IEEE Int. Conf. Image Proc., vol. 1, pp. 630–634, October 1998, Chicago

Semi-iterative inference with hierarchical models

A. Chardin P. Pérez
IRISA/INRIA-Rennes
Campus de Beaulieu, F-35042 Rennes cedex, France.
e-mail : {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 a synthetic image demonstrate the gains provided in terms of both computational efficiency and the quality of results. Then experiments on real aerial images illustrate the ability of hybrid models to perform the multiresolution and multispectral image fusion.

1 Background: hierarchical energy-based models

Many inverse problems from image analysis can be managed by designing and minimizing 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 (in this case, the minimizer corresponds to the maximum *a posteriori* estimate) as well as in partial differential equation (PDE)-based approaches, where, in the first stage, the energy depends on a continuous function x.

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 minimization. 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 minimization algorithms of generic use, or for specific models allowing non-iterative or more efficient inference.

So far, the more fruitful approaches in both these cases have relied on some notion of *hierarchy* (see [5] for a recent review). 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 the 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 [2, 6]. 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 [3, 7, 8, 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). As a consequence, we shall see that the energy minimization can be achieved by mixing a non-iterative inference on sub-trees with iterative deterministic inference of reduced cost at the top of the structure.

2 Hierarchical model and semi-iterative inference

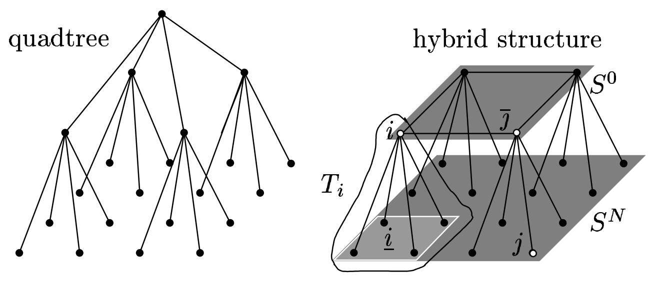


Figure 1: Two hierarchical structures

Let S^0 be a rectangular grid with a 1st-order neighborhood, and S^n $(0 < n \leq N)$ the grid which is $2^n \times 2^n$ times larger. Each site i of S^n has four natural correspondents in S^{n+1} (provided that $n < N$), its children, forming site set $\underline{i}$, and one natural correspondent in S^{n-1} (provided that $n > 0$), its parent, denoted as $\bar{\imath}$. We shall also denote T_i the subtree rooted at i (Fig.1). Vectors x and y are now indexed by the nodes of $S = \bigcup_{n=0}^{N} S^n$. Given this graphical structure consider an energy function of the following form:

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

where $< i,j >$ 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 and the unknown one x_i.

On each tree T_i, $i \in S^0$, considered independently from the others, the exact global minimization can be computed with fixed complexity per node using an extension of the chain-based Viterbi algorithm [4]. This algorithm will be actually used here on each tree, but an iterative ICM-type procedure [1] must be introduced between the two passes in order to take into account the non-causal term involving all the roots of these trees (i.e., sites of the coarse grid S^0). The whole procedure goes as follows:

Semi-iterative energy minimization

▲ upward sweep

Leaves ($i \in S^N$)

$$V_i(x_{\bar{\imath}}) \triangleq \min_{x_i}[l_i(x_i,y_i) + w_i(x_i,x_{\bar{\imath}})]$$
$$x_i^*(x_{\bar{\imath}}) \triangleq \arg\min_{x_i}[l_i(x_i,y_i) + w_i(x_i,x_{\bar{\imath}})]$$

Recursion (for $n = N-1 \dots 1$, $i \in S^n$)

$$V_i(x_{\bar{\imath}}) \triangleq \min_{x_i}[l_i(x_i,y_i) + w_i(x_i,x_{\bar{\imath}}) + \sum_{j\in\underline{i}} V_j(x_i)]$$
$$x_i^*(x_{\bar{\imath}}) \triangleq \arg\min_{x_i}[l_i(x_i,y_i) + w_i(x_i,x_{\bar{\imath}}) + \sum_{j\in\underline{i}} V_j(x_i)]$$

◀▶ coarse ICM:

initialization by

$$x_i = \arg\min_{\lambda}[l_i(\lambda,y_i) + \sum_{j\in\underline{i}} V_j(\lambda)]$$

update all sites of S^0 a few times in turn, for energy

$$\sum_{<i,j>\in S^0} v_{ij}(x_i,x_j) + \sum_{i\in S^0}[l_i(x_i,y_i) + \sum_{j\in\underline{i}} V_j(x_i)]$$

$\Rightarrow \hat{x}_i,\ \forall i \in S^0$

▼ downward sweep (for $n = 1 \dots N$, $i \in S^n$)

$$\hat{x}_i = x_i^*(\hat{x}_{\bar{\imath}})$$

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 $V_i(x_{\bar{\imath}})$, which appear in the upward sweep, progressively collect dependencies with respect to the data, even though this is not made explicit by the misuse of notation: $V_i(x_{\bar{\imath}})$ (as well as $x^*(x_{\bar{\imath}})$) actually depends on y_{T_i}. This means that ICM at the level 0 and downward sweep provide inferences based on all data.

3 Supervised classification comparisons

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

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

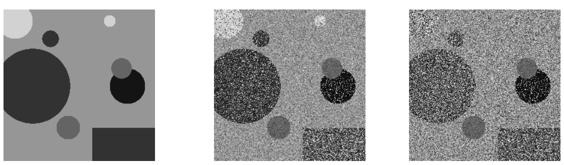
$$w_i(x_i, x_{\bar{i}}) \triangleq \beta[1 - \delta(x_i, x_{\bar{i}})],$$

along with Gaussian likelihoods

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

For $N = 0$ this is a standard non-hierarchical model using iterative ICM, 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 minimization.

3.1 Synthetic images



synthetic image   moderate noise   larger noise

Figure 2: Synthetic data

The experiments were carried out on a 256×256 synthetic image involving 5 classes. We applied an additive Gaussian white noise with a different standard deviation for each class. First, we worked on a fairly corrupted image, and then paid attention to how our algorithms behave for a noisier image (see Fig.2). The gray level means and variances $(\mu_k, \sigma_k^2)_{k=1}^5$ being thus known, 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 Sun Sparc 5).

The results are also compared to the multigrid method [6], where the non-hierarchical energy is minimized within the set of configurations which are piecewise constant over $2^{N-n} \times 2^{N-n}$ blocks, for $n = 0 \dots N$; equivalently

$$U^n(x^n, y) \triangleq \sum_{<i,j> \in S^n} 2^{N-n} \alpha[1 - \delta(x_i^n, x_j^n)] + \sum_{j \in T_i \cap S^N} l_j(x_i^n, y_j)$$

is minimized, where x^n is defined on S^n. Let us note that U^N corresponds to the energy which is manipulated at the coarsest level of the semi-iterative approach for $\beta \to +\infty$ (in this case the optimal configuration x is constant over each tree T_i, $i \in S^0$, and $\sum_{j' \in \underline{i}} V_{j'}(x_i) = \sum_{j \in T_i \cap S^N} l_j(x_i^n, y_j)$, $\forall i \in S^0$). To confirm this statement, we just have to compare the coarsest estimates provided by the two methods, taking the same number of levels (here $N = 2$), the same α and γ, and β very large for the hybrid energy. First, we can examine the initialization (Fig.3(a)) of the coarse ICM for those methods and then the results of the algorithm (Fig.3(b)). Since, in our example, data are only available at the finest level, the initialization of the multigrid coarse ICM is given by the minimization of $\sum_{j \in T_i \cap S^N} l_j(x_i^n, y_j)$ 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 $i \in S^0$. Now, it is easy to see that the images are strictly the same for both the initialization and the final coarse classification. Accordingly, the multigrid approach and the semi-iterative procedure with $\beta \to +\infty$ really behave in the same way.

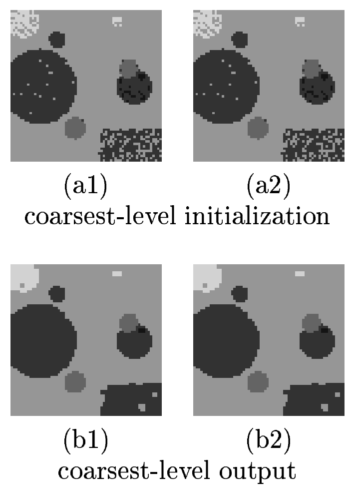
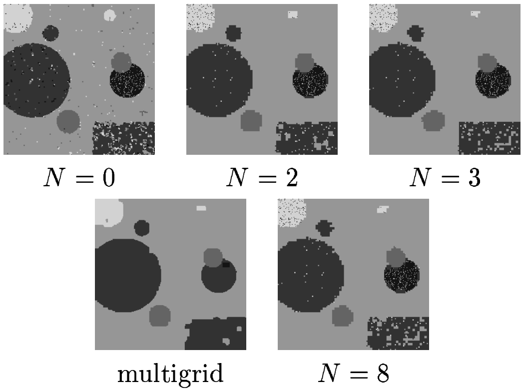


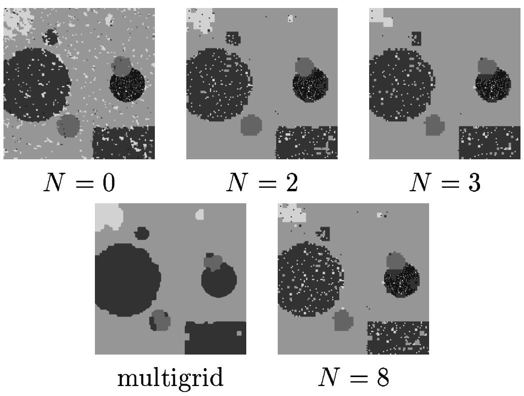
Figure 3: comparative results at the coarsest level between the multigrid method (a1-b1) and the semi-iterative method (a2-b2) with $N = 2$ and $\beta \to +\infty$

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 are of an almost equivalent quality. Whereas the degradation of the synthetic image increases, the three hierarchical models and the multigrid approach seem more robust to noise than the plain iterative ICM which greatly depends on its initialization. Moreover, it can be noticed that the non- or semi-iterative algorithms take much fewer cpu times.



Model	misclassif.	cpu time
iterative $N = 0$	5.3%	20s
semi-iterative $N = 2$	4.15%	8.3s
semi-iterative $N = 3$	4.2%	8.2s
non-iterative $N = 8$	4.95%	8.2s
multigrid	5.9%	35s

Figure 4: comparative results for the classification problem with a moderately corrupted image



Model	misclassif.	cpu time
iterative $N = 0$	9.65%	21s
semi-iterative $N = 2$	6.6%	8.2s
semi-iterative $N = 3$	7.9%	8.2s
non-iterative $N = 8$	8.2%	8.2s
multigrid	7.3%	36s

Figure 5: comparative results for the classification problem with a more corrupted image

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. 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. Thus, the use of a coarse ICM does not seem to imply an extra computational load, while improving the results. On the other hand, in the hybrid restoration of the noisier image with $N = 3$, a disk has totally disappeared. This might be explained by the fact that, when the noise level becomes large, this disk is represented by only one or two sites at the coarsest level and then the coarse ICM smoothes out the structure as if it was part of the noise. That is why the structure with $N = 3$ is less robust than with $N = 2$. A last critical remark: as expected, 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.

3.2 Real aerial images

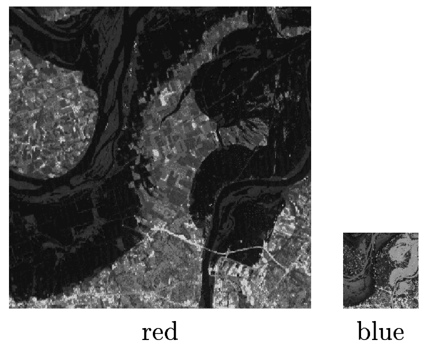


Figure 6: Real aerial images

The previous algorithms were applied to aerial images (in the visible spectrum) of Saint-Louis region during the historic rise in the Mississipi and the Missouri river levels in July 1993. The two available images (Fig.6) are defined with two different resolutions and wavelengths (512×512 red and 128×128 blue). We took into account four classes which are homogeneous in terms of grey levels: one class for the Mississipi, one for the Missouri, one for the urban area (roads, buildings,...) and one last for the left-over land (forests, fields,...).

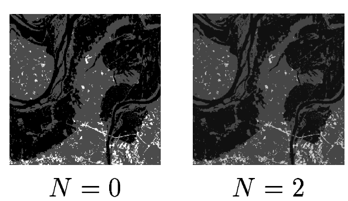


Figure 7: comparative results for the classification problem with the 512×512 red image from figure 6.

If we only consider the red 512×512 image, the two rivers are not distinguished (see Fig.7) whatever method we used. Unlike the tree-based models, the grid-based models do not allow the integration of multiresolution data. Consequently, as concerns the classification based on both images, we can only compare the non-iterative (e.g., $N = 2$ and $N = 3$) and the semi-iterative ($N = 9$, since the size of S^N is $2^9 \times 2^9$) inferences. $N = 3$ and $N = 9$ provide quite similar results with an equivalent cpu time; thus we just present the classifications for the semi-iterative algorithm with $N = 3$ in Fig.8. The quality of the hybrid estimates with $N = 2$ (i.e., with the coarsest level corresponding to the coarse data resolution) appears to be higher: they exhibit less isolated misclassified pixels inside a large homogeneous area. However, the compensation is that the 128×128 coarse ICM slightly increases the computational load (28s on a Sun Sparc 5 compared with 23s for $N = 3$). These results support the idea that we do not need to go up to the one-pixel root of the quadtree and that we should be able to find an optimal number of levels in the hybrid structure with respect to both the quality of results and the computational load.

The figure 8 shows the pyramid-shaped structure of data and classifications. It can be noticed that the estimates are gradually refined throught a progressive integration (and thus fusion) of the multiresolution data. Hierarchical modeling allows us to benefit in a natural and easy way from multiresolution data and provides cascades of estimates which all depend on the whole data.

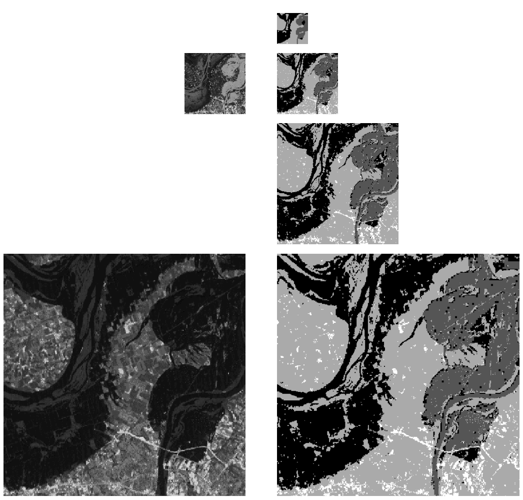


Figure 8: Results for the semi-iterative algorithm with $N = 3$ (right) and two multiresolution multispectral images (left). Cpu times on a Sun Sparc 5 workstation for $N = 3$ et $N = 9$: 23s.

4 Extension

The hybrid structure we have introduced presents an interesting compromise for hierarchical image analysis. With this structure we now plan to deal with the critical problem of parameter estimation, by devising specific EM-like algorithms on it. 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, as already done with trees [3, 7]; and the parameters of the prior model (N, α, β).

Another prospect is the study of a continuous Gaussian counterpart of our hybrid model, as an extension to the tree-based approach of Luettgen *et al.* [8, 9].

References

[1] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", *J. Royal Statist. Soc.*, Vol. B36, pp. 192-236, 1974.

[2] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13(2), pp. 99-113, 1991.

[3] C. Bouman and M. Shapiro, "A multiscale image model for Bayesian image segmentation", *IEEE Trans. Image Processing*, Vol. 3(2), pp. 162-177, 1994.

[4] G. D. Forney, "The Viterbi algorithm", *Proc. IEEE*, Vol. 61(3), pp. 268-278, 1973.

[5] C. Graffigne, F. Heitz, P. Pérez, F. Prêteux, M. Sigelle and J. Zérubia, "Hierarchical and statistical models applied to image analysis: a review", submitted to *IEEE Trans. Inf. Theory*, available at `ftp://gdr-isis.enst.fr/pub/publications/Rapports/GDR/it96.ps`.

[6] F. Heitz, P. Pérez and P. Bouthemy, "Multiscale minimization of global energy functions in some visual recovery problems", *CVGIP : Image Understanding*, Vol. 59(1), pp. 125-134, 1994.

[7] J.-M. Laferté, F. Heitz, P. Pérez and E. Fabre, "Hierarchical statistical models for the fusion of multiresolution image data", *Proc. Int. Conf. Computer Vision*,Cambridge, June 1995.

[8] M. Luettgen, W. Karl and A. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow", *IEEE Trans. Image Processing*, Vol. 3(1), pp. 41-64, 1994.

[9] M. Luettgen and A. Willsky, "Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination", *IEEE Trans. Image Processing*, Vol. 4(2), pp. 194-207, 1995.

[10] P. Pérez and J.-M. Laferté, "Noniterative manipulation of discrete energy-based models for image analysis" *Proc. Int. Work. Energy Minimization Methods in Comp. Vision and Pattern Rec.*, May 1997, extended version at `ftp://ftp.irisa.fr/techreports/1996/PI-1058.ps.gz`.