge sets of configurations.

It turns out that for most energy-based models suitable for image analysis problems, one has to devise deterministic or stochastic iterative algorithms exploiting the locality of the model in order to conduct the MAP or MPM inference. While permitting tractable single-step computations, the locality results in a very slow propagation of information. As a consequence, these iterative procedures may converge very slowly. This motivates the search either for improved algorithms of generic use, or for specific models allowing non-iterative or more efficient inference.

So far, the more fruitful approaches in both cases have relied on some notion of *hierarchy*. Hierarchical models or algorithms allow the information to be integrated in a progressive and efficient way (especially in the case of multiresolution data, when images come into a hierarchy of scales) providing gains in terms of both computational efficiency and quality of results.

Algorithm-based hierarchical approaches are usually related to well-known *multigrid* resolution techniques from Numerical Analysis, where an increasing sequence of nested spaces is explored in a number of possible ways. The particular case of *coarse-to-fine* exploration has been successfully extended to discrete image models [3, 8]. Within this framework, reduced versions of an original (spatial) model can be deduced in a consistent way (the form of the energy and associated parameters are deduced at once). The “stack” of models thus obtained can then be used for inference purposes, the estimate *iteratively* obtained at a given level being used as an initialization for the processing at the next level.

On the other hand, model-based hierarchical approaches aim at defining a new global hierarchical model which has nothing to do with any original (spatial) model. It has to be manipulated as a whole, but according to procedures of reduced complexity. These models usually lie on the nodes of a quad-tree whose leaves fit the pixels of (maximum resolution) images [4, 9, 10]. In this case, the peculiar dependency structure, like in case of Markov chains, allows *non-iterative* inference procedures made of two sweeps: a bottom-up sweep propagating all information to the root, and a top-down one which in turn allows optimal estimate to be obtained at each node given *all the data*.

One of the drawbacks of these tree-based approaches lies in the structural constraints they impose: first of all they might appear artificial for certain types of problems or data; in any case the relevance of the inferred variables at coarsest levels is not obvious (especially at the root). Second, the complete tree-structure is cumbersome in case of large images.

To circumvent this, here we propose a hierarchical model based on a “hybrid” structure which combines a spatial grid of reduced size at a coarser level with “sub-trees” appended below it, down to the finest level (see Fig.1). This paper investigates the use of the MAP estimator and of the MPM estimator, known as more reliable than MAP, on the hybrid structure. It should be noticed that

MPM estimator relies on the computation of posterior marginals, which are a key ingredient for EM-type parameter estimation techniques.

The section 2 describes the hybrid structure and its associated energy function. In the section 3, we explain the procedures to achieve the MAP and MPM estimates by mixing a non-iterative inference on sub-trees with iterative deterministic inference of reduced cost at the top of the structure. The section 4 illustrates these procedures for image classification problems with synthetic and real images.

2 Hierarchical Model and energy function

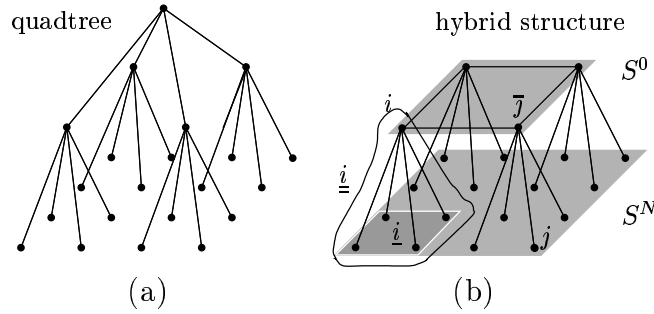


Fig. 1. Two hierarchical structures: (a) quadtree with three levels; (b) truncated tree with two levels.

The hierarchical model we use is based on a hybrid structure [6]. One example is shown in Fig.1(b) for a single level below the coarsest grid. To describe this graph, we shall introduce some notations.

First, we define the coarsest level S^0 as a rectangular grid with a 1st-order neighborhood. Then each site of S^0 initiates a quadtree, so that the grid S^n ($0 < n \leq N$) made up by the nodes at the level n is $2^n \times 2^n$ times larger than S^0 . Now each site i of S^n has four natural correspondents in S^{n+1} (provided that i does not belong to the finest level S^N), its children, forming site set \underline{i} , and one natural correspondent in S^{n-1} (provided that i does not belong to the coarsest level S^0), its parent, denoted as \bar{i} . Finally, the site set forming the tree rooted at i is denoted $\underline{\underline{i}}$ (Fig.1). Vectors x and y are now indexed by the nodes of $S \triangleq \bigcup_{n=0}^N S^n$.

Given this graphical structure consider an energy function of the following form:

$$U(x, y) \triangleq \sum_{\langle i, j \rangle \in S^0} v_{ij}(x_i, x_j) + \sum_{i \notin S^0} w_i(x_i, x_{\bar{i}}) + \sum_{i \in S} l_i(x_i, y_i), \quad (1)$$

where $\langle i, j \rangle$ designates pairs of neighbors in S^0 , v_{ij} and w_i are local functions capturing respectively the spatial prior and the hierarchical prior (they will usually encourage identity between neighbors and between parents and children, resp.), and l_i expresses the point-wise relation between the observed variable y_i ¹ and the unknown one x_i . The MAP estimator for this model amounts to minimizing $U(x, y)$ in (1) with respect to x for a given y .

From a probabilistic point of view, the associated posterior distribution of (x, y) is:

$$P(x|y) \propto \prod_{\langle i, j \rangle \in S^0} \underbrace{g_{ij}(x_i, x_j)}_{\triangleq \exp\{-v_{ij}(x_i, x_j)\}} \times \prod_{i \notin S^0} \underbrace{f_i(x_i, x_{\bar{i}})}_{\triangleq \exp\{-w_i(x_i, x_{\bar{i}})\}} \times \prod_{i \in S} \underbrace{h_i(x_i, y_i)}_{\triangleq \exp\{-l_i(x_i, y_i)\}} . \quad (2)$$

The MPM estimator requires to deduce from this global posterior distribution, each of the local posterior distributions $P(x_i|y)$, for $i \in S$.

3 Semi-iterative inferences

3.1 MAP computation

The global minimization of $U(x, y)$ w.r.t. x can be written in a way that distinguishes variables on S^0 (interacting in a non-causal fashion) from the others (interacting in a tree-based causal fashion):

$$\min_x U(x, y) = \min_{x_{S^0}} \left\{ \sum_{\langle i, j \rangle \in S^0} v_{ij}(x_i, x_j) + \sum_{i \in S^0} l_i(x_i, y_i) + \sum_{i \in S^0} \min_{x_{\underline{i} \setminus \{i\}}} \sum_{j \in \underline{i} \setminus \{i\}} [w_j(x_j, x_{\bar{j}}) + l_j(x_j, y_j)] \right\}. \quad (3)$$

Each tree-based minimization w.r.t. $x_{\underline{i} \setminus \{i\}}$, for $i \in S^0$ in the left-hand-side can be conducted exactly, with fixed complexity per node, using an extension of the chain-based Viterbi algorithm [7]. The first (upward) sweep of this algorithm computes recursively the optimal value $x_i^*(x_{\bar{i}})$ at a node i given the value of its parent, and the associated value $V_i(x_{\bar{i}})$ for the energy term that concerns \underline{i} :

$$V_i(x_{\bar{i}}) \triangleq \min_{x_{\underline{i}}} \sum_{j \in \underline{i}} [w_j(x_j, x_{\bar{j}}) + l_j(x_j, y_j)] \quad (4)$$

$$= \min_{x_i} [w_i(x_i, x_{\bar{i}}) + l_i(x_i, y_i) + \sum_{j \in \underline{i}} V_j(x_i)] \quad (5)$$

¹ In the following we consider that measurements are available at each node $i \in S$, to take into account the case of multiresolution images. Derivations are readily adapted to the cases where data are only available on a subset of S . In practice data are, for instance, often defined at the finest level S^N only. In this paper our experimentations were led on monoresolution images which were either monospectral (synthetic cases) or multispectral (real satellite data).

$$\text{and } x_i^*(x_{\bar{i}}) \triangleq \arg \min_{x_i} [w_i(x_i, x_{\bar{i}}) + l_i(x_i, y_i) + \sum_{j \in \underline{i}} V_j(x_i)]. \quad (6)$$

Once this recursion is completed it remains to perform in (3) the minimization w.r.t. x_{S^0} , which amounts to

$$\min_{x_{S^0}} \sum_{\langle i, j \rangle \in S^0} v_{ij}(x_i, x_j) + \sum_{i \in S^0} [l_i(x_i, y_i) + \sum_{j \in \underline{i}} V_j(x_i)].$$

Apart if S^0 is very small, an iterative ICM-type procedure [2] must be introduced here due to the non-causal term $\sum_{\langle i, j \rangle \in S^0} v_{ij}(x_i, x_j)$. This provides estimates \hat{x}_i , for $i \in S^0$, from which all the other estimates are recursively recovered thanks to functions x_j^* : $\hat{x}_j = x_j^*(\hat{x}_{\bar{j}})$. The whole procedure goes as follows:

Semi-iterative energy minimization	
▲ upward sweep	
Leaves ($i \in S^N$)	
	$\begin{cases} V_i(x_{\bar{i}}) \triangleq \min_{x_i} [w_i(x_i, x_{\bar{i}}) + l_i(y_i, x_i)] \\ x_i^*(x_{\bar{i}}) \triangleq \arg \min_{x_i} [w_i(x_i, x_{\bar{i}}) + l_i(y_i, x_i)] \end{cases}$
Recursion (for $n = N - 1 \dots 1, i \in S^n$)	
	$\begin{cases} V_i(x_{\bar{i}}) \triangleq \min_{x_i} [w_i(x_i, x_{\bar{i}}) + l_i(x_i, y_i) + \sum_{j \in \underline{i}} V_j(x_i)] \\ x_i^*(x_{\bar{i}}) \triangleq \arg \min_{x_i} [w_i(x_i, x_{\bar{i}}) + l_i(x_i, y_i) + \sum_{j \in \underline{i}} V_j(x_i)] \end{cases}$
◀ coarse ICM: update a few times all sites of S^0 in turn, for energy	
	$\begin{aligned} & \sum_{\langle i, j \rangle \in S^0} v_{ij}(x_i, x_j) + \sum_{i \in S^0} [l_i(x_i, y_i) + \sum_{j \in \underline{i}} V_j(x_i)] \\ & \Rightarrow \hat{x}_i, \forall i \in S^0 \end{aligned}$
▼ downward sweep	
	$\hat{x}_i = x_i^*(\hat{x}_{\bar{i}})$

If an exact minimization can be obtained at the coarsest level (which is especially the case for the complete tree where S^0 reduces to a single site), the final estimate \hat{x} is exactly the global minimizer of U . Note that the functions $V_i(x_{\bar{i}})$, which appear in the upward sweep, progressively collect dependencies with respect to the data, even though this is not made explicit by abuse of notation: $V_i(x_{\bar{i}})$ (as well as $x^*(x_{\bar{i}})$) actually depends on y_i . This means that ICM at the level 0 and downward sweep provide inferences based on all data.

When y is only attached to S^N , this procedure can be compared to the multigrid method in [8], where the non-hierarchical energy

$$U^N(x^N, y) = \sum_{\langle i, j \rangle \in S^N} v_{ij}^N(x_i, x_j) + \sum_{i \in S^N} l_i(x_i, y_i) \quad (7)$$

is minimized within the set of configurations which are piece-wise constant over $2^{N-n} \times 2^{N-n}$ blocks, for $n = 0 \dots N$; equivalently

$$U^n(x^n, y) = \sum_{\langle i, j \rangle \in S^n} 2^{N-n} v_{ij}^N(x_i^n, x_j^n) + \sum_{j \in \underline{i} \cap S^N} l_j(x_i^n, y_j) \quad (8)$$

is minimized, where x^n is defined on S^n . Let us note that U^0 , thus defined for $n = 0$ in the multigrid approach, corresponds to the energy which is manipulated at the coarsest level of the semi-iterative approach for $w_i(x_i, x_\tau) \triangleq \beta[1 - \delta(x_i, x_\tau)]$ with $\beta \rightarrow +\infty$ (in this case the optimal configuration x is constant over each tree \underline{i} , $i \in S^0$, and $\sum_{j' \in \underline{i}} V_{j'}(x_i) = \sum_{j \in \underline{i} \cap S^n} l_j(x_i^n, y_j)$, $\forall i \in S^0$). Moreover, if data are only available at the finest level, the initialization of the multigrid coarse ICM is given by the minimization of $\sum_{j \in \underline{i} \cap S^n} l_j(x_i^n, y_j)$ for each site $i \in S^0$ while the initialization of the hybrid coarse ICM is given by the minimization of $\sum_{j \in \underline{i}} V_j(x_i)$ for each site $i \in S^0$. To confirm this statement, we will just have to compare the coarsest initializations and estimates provided by the two methods, taking the same number of levels, the same functions v_{ij} and l_i , and $w_i(x_i, x_\tau) \triangleq \beta[1 - \delta(x_i, x_\tau)]$ with β very large for the hybrid energy.

3.2 MPM computation

In the case of a complete tree where S^0 reduces to a single site, the exact MPM estimates can be computed on each node [9] using an extension of Baum-Welch algorithm on a chain [1]. In the general case, the downward recursion is now based on the following relation:

$$\forall i \notin S^0, P(x_i|y) = \sum_{x_\tau} P(x_i|x_\tau, y)P(x_\tau|y),$$

where

$$P(x_i|x_\tau, y) = P(x_i|x_\tau, y_{\underline{i}})$$

due to separation property. The use of this recursion requires that a previous upward sweep provides $P(x_i|x_\tau, y_{\underline{i}})$ for $i \notin S^0$ and $P(x_i|y)$ for $i \in S^0$. This is achieved by successively summing out x_i 's for all $i \notin S^0$. The recursion is based on:

$$\begin{aligned} P(x_i|x_\tau, y_{\underline{i}}) &\propto f_i(x_i, x_\tau)h_i(x_i, y_i) \sum_{x_{\underline{i} \setminus \{i\}}} \prod_{j \in \underline{i} \setminus \{i\}} f_j(x_j, x_\tau)h_j(x_j, y_j) \\ &\propto f_i(x_i, x_\tau)h_i(x_i, y_i) \prod_{j \in \underline{i}} \sum_{x_j} \prod_{k \in \underline{j}} f_k(x_k, x_\tau)h_k(x_k, y_k), \end{aligned} \quad (9)$$

with

$$\begin{aligned} \mathbb{F}_j(x_i) &\triangleq \sum_{x_{\underline{j}}} \prod_{k \in \underline{j}} f_k(x_k, x_\tau)h_k(x_k, y_k) \\ &= \sum_{x_j} f_j(x_j, x_\tau)h_j(x_j, y_j) \prod_{k \in \underline{j}} \mathbb{F}_k(x_j). \end{aligned} \quad (10)$$

This eventually provides the probability

$$P(x_{S^0}|y) = \sum_{x_{S \setminus S^0}} P(x|y).$$

Because of the non-causal structure on S^0 , $P(x_i|y)$ for $i \in S^0$ have to be approximated with the help of a Gibbs sampling of the distribution $P(x_{S^0}|y)$. This procedure provides approximated local posterior marginals in a semi-iterative way. If a MPM estimator is employed, an approximation of it is obtained as a by-product.

Then the whole procedure goes as follows:

<p>Semi-iterative local posterior marginal computation and MPM estimation</p> <p>▲ upward sweep</p> <p>Leaves ($i \in S^N$)</p> $\mathbb{F}_i(x_{\bar{i}}) = \sum_{x_i} f_i(x_i, x_{\bar{i}}) h_i(x_i, y_i)$ <p>Recursion (for $n = N - 1 \dots 1, i \in S^n$)</p> $\mathbb{F}_i(x_{\bar{i}}) = \sum_{x_i} f_i(x_i, x_{\bar{i}}) h_i(x_i, y_i) \prod_{j \in \underline{i}} \mathbb{F}_j(x_i)$ <p>◀ coarse posterior marginal computation:</p> <hr/> <p>draw samples $x_{S^0}(1), \dots, x_{S^0}(m)$ from:</p> $P(x_{S^0} y) \propto \prod_{\langle i, j \rangle \in S^0} g_{ij}(x_i, x_j) \times \prod_{i \in S^0} h_i(x_i, y_i) \times \prod_{i \in S^0} \prod_{j \in \underline{i}} \mathbb{F}_j(x_i)$ <p>approximation of $P(x_i y) \approx \frac{1}{m-k} \sum_{j=k+1}^m \delta[x_i(j), x_i]$</p> <hr/> <p>▼ downward sweep</p> <p>Recursion (for $n = 1 \dots N, i \in S^n$)</p> $P(x_i y) = \sum_{x_{\bar{i}}} P(x_{\bar{i}} y) \frac{f_i(x_i, x_{\bar{i}}) h_i(x_i, y_i)}{\mathbb{F}_i(x_{\bar{i}})} \prod_{j \in \underline{i}} \mathbb{F}_j(x_i)$ <p>MPM at leaves</p> $\hat{x}_i \triangleq \arg \min_{x_i} P(x_i y)$

Here, the same abuse of notation is made as in the MAP procedure: in fact the functions $\mathbb{F}_i(x_{\bar{i}})$ depend on $y_{\underline{i}}$, so that the progressive summations out x_i 's for all $i \notin S^0$ are made with respect to data associated to \underline{i} and the downward sweep provides local posterior marginal based on all data.

4 Supervised Classification Comparisons

To demonstrate the practicability and the relevance of the approach for discrete low-level image analysis, we first report comparative experiments for supervised classification. To this end, we considered a Potts-type prior with potentials

$$v_{ij}(x_i, x_j) \triangleq 2^N \alpha [1 - \delta(x_i, x_j)] , \quad (11)$$

$$w_i(x_i, x_{\bar{i}}) \triangleq \beta [1 - \delta(x_i, x_{\bar{i}})] , \quad (12)$$

along with Gaussian likelihoods

$$l_i(x_i = k, y_i) \triangleq \begin{cases} \frac{(y_i - \mu_k)^2}{2\sigma_k^2} + \log(\sigma_k) & \text{if } i \in S^N, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

Comparative experiments were led for $N \in \{0, 2, 3, 8\}$ for the synthetic scene and for $N \in \{0, 3, 4, 8\}$ for the real spot scene. For $N = 0$ the inferences are based on standard non-hierarchical models using iterative ICM for the approximation of the MAP or an iterative Gibbs sampler for the approximation of the MPM. While $N = 8$, when the size of S^N is $2^8 \times 2^8$, corresponds to the complete tree ($|S^0| = 1$) allowing an exact non-iterative computations. $N = 2, 3, 4$ correspond to three-, four- and five-level hybrid structures on which semi-iterative MAP and MPM inference are performed.

4.1 Synthetic images

First, the experiments were carried out on a 256×256 synthetic image involving 5 classes (Fig. 2). We applied an additive Gaussian white noise with a different standard deviation for each class, thus the gray level means and variances $(\mu_k, \sigma_k^2)_{k=1}^5$ are known.

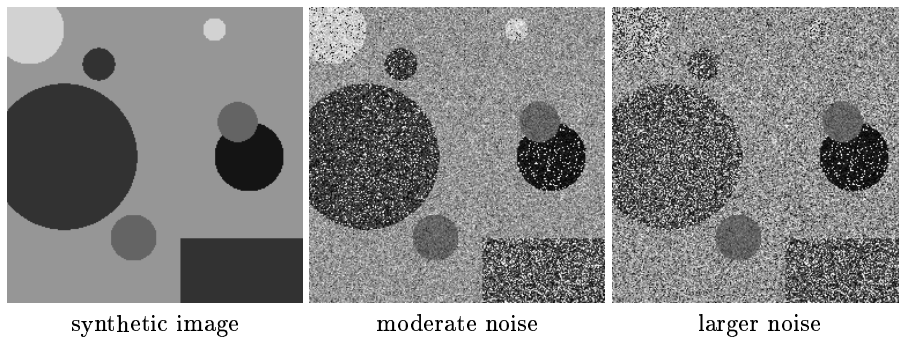


Fig. 2. Synthetic data.

First we stress on the MAP procedure to support experimentally the statement given in the section 3, according to which the semi-iterative MAP inference with data at the finest level only tends to constrained non-hierarchical minimization when $\beta \rightarrow +\infty$. It is easy to see (Fig.3) that both the initializations and the results of the coarse ICM for the multigrid method and the semi-iterative one are the same as soon as β is large enough (greater than 80). Accordingly, the multigrid approach and the semi-iterative procedure with $\beta \rightarrow +\infty$ really behave in the same way.

Secondly, we worked on a moderately corrupted and a fairly corrupted images, and then paid attention to how our algorithms behave for a noisier image (see Fig.2). The obtained classification results are shown in Fig.4 and in Fig.5 with their respective percentages of misclassification and cpu times in seconds (on a 170 MHz Ultra Sparc I workstation) in tables 1 and 2.

As can be seen from the figures 4 and 5, the hierarchical models provide good results. As the noise level is low, the resulting classifications of all the methods



Fig. 3. comparative results at the coarsest level between the multigrid method (a1-b1) and the semi-iterative method (a2-b2) with $N = 2$ and $\beta \rightarrow +\infty$.

are of an almost equivalent quality. Whereas the degradation of the synthetic image increases, the hierarchical models and the multigrid approach seem more robust to noise than the plain non-hierarchical iterative procedures. Moreover, it can be noticed that the non- or semi-iterative algorithms take much fewer cpu time: for the MAP, they take three times less cpu time than the ICM, and for the MPM, more than forty times less cpu time than the Gibbs sampler. For the estimation of the posterior marginals in the sampling procedure, we fit the number of retained samples ($m - k$, see the description of the algorithm) to the size of the concerned grid (200 for the complete grid and 20 for the coarse grid of the hybrid structure).

To go further in the comparisons, we can focus on the results for the complete tree ($N = 8$) and for the two examples of the hybrid structure for each estimator. First, if we just look at the tables, we can say that the semi-iterative estimation provides slightly better classification than a non-iterative one for a comparable cpu time. Moreover, the MPM classifications are slightly better than the MAP ones, especially for $N = 2$, but they enquire more calculations because the MAP downward recursion is simpler (one has only to read look-up tables built in the upward recursion, whereas there are some calculations during the MPM downward step). Anyway, whatever estimator is used, the use of a coarse iterative procedure on top of the hybrid structures does not seem to imply an extra computational load, while improving the results.

As far as the visual aspect is concerned, the semi-iterative classifications still reveal a blocky aspect (as observed in all tree-based approaches). But these artifacts are less and less pronounced as the number of levels in the hybrid structure decreases.

On top of the fact that the MPM is doing better than the MAP, the hierarchical MPM provide a measure of relative confidence associated to the estimated value at each site, through the entropy

$$c_i \triangleq \sum_{x_i} P(x_i|y) \log P(x_i|y).$$

These measures (Fig.6) allow us to better appreciate the quality of the obtained estimates and to use them in consequence.

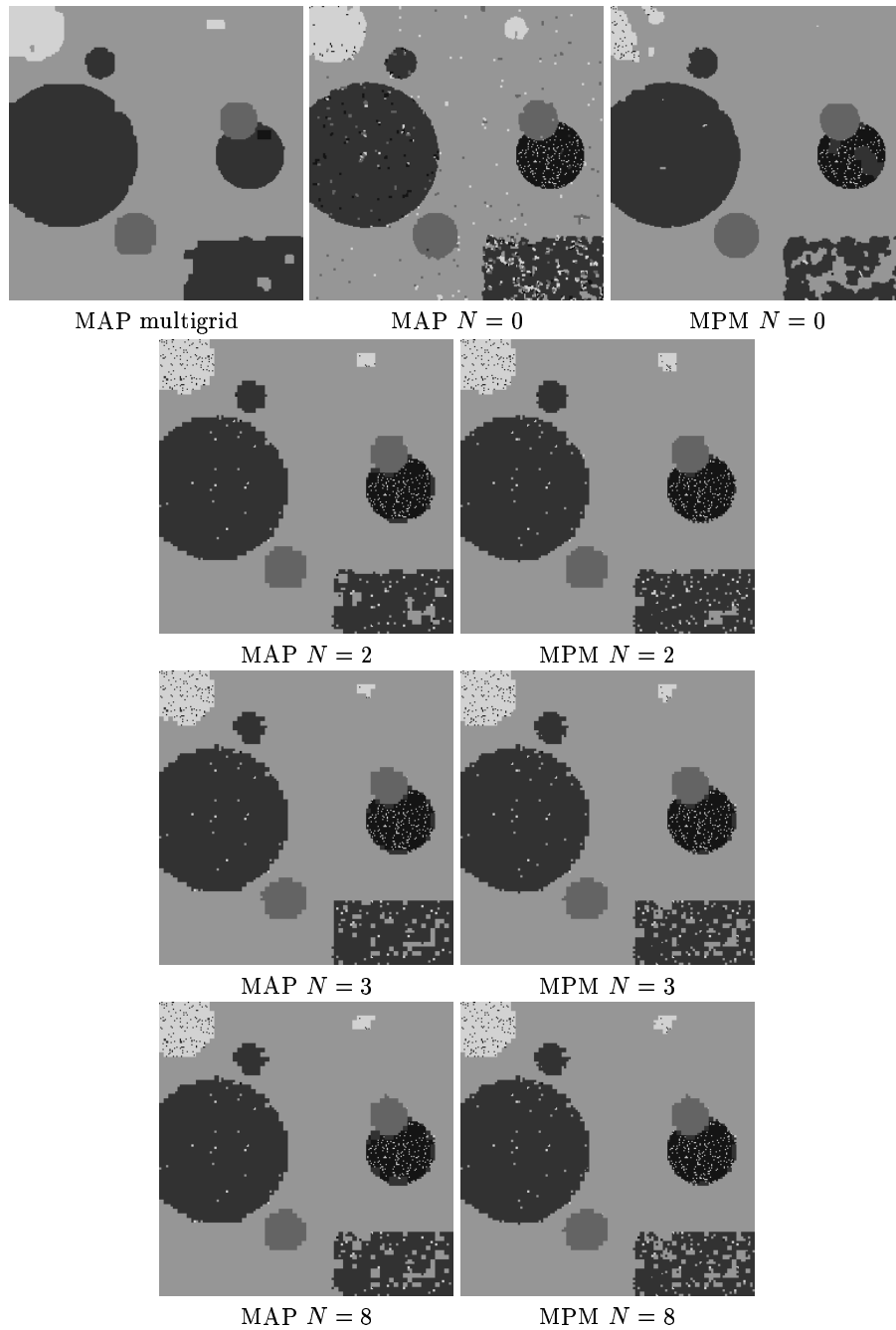


Fig. 4. Comparative results for the classification problem with a moderately corrupted image.

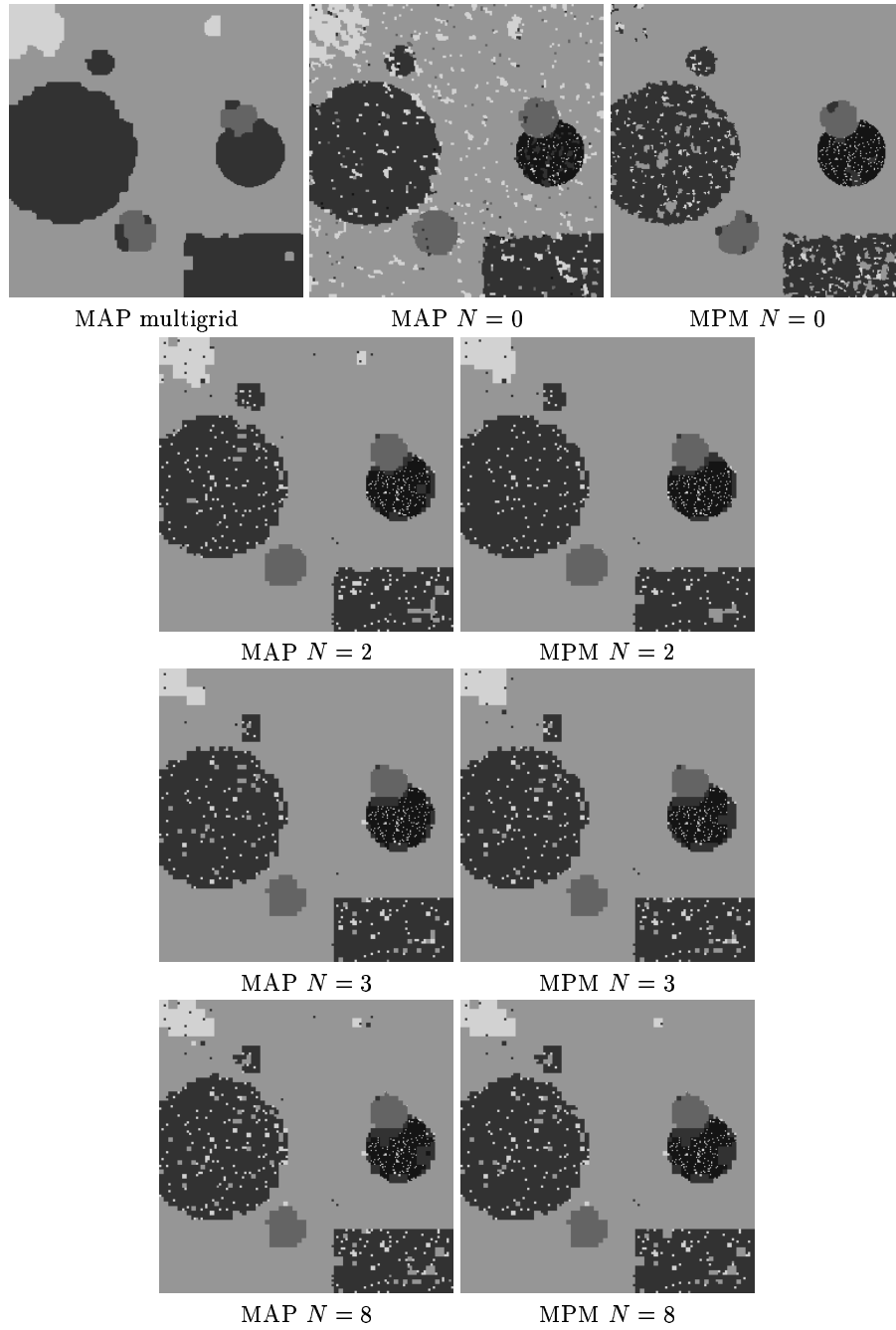


Fig. 5. Comparative results for the classification problem with a more corrupted image.

Model	MAP		MPM	
	misclassification	cpu time	misclassification	cpu time
iterative $N = 0$	5.3%	11s	7%	7min
semi-iterative $N = 2$	4.15%	3.6s	3.6%	10.5s
semi-iterative $N = 3$	4.2%	3.6s	4.5%	9.3s
non-iterative $N = 8$	4.8%	3.6s	4.7%	9s
multigrid	5.9%	11.6s		

Table 1. Performances of the different algorithms with the synthetic image in fig.4

Model	MAP		MPM	
	misclassification	cpu time	misclassification	cpu time
iterative $N = 0$	9.65%	11s	10%	7min
semi-iterative $N = 2$	6.6%	3.6s	6.2%	10.5s
semi-iterative $N = 3$	7.9%	3.6s	7.9%	10s
non-iterative $N = 8$	8%	3.6s	8%	9s
multigrid	7.3%	11.6s		

Table 2. Performances of the different algorithms with the synthetic image in fig.5

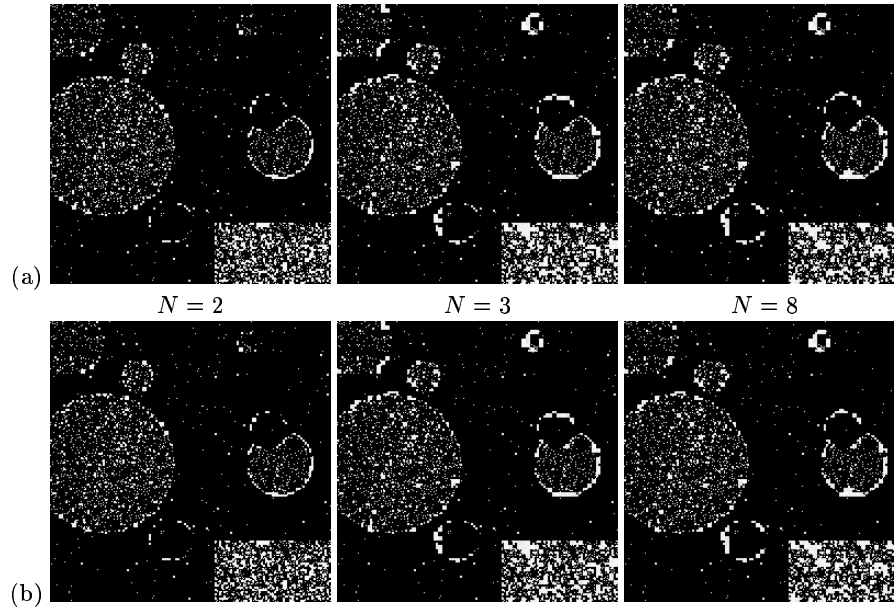


Fig. 6. Confidence maps associated to classification (a) of the moderately noisy image and (b) of the noisier image.

4.2 Real satellite images

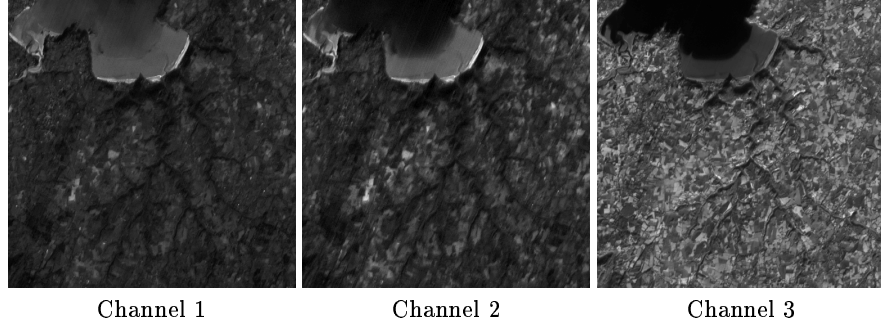


Fig. 7. 512×512 Spot images (courtesy of Costel, University of Rennes 2 and GSTB).

The previous algorithms were applied to SPOT satellite images provided by the Costel laboratory (University of Rennes 2) in the context of remote sensing researches. The scene (Fig.7) is composed of three 512×512 images with different wavelengths in the visible spectrum and represents the Bay of Lannion in France in December 1996. The goal of this study was to determine the land cover of this area. So as to reach this aim, the geographers of the Costel laboratory built a list of eight classification categories:

- Sea and water
- Sand and bare soil
- Urban area
- Forests and heath
- Temporary meadows
- Permanent meadows
- Colza
- Vegetables.

Thanks to both tests on the lands and photointerpretations, they were also able to provide samples of these eight categories on the three SPOT images of the scene. Some of them were used to learn the gray levels means and the variances of each category for each image, so that we could perform supervised classifications, and the left samples were kept to assess the accuracy of the classifications.

As for the model, we considered the three channels as independant. As a consequence, the form of the relation between the observed variables (here the multispectral scene) and the unknown variables (class label at each pixel) becomes:

$$l_i(x_i = k, y_i^c \ c \in \{1, 2, 3\}) \triangleq \sum_{c=1}^{c=3} \left(\frac{(y_i^c - \mu_{k_c})^2}{2\sigma_{k_c}^2} + \log(\sigma_{k_c}) \right) \quad \text{if } i \in S^N, \quad (14)$$

where $(\mu_{k_c}, \sigma_{k_c})$ are the gray level mean and variance of the class k within channel number $c \in \{1, 2, 3\}$.

The eight algorithms provide quite similar results of a good quality (see Fig.8 and Tab.3). About 92% of the pixels of the samples are well classified. The three hierarchical MAP inferences are achieved in almost twice less cpu time than the iterative ICM. In comparison with the iterative MPM which takes more than one hour to be performed, the three hierarchical MPM inferences need reasonable cpu time. This is encouraging for the EM-type parameter estimation algorithms for which the form of one recursion step is close to the MPM algorithm form.

Model	MAP		MPM	
	misclassification	cpu time	misclassification	cpu time
iterative $N = 0$	8%	150s	8%	1h10min
semi-iterative $N = 3$	7.5%	85s	7.5%	165s
semi-iterative $N = 4$	7.5%	85s	7.5%	165s
non-iterative $N = 8$	8%	65s	7.8%	160s

Table 3. Performances of the different algorithms with the real multispectral images in fig.7.

5 Conclusion and extensions

In this paper, we presented a hybrid hierarchical structure which is an interesting compromise between standard spatial models and hierarchical models based on a complete quadtree. We introduce two inference algorithms devoted to this new structure: the first one computes the MAP estimate and the second one computes the local posterior marginals and the MPM estimate.

With this structure we now plan to deal with the critical problem of parameter estimation: using marginal computation introduced here, we should be able to design specific EM-like algorithms on the hybrid structure, as already done with trees [4, 9] and with spatial grids [5]. In the classification problem, this will concern both the data parameters (number of classes, gray level means and variances of each class) whose automatic estimation will allow unsupervised classification, and the parameters of the prior model (α, β) . We also plan to address the issue of automatically estimating the optimal number of levels in the structure.

References

1. L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.*, Vol 41: pp. 164–171, 1970.

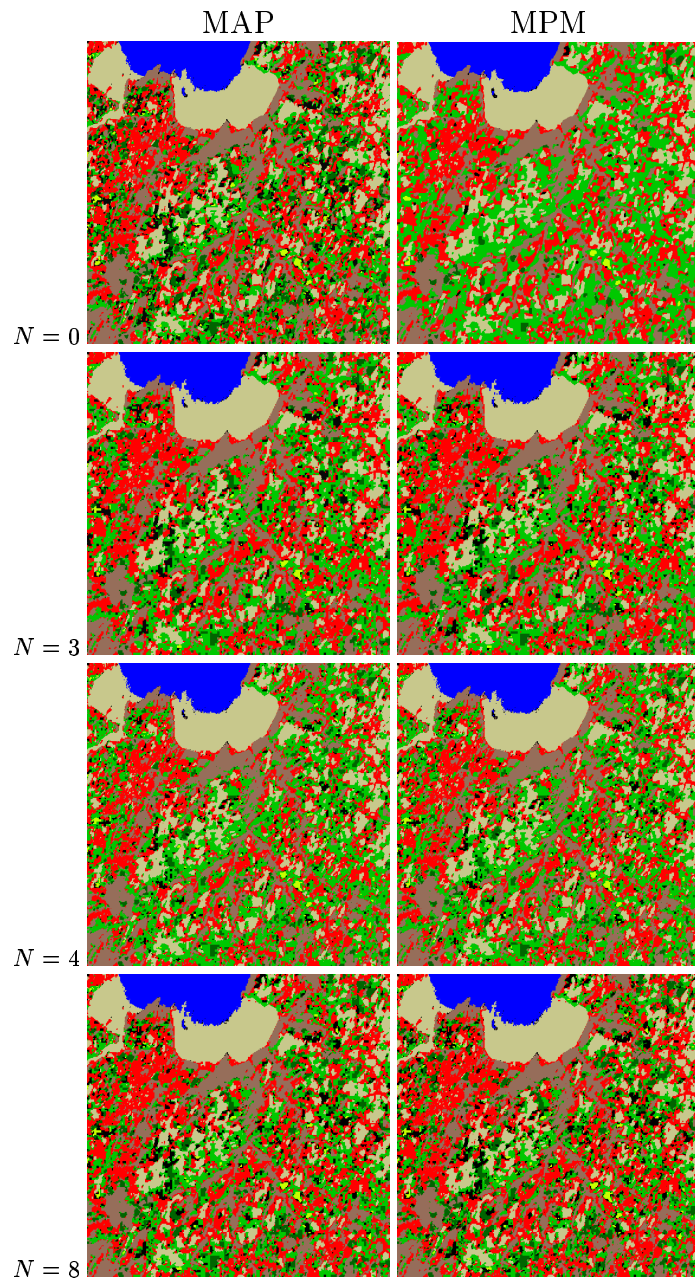


Fig. 8. Comparative results for the classification problem with multispectral spot images (legend: sea and water in blue, sand and bare soil in beige, urban area in red, forests and heath in maroon, temporary meadows in pale-green, permanent meadows in dark-green, colza in yellow and vegetables in black).

2. J. Besag. Spatial interaction and the statistical analysis of lattice systems. *J. Royal Statist. Soc.*, 36, Série B:192–236, 1974.
3. C. Bouman and B. Liu. Multiple resolution segmentation of textured images. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No. 2 : pages 99–113, Février 1991.
4. C. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. on Image Processing*, 3, No. 2 :162–177, March 1994.
5. B. Chalmond. An iterative Gibbsian technique for reconstruction of m-ary images. *Pattern Recognition*, Vol. 22, No 6: pages 747–761, 1989.
6. A. Chardin and P. Pérez. Semi-iterative inference with hierarchical models. In *Int. Conf. on Image Processing*, pages 630–634, Chicago, USA, october 1998.
7. G.D. Forney. The Viterbi algorithm. *Proc. IEEE*, Vol. 13: pages 268–278, March 1973.
8. F. Heitz, P. Pérez, and P. Bouthemy. Multiscale minimization of global energy functions in some visual recovery problems. *CVGIP : Image Understanding*, Vol. 59, No 1: pages 125–134, Jan. 1994.
9. J-M Laferté, P. Pérez, and F. Heitz. Discrete markov image modeling and inference on the quad-tree. *IEEE Trans. Image Processing*, Accepted for publication, 1999.
10. M. Luetttgen, W. Karl, and A. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. *IEEE Trans. Image Processing*, 3(1):41–64, 1994.