Hierarchical Markov random field models applied to image analysis: a review

Christine Graffigne [1], Fabrice Heitz [2], Patrick Pérez [3],

Françoise Prêteux [4], Marc Sigelle [5], and Josiane Zerubia [6]

[1] *Université René Descartes*, 45 rue des Saints Pères, 75270 Paris Cédex 06, France
[2] *ENSPS/LSIIT*, Bd Sébastien Brant, 67400 Illkirch, France
[3] *IRISA/INRIA*, Campus de Beaulieu, 35042 Rennes Cédex,France
[4] Dpt. Signal & Image Proc., *INT*, 9 rue Charles Fourier, 91011 Evry Cédex, France
[5] Dpt. Image, *ENST*, 46 rue Barrault, 75634 Paris Cédex 13, France
[6] *INRIA*, BP 93, 2004 Route des lucioles, 06902 Sophia-Antipolis Cédex, France

ABSTRACT

The need for hierarchical statistical tools for modeling and processing image data, as well as the success of Markov Random Fields (MRFs) in image processing, have recently given rise to a significant research activity on hierarchical MRFs and their application to image analysis problems. Important contributions, relying on different models and optimization procedures, have thus been recorded in the literature. This paper presents a synthetic overview of available models and algorithms, as well as an attempt to clarify the vocabulary in this field. We propose to classify hierarchical MRF-based approaches as *explicit* and *implicit* methods, with appropriate subclasses. Each of these major classes is defined in the paper, and several specific examples of each class of approach are described.

Keywords: Statistical image models, hierarchical models, Markov random fields, multigrid-, multiscale-, multiresolution-, pyramidal-, models and algorithms.

1 INTRODUCTION TO HIERARCHICAL MRFs

Multiresolution techniques first aimed at increasing the computational speed of relaxation algorithms in numerical analysis [48]. Since the pioneering work of Rosenfeld and Thurston in computer vision and image processing [94], hierarchical image description through different semantic levels or primitives has revealed to be more and more important due to Marr's paradigm preeminence [69]. Considerable work has been achieved, dealing with hierarchical image representation and analysis, mostly in a deterministic context [70–72,79–81].

Hierarchical approaches are today involved in many applicative problems of image analysis such as restoration, segmentation, motion estimation, coding, synthesis, stereo-vision... by means of "multigrid" techniques, "multiresolution" approaches, "multiscale" representations or in a more general way, "hierarchical" models. Although the definition of these different terms seems to be context-dependent and strongly overlapping, the following

0-8194-1927-3/95/$6.00

general motivations for using hierarchical methods in image processing are encountered:

- like in numerical analysis, there is a strong need for decreasing the computational complexity of most image processing algorithms [97];
- the increase of available data in imagery due to technological advances has resulted in various resolution data images of a single scene that have to be merged into a single representation: this is the case of multisource/multispectral imagery for example;
- physical phenomena such as fractal signals, $1/f$ stochastic processes, phase transitions and turbulence possess an intrinsic multiscale nature that has to be modeled;
- most advanced models now used in image processing lead to hard non linear optimization problems (involving many local minima). Hierarchical image representations and image processing algorithms are known to partially alleviate this problem, by smoothing out local minima and allowing faster propagation of interactions between local and global primitives (pixels, regions, interpretation labels, graphs...). This results in faster convergence towards solutions of improved quality as well as a reduced sensitivity to initial configurations.

In addition, the recent success of stochastic models and methods in image processing, in particular Markov Random Field (MRF) representations [23], gave birth to a wide variety of hierarchical MRF approaches and multigrid stochastic algorithms in image analysis [6,11,13,20,45,58,90]. Hierarchical approaches can in this respect, be viewed along three complementary levels:

1. *hierarchical decomposition of data*: pyramid transformation of initial data, wavelet decomposition, integration of various hierarchical information sources;
2. *deterministic or stochastic multiscale image modeling*: wavelet-based models, multiresolution stochastic processes, etc;
3. *hierarchical algorithm definition*: multigrid techniques, renormalization group-based algorithms, etc.

This field has been very active in the past few years, and therefore a synthetic overview as well as a vocabulary clarification appear to be useful. In this paper, we will use the term "hierarchical model" because of its generality, allowing to take into account the specificities of each of the approaches mentioned above. A critical review of the bibliography in this field led us to distinguish two types of hierarchies according to the fact that this term qualifies directly the modeling problem or only its (algorithmic) resolution.

The first case will be referred to as **explicit hierarchy** because the hierarchy is integrated in the model definition. On the opposite, the second case will be referred to as **induced hierarchy** since the hierarchy derives from the use of a stochastic or non stochastic transformation, aiming at the ease of problem resolution and algorithmic implementation. We notice that whatever the hierarchy type, i.e. explicit or induced hierarchy, the fundamental properties rely on a *partial or a total order relationship* operating on various spaces (see Fig. 1).

Induced hierarchical approaches have first been used to simplify the optimization problems in image processing. These approaches include the renormalization techniques and related approaches [44,45,52], or the decomposition by sequences of constrained configuration spaces [12,50,53].

Most of the image processing tasks modeled through Markov random fields (MRFs) lead to hard non-linear optimization problems. This has motivated the development of techniques allowing to ease the search for an optimal solution. To this end, by using a stochastic or deterministic transformation operating on the set of sites (renormalization group techniques) and transformations operating on the configuration space in terms of restrictions, a sequence of ordered configuration spaces is constructed. Assuming that the effect of such a transformation

on the initial optimization problem can be tracked, one gets not a single optimization problem but a sequence of nested optimization tasks whose complexity decreases with the size of the configuration space.

Within the induced hierarchical framework, considering such transformations, leads to approximate the initial optimization problem (e.g., renormalization group techniques) or to exactly solve the optimization problem (e.g., constrained spaces) by using an iterative resolution (Fig. 1).

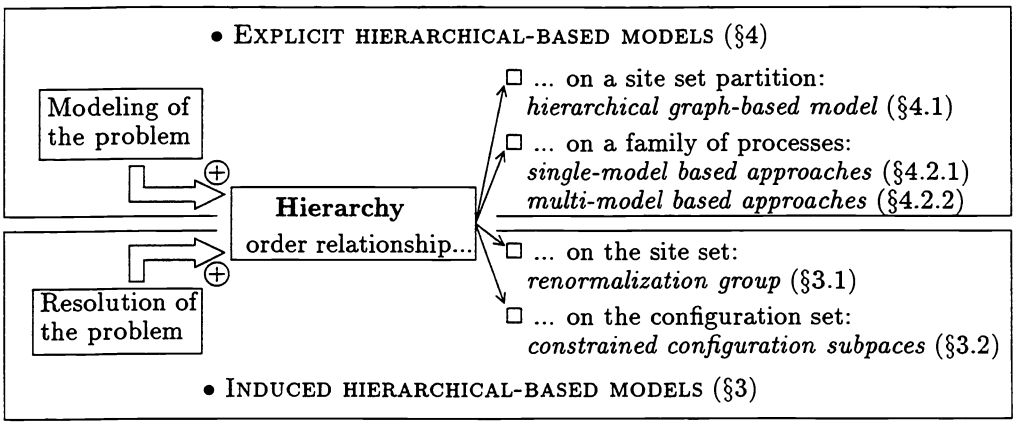

Figure 1: Hierarchical approach synopsis:
• Explicit hierarchical-based models: original modeling + order relationship on a given set (sites, process family);
• Induced hierarchical-based models: optimization procedure + order relationship on a given set (sites, configurations).

The generality of the hierarchical approaches appeared only later when used in the modeling itself, yielding an **explicit hierarchical representation** (Fig. 1). In this context, the order relationship operates on the graph sites on which the model is defined (for instance, quad-tree models [15,67]) or on a family of stochastic processes defined on various sets of sites which may be deterministically or stochastically constructed (for instance, conditional models [90,92]), yielding a wide range of possibilities.

In the first case, a (unique) stochastic process is defined on a set of sites S which is itself divided into a family of grids related by a partial order relation of type "parent-child". In general, the information finally selected is the one obtained at the "maximal" level of the partition (i.e., at the "finest" resolution). Here, the hierarchy follows in fact from a property of the graph of sites, and we will say that these are models defined on hierarchical graphs.

In the second case, a family of processes is constructed, each process being defined on a set of sites which may be variable and even conditionally defined. The stochastic processes are successively defined by an initial process and a stochastic or deterministic definition of each process conditionally to the "previous" one.

We feel that the capacity of hierarchical models has certainly not yet been fully explored. This paper is devoted to a comprehensive bibliography of existing literature on hierarchical MRFs in image processing and to a classification of the hierarchical models based on:

1. the induced or explicit character of the hierarchy;
2. the nature of the space on which the order relationship operates.

Section 2 introduces basic mathematical concepts for defining MRFs on finite simple nondirected graphs. Specifically, graphs and their neighborhood systems, as well as trees and pyramidal structures are presented. The

Markovian property is then recalled together with the standard use of MRFs in terms of optimization-based Bayesian estimation. In Section 3, we mathematically define the concept of induced hierarchical-based models. We distinguish the case for which the order relationship is directly applied to a set of sites (illustrated by the renormalization group approach) from the one for which ordered constrained configuration subspaces are defined. Section 4 deals with explicit hierarchical-based models. The hierarchical graph-based model is first defined and discussed. The ordered families of models are introduced and the specific mono- or multi-model based approaches are presented.

2 MATHEMATICAL STRUCTURES: GRAPHS AND MRFs

A graph structure defined on a set of image sites allows to support representation of local properties in images. More specifically, finite simple nondirected graphs provide the natural tool to support the definition of the local Markov property.

Besides, pyramidal structures allow to represent image features at different scales or resolutions. Many variations around this idea have been introduced in the literature: quad-trees, pyramids, adaptive pyramids, stochastic pyramids (see [54,77]). A proper definition of these structures is allowed within the graph framework.

For these two reasons, we present MRFs as stochastic processes indexed by sites of finite simple nondirected graphs.

2.1 Graphs and pyramidal structures

Neighborhood system and graph A *simple finite nondirected graph* $G = (S, \mathcal{V})$ is defined by its *finite* site set S and a collection $\mathcal{V} = \{\mathcal{V}_s, s \in S\}$ of nonempty subsets of S such that:

$$\begin{aligned} &(i) && \forall s \in S,\ s \notin \mathcal{V}_s, \\ &(ii) && \forall \{s_1, s_2\} \subset S,\ s_1 \in \mathcal{V}_{s_2} \Leftrightarrow s_2 \in \mathcal{V}_{s_1}. \end{aligned}$$

This collection $\mathcal{V}$ is called a *neighborhood system*, and for each site s, $\mathcal{V}_s$ is the set of its neighbors (or neighborhood). The graph is *simple* since there is no loop (edge whose endpoints coincide), and it is *nondirected* due to the symmetric neighborhood relationship.

A *clique* of G is a nonempty subset c of S such that $|c| = 1$, or $|c| > 1$ and any two distinct sites of c are neighboring. In this paper, the clique set of G will be denoted by C.

In image analysis, the most common structure is the bidimensional lattice (i.e. portion of $\boldsymbol{Z}^2$) equipped with a *4-connexity* (or *8-connexity*) *neighborhood system.*

Trees A *chain* of G is a sequence $(s_1, \ldots, s_n)$ of successively neighboring sites ($s_{i+1} \in \mathcal{V}_{s_i}$ for $i = 1, \ldots, n-1$). If $s_1 = s_n$, the chain is called a *cycle.* The graph is *connected* if for any two sites s and t, there exists a finite chain joining them. In the remainder G will always be meant to be connected.

A *tree* is a connected graph without cycles. One can show that a graph is a tree if and only if there exists one and only one chain between any two sites. Some reference site (the *root*) being chosen, trees naturally support a simple hierarchical structure by associating with each site s the length, denoted by $l(s)$, of the chain joining it to the root. If $l(s) = k$, s has a unique neighbor t, called its *parent*, such that $l(t) = k-1$. The remaining neighbors, called its *children*, are at "distance" $k+1$ to the root.

In image analysis, a simple *pyramidal structure* is widely used: it is the *quad-tree* in which each site (apart from those of the finest "level") has 4 children (Fig. 2.a). A similar structure, the *dyadic tree*, is used for monodimensional signal processing.

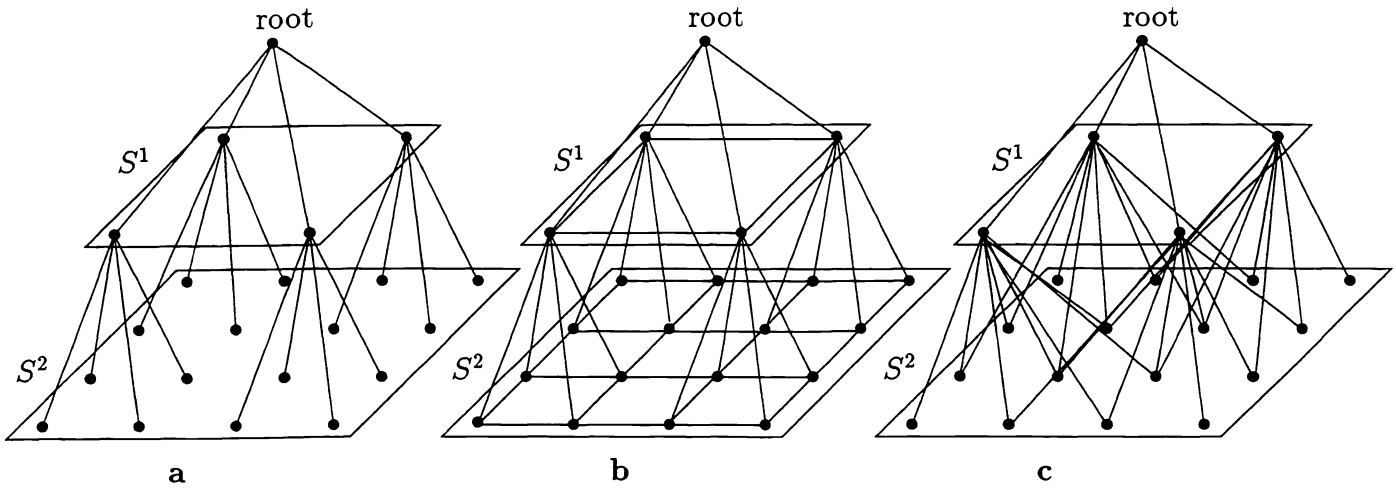

Figure 2: **a**– quad-tree, **b** and **c**– examples of pyramidal graphs based on the quad-tree

Pyramidal graphs More sophisticated pyramidal structures have been defined by adding extra edges to a tree. Level k of a tree being the subset $S^k \triangleq \{s \in S : \ l(s) = k\}$, all sites of a given level are mutually non neighboring. The extra connections in pyramidal graphs may concern only pairs of sites of the same level. For example, each level may be turned into a connected subgraph [58,101] (Fig. 2.b). The extra edges may also be added between consecutive levels, increasing for each site the number of children, and the number of fathers as well [16] (Fig. 2.c). In any case, the new graph thus obtained is not a tree any more.

2.2 Markov random fields on graphs

Definition S is still a finite set of sites and $\mathcal{V}$ a neighborhood system on it. Let A be a finite set, representing the possible values for the considered image features (grey levels for example). A *random field* on S with state space A is a collection $X = \{X_s, s \in S\}$ of A-valued random variables defined on some probability space $(\Omega, \mathcal{A}, p)$ and indexed by sites of S. A realization of X (usually called a *configuration*) will be denoted by $x = \{x_s, s \in S\}$. The (finite) set of all possible configurations is $\Lambda \triangleq A^S$. A random field X is a *Markov Random Field* with respect to $\mathcal{V}$ (or on graph $G = (S, \mathcal{V})$) if and only if, $\forall s \in S, \ \forall x \in \Omega$:

$$(i) \quad \mathsf{Pr}\{X = x\} > 0,$$
$$(ii) \quad \mathsf{Pr}\{X_s = x_s | X_r = x_r, \ r \in S - \{s\}\} = \mathsf{Pr}\{X_s = x_s | X_r = x_r, \ r \in \mathcal{V}_s\}.$$

The Markov property (ii) expresses that values taken by neighboring random variables of some X_s suffice to get access to the local conditional distribution of X_s given realizations of all others.

An important characterization of MRFs is given by Hammersley and Clifford's theorem [9], in terms of *global* distribution $\mathsf{Pr}\{X = x\}$. Namely, X is a MRF with respect to $\mathcal{V}$ if and only if it follows a *Gibbs distribution*

$$\mathsf{Pr}\{X = x\} = \frac{1}{Z} \exp\{-U(x)\}, \ \forall x \in \Omega,$$

where Z is a normalization constant and function U, called the *energy function*, exhibits a decomposition of the form:

$$U(x) = \sum_{c \in C} V_c(x).$$

Function V_c is called a *clique potential (function)* associated with c and *depends only* on variables $\{x_s, s \in c\}$.

Estimation issue The observed data (images) being viewed as a realization y of some random field Y, and $X = \{X_s, s \in S\}$ being the random field modeling the "hidden" variables of interest (called *labels*), one looks for the "best" configuration $\hat{x}$, in terms of Bayesian estimation. The most widely used estimation criterion is the *maximum a posteriori* (MAP) estimator for which the best estimate is the most probable given $Y = y$. With proper assumptions on the observation likelihood $\Pr\{Y = y|X = x\}$ and on a prior distribution (assumed to be Markovian), X turns out to be Markovian given $Y = y$:

$$\Pr\{X = x|Y = y\} = \frac{1}{Z(y)} \exp\{-\underbrace{\sum_{c \in C} V_c(x; y)}_{U(x;y)}\}.$$

Therefore the MAP estimator leads to a *global minimization* problem:

$$\hat{x} = \underset{x \in \Lambda}{\operatorname{argmin}}\, U(x; y). \tag{1}$$

This minimization is an intricate issue since Λ is usually extremely large (or of high dimension), and the objective function to be minimized is often non-convex with numerous local minima. The research of the global minimum (or of some "satisfactory" local minimum) is in general carried out with *iterative relaxation* algorithms which take advantage of the local dependency structure of the energy to operate low-cost local updatings (typically, one variable at a time) [9,10,42].

3 INDUCED HIERARCHICAL-BASED MODELS

In this class of models, a cascade of reduced optimization problems is *derived* from the original one (Eq. 1), either by deriving new random fields defined on a hierarchy of *ordered site sets*, or by restricting the original problem to an *ordered sequence of configuration subspaces* (Fig. 3).

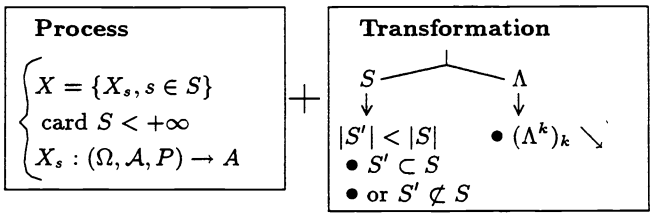

Figure 3: Induced hierarchical-based model

3.1 Renormalization group: ordered relationship on sites

A general, consistent definition of the renormalization method widely used in Statistical Mechanics has been given in [45], on the basis of hierarchical conditional random processes. Given site sets $S^0, S^1, \ldots S^n \equiv S$ (with $|S^0| < |S^1| < \ldots < |S^n|$), new label fields ($X^{n-1}$ on S^{n-1},...,X^0 on S^0) are defined from $X^n \equiv X$ through transition probability laws $\Pr\{X^{k-1} = x^{k-1}|X^k = x^k\}$ between consecutive random processes (Fig. 4.a).

Decimation and block scalings are particular cases of this general procedure, although the most important ones. Decimation successively restricts the random process to a decreasing sequence of subsets of the original

lattice S ($S^{k-1} \subset S^k$):

$$\mathsf{Pr}\{X^{k-1} = x^{k-1} | X^k = x^k\} = \prod_{s \in S^{k-1}} \delta(x_s^{k-1}, x_s^k),$$

whereas, in block scaling, each set S^k is partitioned into "blocks", each of them corresponding to a unique site of S^{k+1}. Then for each site s of S^{k-1}, the attached random variable X_s^{k-1} is only function of components of X^k lying within the associated block $B^k(s) \subset S^k$. For instance, block averaging is defined as the following process:

$$\mathsf{Pr}\{X^{k-1} = x^{k-1} | X^k = x^k\} = \prod_{s \in S^{k-1}} \delta\left(x_s^{k-1}, \frac{1}{|B^k(s)|} \sum_{r \in B^k(s)} x_r^k\right).$$

In all cases, the induced posterior probability law at level $k-1$ is given by:

$$\mathsf{Pr}\{X^{k-1} = x^{k-1} | Y = y\} = \sum_{x^k} \mathsf{Pr}\{X^{k-1} = x^{k-1} | X^k = x^k) \times \mathsf{Pr}\{X^k = x^k | Y = y\}.$$

Such scaling transformations, with particular attention paid to the conditions of existence of a restricted MRF and to the important case of Gaussian MRFs, are thoroughly considered and described in [44,52,64,88]. It is shown that in general $\mathsf{Pr}\{X^k = x^k | Y = y\}$, $k < n$, is not a local Gibbs distribution: long range interactions have appeared through the renormalization process. For practical convenience, the distribution has to be *approximated* by a local Gibbs law $\mathsf{Pr}\{X^k = x^k | Y^k = y^k\} \propto \exp -U^k(x^k; y^k)$ where y^k is a convenient reduction of y. To our knowledge, only three MRF models transform into MRF by decimation renormalization (allowing *exact* calculations): 2D 4-connexity Ising and Potts models [44] and 2D, 4-connexity Gaussian models [32,39].

The hierarchical image processing method based on these transformations can be described as follows: first, transforms of the original energy function are derived or approximated at all scales *if possible* (bottom-up); then, the algorithm itself results in successive multiscale coarse-to-fine optimizations with top-down "projections":

$$\hat{x}^k = \underset{x^k \in \Lambda^k(\hat{x}^{k-1})}{\operatorname{argmin}} U^k(x^k; y^k), \text{ with } \Lambda^k(x^{k-1}) \triangleq \underset{x^k}{\operatorname{argmax}} \mathsf{Pr}\{X^{k-1} = x^{k-1} | X^k = x^k\}$$

Stochastic optimization methods such as simulated annealing (SA) are generally used to this end although deterministic schemes have also been proposed [45,86]. Significant gains on CPU time have been observed, with respect to standard (non-hierarchical) optimization techniques.

Applications include binary [45] and gray level image restoration [39,84,85], image segmentation [76], motion segmentation [89], and texture parameters estimation for classification purposes [32].

3.2 Ordered-constrained configuration spaces

Inspired from multigrid techniques in numerical analysis [48], the use of nested constrained configuration subspaces has been proposed independently in [13,50,51,87]. This method allows to define consistent hierarchical MRFs starting from a single full resolution MRF. Their principle consists in solving a sequence of global optimization problems deduced from the original one (Eq. 1) by constraining visited configurations to belong to a sequence of embedded configuration subspaces $\Lambda^0 \subset \Lambda^1 \subset \ldots \Lambda^n \equiv \Lambda$:

$$\text{search } \underset{x \in \Lambda^k}{\arg\min}\, U(x; y) \text{ from } k = 0 \text{ to } k = n.$$

The constraints are chosen to be consistent with the prior distribution (e.g., piecewise constancy for smoothing prior) and such that dimension of Λ^k decreases with k. Then $x \in \Lambda^k$ is completely described through a reduced set of variables $x^k = \{x_s^k, s \in S^k\}$ (with $|S^k| < |S|$) interacting with the original data y according to the deduced "coarse" energy function $U^k(x^k; y) \triangleq U(x; y)$ (see Fig. 4.b for the structure corresponding to piecewise

constant configurations over $2^{n-k} \times 2^{n-k}$ blocks of S) [50,51,87]. Then one has to solve the following sequence of optimization problems:

$$\text{search } \underset{x^k}{\text{argmin}}\, U^k(x^k; y) \text{ from } k = 0 \text{ to } k = n.$$

In contrast with the renormalization group transform approach presented before, the induced models defined by coarse energy functions exhibit a Markovian locality similar to the one of the original model. As for the associated estimation algorithms, "coarse-to-fine" strategies are naturally used: the estimated field at level k is associated (thanks to the model definition) with a *unique* configuration at level $k+1$ of the hierarchy. This configuration is used as an initial guess for relaxation at this level. Experiments show that even when deterministic optimization algorithms are used, they lead to estimation results close to those obtained by stochastic methods while they converge 60 to 100 times faster. Such methods have been applied to motion analysis [12,50,87], image classification [57,58], image restoration [51,53,55] and texture segmentation [13].

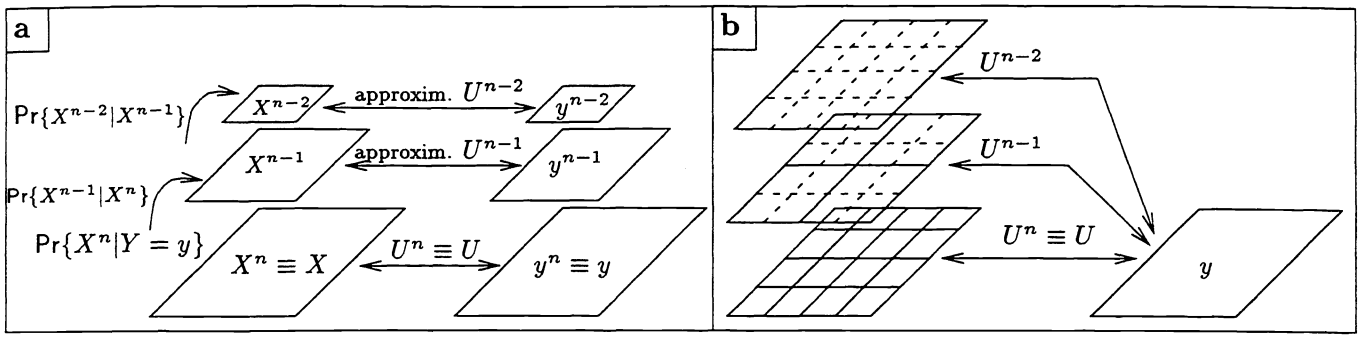

Figure 4: a- renormalization group, b- constrained configuration subspaces

4 EXPLICIT HIERARCHICAL-BASED MODELS

As previously mentioned in Section 1 (see Fig. 1), explicit hierarchical-based models rely on an explicit order relationship. This relationship operates (a) either on the site set on which the global and unique model is defined, leading to a site set partition, (b) or on a family of processes. The corresponding models will respectively be referred to as *hierarchical graph-based models* and *ordered family of models.*

4.1 Hierarchical graph-based models

In this class of approaches, a *unique* Markov model $X = \{X^0, \ldots, X^n\}$ is defined on a *hierarchically partitioned* graph $G = (S = \cup_{k=0}^{n} S^k, \mathcal{V})$ (e.g., a dyadic tree, a quad-tree or a pyramidal graph like those of Fig. 2, etc.), with $|S^k| < |S^{k+1}|$, and X^k being the restriction of X to site subset S^k. This partition of S induces a hierarchical partition of the neighborhood of each site (Fig. 5):

$$\forall s \in S, \mathcal{V}_s = \bigcup_{k=0}^{n} \underbrace{\mathcal{V}_s \cap S^k}_{\triangleq \mathcal{V}_s^k}.$$

Some particular properties of the model under concern can be viewed in terms of this neighborhood partitions. However, the graph is always chosen such that the upper neighborhood $\mathcal{V}_s^{k-1}$ of $s \in S^k$ (if $k > 0$) and the lower neighborhood $\mathcal{V}_s^{k+1}$ of $s \in S^k$ (if $k < n$) are nonempty.

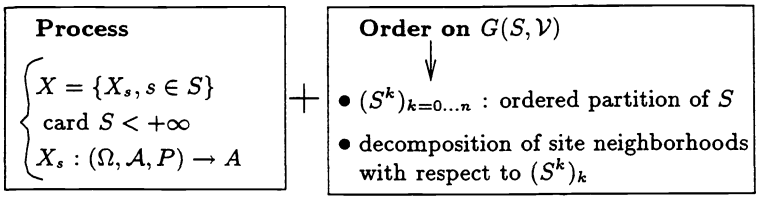

Figure 5: Hierarchical graph-based models

The Markov model is specified on the hierarchical graph by defining a global energy function $U(x^0, \ldots, x^n; y)$:

- either directly, as a proper sum of local potentials. In [58,101] for instance, sites have neighbors at their own level (see Fig. 2.b for an illustration of the kind of structure under concern): $\forall s \in S^k$, $\mathcal{V}_s^k \neq \emptyset$;
- or by mean of transition probabilities $\mathsf{Pr}\{X^k = x^k | X^{k-1} = x^{k-1}\}$ which express causal statistical interactions between different levels [14–16,65,67,35,36], along with the coarsest prior distribution $\mathsf{Pr}\{X^0 = x^0\}$ (Fig. 6.a). The prior distribution is then defined as:

$$\mathsf{Pr}\{X = x\} = \mathsf{Pr}\{X^0 = x^0\} \prod_{k=1}^{n} \mathsf{Pr}\{X^k = x^k | X^{k-1} = x^{k-1}\}.$$

  As a result of this second construction, a site has no neighbors at its own level: $\forall s \in S^k$, $\mathcal{V}_s^k = \emptyset$ (the quad-tree of Fig. 2.a is an illustration of this case). Gaussian models [67,65] as well as non-linear Markov models have been defined this way on quad-trees [67] for image analysis and on dyadic trees [6–8,26–28] for monodimensional signal processing, using state models [6,65,67]. In this particular case, the only nonempty neighborhood parts of $s \in S^k$ are $\mathcal{V}_s^{k+1}$ (set of children), and $\mathcal{V}_s^{k-1}$ which contains an unique site, the parent of s (see Section 2). A simple linear model of that type is provided by the recursion $X_s^k = X_{\text{parent}}^{k+1} + W_s$, where W stands for a white noise indexed by elements of S.

As for the interaction with data, observations y are plugged in either at only one level (usually the larger one, S^n, corresponding to the finest resolution), or at all of them, in case of multiresolution data (Fig. 6).

The kind of estimation algorithm associated with this class of Markov models essentially depends on the graph structure and on the characteristics of the model (causal/non-causal, linear/non-linear model). In general, the optimization of the global energy requires some iterative relaxation scheme [58,101]. Nevertheless, when the structure of the hierarchical model is related to a tree (dyadic tree for monodimensional signals [6] or quad-tree for an image [15,65,67]), the absence of cycles allows the design of *non-iterative* estimation algorithms [35,36]. Viterbi-like estimation algorithms [15,63], Kalman filter-like algorithms [65,67] as well as more specific non-linear algorithms [67] have been described in this framework. The computational load can be reduced significantly by using causal models rather than standard MRFs defined on a lattice, as illustrated in [65,67]. Let us however notice that the use of a quad-tree graph, as observed by several authors [15,67], induces blocky-shaped artefacts in the final estimates. More sophisticated (unfortunately non-causal) models [15,16] have to be defined to alleviate this effect.

The main applications related to hierarchical models defined on a tree are supervised classification [15], texture analysis [65,67], optical flow measurement in a video sequence [67] and multisensor-multiresolution image data fusion [63]. As for the hierarchical models defined by a pyramidal graph, the first application has been supervised [58,101] and unsupervised [59] image classification. Another potential field of application is multisensor data fusion.

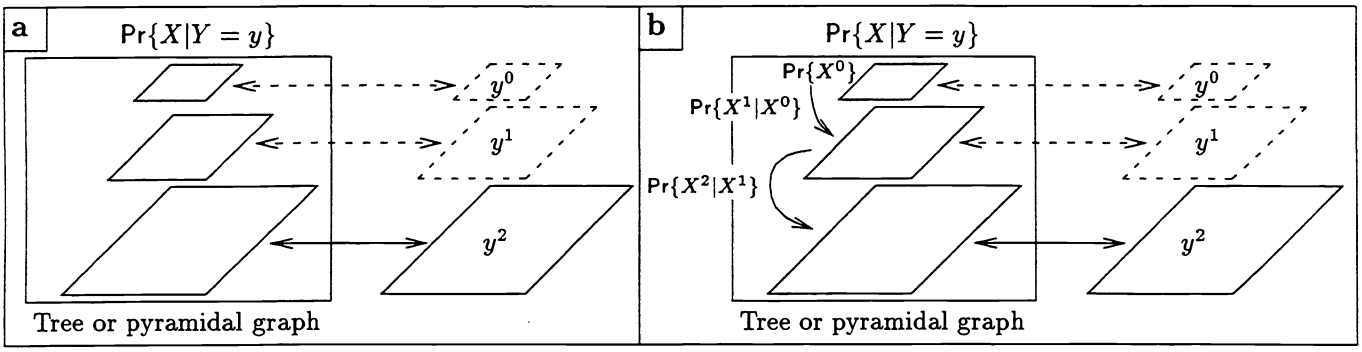

Figure 6: **a**- generic hierarchical graph-based model, **b**- hierarchical graph-based model causaly defined (from 0 to n)

4.2 Ordered family of models

We consider here an ordered family of processes $(X_k)_k$:

$$\begin{cases} X^k = \{X_s^k, s \in S^k\} \\ X_s^k : (\Omega^k, \mathcal{A}^k, P^k) \longrightarrow A^k \end{cases}$$

In the literature, degenerated cases of such family models are proposed, all the processes following the "same" law. Such hierarchical approaches will be referred to as *single-model based approaches*, the general case being called *multi-model based approaches.*

4.2.1 Single-model based approaches

The single-model approaches essentially rely on an heuristic association of MRF models with multiresolution decomposition of image data. Multiresolution image transforms based on Gaussian filters [4,5,54,73,74] or wavelet decompositions [3,11,20,71,72] are used to decompose the image data y on a pyramidal structure, providing a data set $(y^0, y^1, \ldots, y^n \equiv y)$. Given the original MRF model of energy $U(x;y)$, a similar reduced MRF model of energy $U^k(x^k;y^k)$ is then defined at each resolution level k (same neighborhood structure and same potential functions as the original model, i.e. U^k "similar" to U) (Fig. 7). Standard stochastic or deterministic relaxation algorithms provide Bayesian estimates of the label field at each resolution level. These algorithms are generally associated to a coarse-to-fine exploration of the hierarchical structure, starting from the lowest resolution and propagating the estimates from coarse to fine scales.

Two techniques for propagating information through scale have been described. In most of the cases, the final estimate obtained at a given level is interpolated to be used as an initialization for the relaxation process at the next finer level [3,5,4,11,20,73,74]

$$\hat{x}^k = \underset{x^k}{\operatorname{argmin}}\, U^k(x^k;y^k) \quad \text{and} \quad \text{Interpol}_{k\to k+1}(\hat{x}^k) = \text{initial guess at level } k+1.$$

A second technique consists in computing an *incremental* label field at each resolution level: the solution obtained at a given scale is incrementally refined at the next resolution level, and the final full resolution estimate $\hat{x}$ is the sum of properly interpolated estimates of all levels:

$$\hat{x} = \hat{x}^n + \sum_{k=0}^{n-1} \text{Interpol}_{k\to n}(\hat{x}^k).$$

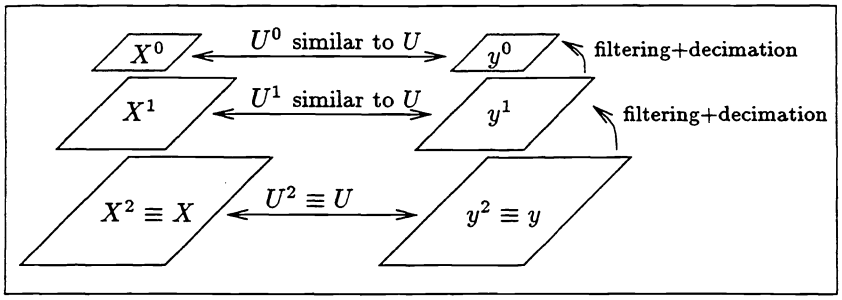

Figure 7: Heuristic cascade of reduced versions on the original model

For example, the 2D velocity field estimate in a sequence of moving images has to be decomposed into various amplitude scales to recover large displacements using differential methods [49,61].

In some applications, methods for adjusting the parameters of the model through scale have been suggested [73]. These methods remain however generally ad hoc, and the derivation of mathematically consistent methods for computing the model parameters at each scale, is not addressed in this class of approaches.

Applications of the single model approach include stereovision [5], optical flow computation [11,49,61], image coding [3], image restoration [20] and region-based image segmentation [73,74].

4.2.2 Multi-model based approaches

The most general hierarchical Markovian model consists of a family of stochastic processes, not necessarily all Markovian.

The Markov models are then defined on graphs that may differ from one level to another one, and use a large variety of potentials. The primitives modeled at each level of the hierarchy (such as micro and macro-textures, labels, regions, motion parameters,...) are in general different. The relationship between the hierarchy levels can be either of a conditional probabilistic type, or of a sequential deterministic type. They can also involve various levels of interpretation such as semantic and symbolic descriptions.

More precisely, in the studied papers, a finite sequence of processes $(X^k)_k$ is constructed, X^{k+1} being defined conditionally to X^k mainly in two ways: either the law of X^{k+1} given X^k is specified, or the graph structure associated with X^{k+1} is defined conditionally to X^k in a deterministic way.

A hierarchy defined by a stochastic conditional dependency is proposed in [90]. In this article, two processes are involved: a Boolean process and a Markovian one.

A sequential aspect relying on an alternative use of two levels of interpretations is illustrated in [2] and [91]. In both papers, the first process is defined on the lattice associated with the image data, while the second one is defined on a graph which depends on the current configuration of the first process, thus dynamically evolving.

Finally, an example of successive use of stochastic models leading to higher interpretation levels is presented in [92].

Applications of the multi-model based approaches include restoration and segmentation [2,91], texture modeling and synthesis [90] as well as 3D object recognition [92].

5 CONCLUSION

In this article, we have proposed a unified framework within MRF context for so-called multiresolution, multigrid and multiscale approaches in image analysis. The concept of "hierarchical models" based on a notion of order relationship has been introduced, and a classification into two types of hierarchies, namely *induced* and *explicit hierarchies*, has been presented. The literature mentioned in the present paper has been critically analysed with respect to the developed classification. Specificities and characterization of the major contributions within the following subclasses have been mathematically discussed:

- the approaches related to the renormalization group transformations;
- the optimization schemes based on an exploration of constrained configuration subspaces;
- the models relying on a hierarchical partition of the Markovian graph structure;
- the ordered families of processes within a general of degenerated context.

Let us notice that this review mainly takes into account contributions up to end 1994. Nevertheless, new papers around hierarchical MRF models have been published. This demonstrates the intense research activities currently developed in the field of hierarchical approaches to statistical image analysis.

6 ACKNOWLEDGMENTS

This work has been supported by the DRED (Direction de la Recherche et des Études Doctorales) in the context of a national program with CNRS GdR 134 (project "Hierarchical Markov models for image analysis").

7 REFERENCES

[1] R. AZENCOTT, B. CHALMOND, and P. JULIEN, "Bayesian 3D path search and its applications to focusing seismic data," In *Lecture Notes in Statistics*, Springer Verlag, 1991.

[2] R. AZENCOTT and C. GRAFFIGNE, "Non supervised segmentation using multi-level Markov Random Fields," In *Int. Conf. Pattern Rec.*, pp. III.201–III.204, The Hague, The Netherlands, Sept. 1992.

[3] M. BARLAUD, L. BLANC-FÉRAUD, and P. CHARBONNIER, "Image coding using multiresolution Markov random fields," In *Proc. SPIE, Image Processing Algorithms and Techniques III*, San Jose, USA, Feb. 1992.

[4] S. BARNARD, "Stereo Matching," In R. CHELLAPPA and A. JAIN, editors, *Markov Random Fields - theory and application*, pp. 245–272. Academic Press, 1993.

[5] S.T. BARNARD, "Stochastic stereo matching over scale," *Int. J. Comp. Vis.*, Vol. 3, pp. 17–32, 1989.

[6] M. BASSEVILLE, A. BENVENISTE, K.C. CHOU, S.A. GOLDEN, R. NIKOUKHAH, and A.S. WILLSKY, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Information Theory*, Vol. 38, No. 2, pp. 766–784, 1992.

[7] M. BASSEVILLE, A. BENVENISTE, and A.S. WILLSKY, "Multiscale autoregressive processes - Part I: Schur-levinson parametrizations," *IEEE Trans. Signal Processing*, Vol. 40, No. 8, pp. 1915–1934, 1992.

[8] M. BASSEVILLE, A. BENVENISTE, and A.S. WILLSKY, "Multiscale autoregressive processes - Part II: Lattice structures for whitening and modeling," *IEEE Trans. Signal Processing*, Vol. 40, No. 8, pp. 1935–1954, 1992.

[9] J. BESAG, "Spatial interaction and the statistical analysis of lattice systems," *J. Royal Statist. Soc.*, Serie B, Vol. 36, pp. 192–236, 1974.

[10] J. BESAG, "On the statistical analysis of dirty pictures," *J. Royal Statist. Soc.*, Vol. 48, Serie B, No. 3, pp. 259–302, 1986.

[11] L. BLANC-FÉRAUD, M. BARLAUD, and T. GAIDON, "Motion estimation involving discontinuities in a multiresolution scheme," *Optical Engineering*, Vol. 32, No. 7, pp. 1475–1482, 1993.

[12] C. BOUMAN and B. LIU, "A multiple resolution approach to regularization," In *Proc. SPIE Vol. 1001 Visual Communications and Image Processing'88*, pp. 512–520, 1988.

[13] C. BOUMAN and B. LIU, "Multiple resolution segmentation of textured images," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No. 2, pp. 99–113, 1991.

[14] C. BOUMAN, "A multiscale image model for Bayesian image segmentation," Technical Report TR-EE 91-53, Purdue University, December 1991.

[15] C. BOUMAN and M. SHAPIRO, "Multispectral image segmentation using a multiscale model," In *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing*, pp. III.565–III.568, San Francisco, USA, April 1992.

[16] C. BOUMAN and M. SHAPIRO, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Processing*, Vol. 3, No. 2, pp. 162–177, 1994.

[17] P. BOUTHEMY and E. FRANÇOIS, "Motion segmentation and qualitative dynamic scene analysis from an image sequence," *Int. J. Computer Vision*, Vol. 10, No. 2, pp. 157–182, 1993.

[18] P. BOUTHEMY and P. LALANDE, "Recovery of moving object masks in an image sequence using local spatio-temporal contextual information," *Optical Engineering*, Vol. 32, No. 6, pp. 1205–1212, 1993.

[19] B. CHALMOND, "Image restoration using an estimated Markov model," *Signal Processing*, Vol. 15, No. 2, pp. 115–129, 1988.

[20] P. CHARBONNIER, L. BLANC-FÉRAUD, and M. BARLAUD, "Noisy image restoration using multiresolution Markov random fields," *J. of Visual Com. and Image Representation*, Vol. 3, No. 4, pp. 338–346, 1992.

[21] R. CHELLAPPA, "Time series models for multiresolution images," In A. Rosenfeld, editor, *Multiresolution Image Processing and Analysis*, pp. 102–108. Springer Verlag, 1984.

[22] R. CHELLAPPA and S. CHATTERJEE, "Classification of textures using gaussian markov random fields," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 33, No. 4, pp. 959–963, 1985.

[23] R. CHELLAPPA and A.K. JAIN, editors, *Markov Random Fields: Theory and Applications*, Academic Press, Boston, 1993.

[24] R. CHELLAPPA and R.L. KASHYAP, "Texture synthesis using 2-D noncausal autoregressive models," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 33, No. 1, pp. 194–203, 1985.

[25] P.B. CHOU and C.M. BROWN, "The theory and practice of Bayesian image modeling," *Int. J. Comp. Vis.*, Vol. 4, pp. 185–210, 1990.

[26] K.C. CHOU, S.A. GOLDEN, and A.S. WILLSKY, "Multiresolution stochastic models, data fusion and wavelet transforms," *Signal Processing*, Vol. 34, No. 3, pp. 257–282, 1993.

[27] K.C. CHOU, A.S. WILLSKY, and A. BENVENISTE, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Trans. Automatic Control*, Vol. 39, No. 3, pp. 464–478, 1994.

[28] K.C. CHOU, A.S. WILLSKY and R. NIKOUKHAH, "Multiscale systems, Kalman filters and Riccati equations," *IEEE Trans. Automatic Control*, Vol. 39, No. 3, pp. 479–492, 1994.

[29] F.S. COHEN and D.B. COOPER, "Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 9, No. 2, pp. 195–219, 1987.

[30] G.R. CROSS and A.K. JAIN, "Markov random field texture models," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 5, No. 1, pp. 25–39, 1983.

[31] H. DERIN and H. ELLIOT, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 9, No. 1, pp. 39–55, 1987.

[32] X. DESCOMBES, M. SIGELLE, and F. PRÊTEUX, "Estimating Gaussian Markov random field parameters in a non-stationnary framework: application to remote sensing imaging," Submitted to *IEEE Trans. Image Proc.*, 1994.

[33] J.M. DINTEN, "Tomographic reconstruction of axially symmetric objects: Regularization by a markovian modelization," In *Proc. Int. Conf. Pattern Recognition*, volume 2, pp. 153–158, Atlantic City, USA, June 1990.

[34] R.C. DUBES and A.K. JAIN, "Random field models in image analysis," *J. Applied Statistics*, Vol. 16, No. 2, pp. 131–163, 1989.

[35] E. FABRE, *Traitement du signal multirésolution : conception de lisseurs rapides pour une famille de modèles* (in French), PhD thesis, Université de Rennes 1, France, Dec. 1994.

[36] E. FABRE, "New fast smoothers for multiscale models," Technical Report 857, IRISA, Sept. 1994.

[37] P. FLANDRIN, "On the spectrum of fractional Brownian motions," *IEEE Trans. Inf. Theory*, Vol. 35, No. 1, pp. 197–199, 1989.

[38] E. FRANCOIS and P. BOUTHEMY, "Multiframe-based identification of mobile components of a scene with a moving camera," In *proc. IEEE Int. Conf. Computer Vision Pattern Recognition*, pp. 166–172, Hawaii, USA, June 1991.

[39] D. GEIGER and J.E. KOGLER, "Scaling images and image features via the renormalization group," In *Proc. IEEE Int. Conf. Computer Vision Pattern Recognition*, pp. 47–53, New-York, USA, June 1993.

[40] D. GEMAN, S. GEMAN, C. GRAFFIGNE, and D. PONG, "Boundary detection by constrained optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 12, No. 7, pp. 609–628, 1990.

[41] D. GEMAN and G. REYNOLDS, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 14, No. 3, pp. 367–383, 1992.

[42] S. GEMAN and D. GEMAN, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No. 6, pp. 721–741, 1984.

[43] S. GEMAN and D.E. McCLURE, "Statistical methods for tomographic image reconstruction," In *Proc. 46th Session Int. Statistical Institute (ISI)*, Vol 52, 1987.

[44] B. GIDAS, "A multilevel-multiresolution technique for computer vision via renormalization group ideas," In *SPIE Vol. 880, High Speed Computing*, pp. 214–218, 1988.

[45] B. GIDAS, "A renormalization group approach to image processing problems," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No. 2, pp. 164–180, Feb. 1989.

[46] C. GRAFFIGNE and C. LABOURDETTE, "Some segmentation algorithms and their parallel implementation," In *Proc. 49th session Int. Statistical Institute (ISI)*, pp. I199-I215, Firenze, 1993.

[47] C. GRAFFIGNE, F. HEITZ, F. PRÊTEUX, M. SIGELLE, and J. ZERUBIA, *Modèles markoviens hiérarchiques pour l'analyse d'images* (in French), Ed. CNRS-GdR TdSI, November 1994.

[48] W. HACKBUSCH, *Multigrid methods and applications*, Springer-Verlag, Berlin, 1985.

[49] F. HEITZ and P. BOUTHEMY, "Multimodal estimation of discontinuous optical flow using Markov Random Fields," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 15, No. 12, pp. 1217–1232, 1993.

[50] 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, pp. 125–134, 1994.

[51] M. HURN and C. JENNISON, "A study of simulated annealing and a revised cascade algorithm for image reconstruction," Technical Report 93:04, Univ. of Bath, UK, April 1993.

[52] F.C. JENG, "Subsampling of Markov random fields," *J. of Visual Com. and Image Representation*, Vol. 3, No. 3, pp. 225–229, 1992.

[53] C. JENNISON and M. JUBB, "Statistical image restoration and refinement," In M.A. Viergever, editor, *Information Processing in Medical Imaging*, pp. 255–262. Plenum Press, New-York, USA, 1988.

[54] J.M. JOLION and A. ROSENFELD, "*A pyramid framework for early vision*," Kluwer Academic Press, 1994.

[55] M. JUBB and C. JENNISON, "Aggregation and refinement in binary image restoration," In Possolo, editor, *Spatial Statistics and Imaging*, pp. 150–162. Institute of Mathematical Sciences Lecture Notes, Hayward, USA, 1991.

[56] Z. KATO, J. ZERUBIA and M. BERTHOD, "Bayesian image classification using Markov random fields," In *Proc. 12th Int. Maxent Workshop*, pp. 375–381, Paris, France, July 1992.

[57] Z. KATO, M. BERTHOD, and J. ZERUBIA, "A hierarchical Markov random field model and multi-temperature annealing for parallel image classification," Research Report 1938, INRIA, Aug. 1993.

[58] Z. KATO, M. BERTHOD, and J. ZERUBIA, "Multiscale Markov random field models for parallel image classification," In *Proc. Int. Conf. Computer Vision*, pp. 253–257, Berlin, Germany, May 1993.

[59] Z. KATO, J. ZERUBIA, and M. BERTHOD, "Unsupervised parallel image classification using a hierarchical Markovian model," In *Proc. Int. Conf. Computer Vision*, Cambridge, USA, June 1995.

[60] J.E. KOGLER and D. GEIGER, "Scaling images and image features via the renormalization group," Technical report, Universidade São Paulo-LSI/EPUSP, 1993.

[61] J. KONRAD and E. DUBOIS, "Bayesian estimation of motion vector fields," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 14, No. 9, pp. 910–927, 1992.

[62] J.-M. LAFERTÉ, P. PÉREZ, and F. HEITZ, "Global non-linear multigrid optimization for image analysis tasks," In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Adelaïde, Australia, April 1994.

[63] J.-M. LAFERTÉ, F. HEITZ, P. PÉREZ, and E. FABRE, "Hierarchical Statistical Models for the Fusion of Multiresolution Image Data," In *Proc. Int. Conf. Computer Vision*, Cambridge, USA, June 1995.

[64] S. LAKSHMANAN and H. DERIN, "Gaussian Markov random fields at multiple resolutions," In R. Chellappa and A. Jain, editors, *Markov Random Fields - Theory and application*, pp. 131–157. Academic Press, 1993.

[65] M. LUETTGEN, *Image processing with multiscale stochastic models*, PhD thesis, MIT Laboratory of Information and Decision Systems, May 1993.

[66] M. LUETTGEN, W. KARL, A. WILLSKY, and R. TENNEY, "Multiscale representations of Markov random fields," *IEEE Trans. Signal Processing*, Vol. 41, No. 12, pp. 3377–3395, 1993.

[67] M.R. LUETTGEN, W. KARL, and A.S. WILLSKY, "Efficient multiscale regularization with applications to the computation of optical flow," *IEEE Trans. Image Processing*, Vol. 3, No. 1, pp. 41–64, 1992.

[68] M.R. LUETTGEN and A.S. WILLSKY, "Likelihood calculation for a class of multiscale stochastic models," Technical Report CICS-P-370, MIT, April 1993.

[69] D. MAAR, *Vision: a computational investigation into the human representation and processing of visual information*," W.H. Freeman, San-Francisco, 1982.

[70] S. MALLAT, "Multifrequency channel decomposition of images and wavelet models," *IEEE Trans. Acoustic Speech Signal Proc.*, Vol. 37, No. 12, pp. 2091–2110, 1989.

[71] S. MALLAT, "Multiresolution approximation and wavelet orthonormal bases of L^2," *Trans. American Mathematical Society*, Vol. 315, pp. 69–87, 1989.

[72] S. MALLAT, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intel.*, Vol. 11, No. 7, pp. 674–693, 1989.

[73] F. MARQUES, J. CUNILLERA, and A. GASULL, "Hierarchical segmentation using compound Gauss-Markov random fields," In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. III.53–III.56, San Francisco, USA, 1992.

[74] F. MARQUES and A. GASULL, "Stochastic image model for segmentation. Application to image coding," In *Proc. Scandinavian Conf. Image Anal.*, pp. 265–272, Tromso, Norway, 1993.

[75] J.L. MARROQUIN, S. MITTER, and T. POGGIO," "Probabilistic solution of ill-posed problems in computational vision," *J. American Statis. Assoc.*, Vol. 82, pp. 76–89, 1987.

[76] I. MATSUBA, "Renormalization group approach to hierarchical image analysis," In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. 1044–1047, New-York, USA, April 1988.

[77] P. MEER, "Stochastic image pyramids," *Comput. Vision, Graphics, Image Processing*, Vol. 45, pp. 269–294, 1989.

[78] E. MÉMIN, F. HEITZ, and F. CHAROT, "Efficient parallel multigrid relaxation algorithms for Markov Random Field-based low-level vision applications," In *IEEE Conf. Comp. Vis. Patt. Rec.*, pp. 644–648, Seattle, USA, June 1994.

[79] Y. MEYER, *Ondelettes et opérateurs. Volume I : ondelettes*, Hermann, Paris, 1990.

[80] Y. MEYER, *Ondelettes et opérateurs. Volume II : opérateurs de Caderon-Zygmund*, Hermann, Paris, 1990.

[81] Y. MEYER, *Les ondelettes : algorithmes et applications*, A. Colin Ed., Paris, 1992.

[82] J.W. MODESTINO and J. ZHANG, "A Markov random field model-based approach to image interpretation," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol 14, No. 6, pp. 606–615, 1992.

[83] D.W. MURRAY and H. BUXTON, "Scene segmentation from visual motion using global optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol 9, No. 2, pp. 220–228, 1987.

[84] G. K. NICHOLLS, "An order preserving hierarchy of Markov random fields," In *Proc. New Zealand Conf. Image and Vision Computing*, pp. 205–212, Awckland, New Zealand, 1993.

[85] G. K. NICHOLLS and M. PETROU, "A generalisation of renormalization group methods for multiresolution image analysis," In *Proc. Int. Conf. Pattern Recognition*, pp. I.567–I.570, The Hague, The Netherlands, 1992

[86] G. K. NICHOLLS and M. PETROU, "Multiresolution representation of Markov random fields," Technical Report VSSP-TR-3/93, University of Surrey, 1993.

[87] P. PÉREZ and F. HEITZ, "Global bayesian estimation, constrained multiscale Markov random fields and the analysis of visual motion," In *Proc. 12th Int. Maxent Workshop*, pp. 383–388, Paris, July, 1992.

[88] P. PÉREZ and F. HEITZ, "Restriction of a Markov random field on a graph and multiresolution image analysis," Technical Report No. 784, IRISA, January 1994, submitted to *IEEE Trans. Inf. Theory.*

[89] M. PETROU, M. BOBER, and J. KITTLER, 'Multiresolution motion segmentation," In *Proc. Int. Conf. Pattern Recognition*, pp. 379–383, Jerusalem, Israel, Oct. 1994.

[90] F. PRÊTEUX and X. DESCOMBES, "Synthèse et analyse de textures par coopération de procesus multi-échelles" (in French), In *Proc. 8ème congrès AFCET/INRIA Reconnaisance des Formes et Intelligence Artificielle*, pp. 1015–1022, Lyon, France, 1991.

[91] C.S. REGAZZONI, F. ARDUINI, and G. VERNAZZA, "A multilevel GMRF-based approach to image segmentation and restoration," *Signal Processing*, Vol. 34, No. 1, pp. 43–67, 1993.

[92] C.S. REGAZZONI, G. FORESTI, and V. MURINO, "Distributed hierarchical regularization system for recognition of planar surfaces," *Optical Engineering*, Vol. 32, No. 6, pp. 1167–1192, 1993.

[93] A. ROSENFELD, editor, *Multiresolution image processing and analysis*, Springer, New-York, 1984.

[94] A. ROSENFELD and M. THURSTON, "Edge and curve detection for visual scene analysis," *IEEE Trans. Computers*, Vol. 20, pp. 262–569, 1971.

[95] M. SIGELLE and R. RONFARD, "Relaxation of previously classified images by a Markov field technique and its relationship with statistical physics," In *Proc. Scandinavian Conf. Image Anal.*, pp. 387–394, Aalborg, Danmark, August 1991.

[96] J. SUBRAHMONIA, Y.P. HUNG, and D.B. COOPER, "Model-based segmentation and estimation of 3D surfaces from two or more intensity images using Markov random fields," In *Proc. Int. Conf. Pattern Recognition*, volume 1, pp. 390–397, Atlantic City, USA, June 1990.

[97] D. TERZOPOULOS, "Image analysis using multigrid relaxation methods," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 8, No. 2, pp. 129–139, 1986.

[98] A.H. TEWFIK and M. KIM, "Correlation structure of the discrete wavelet transform coefficients of fractional Brownian motion," *IEEE Trans. Inf. Theory*, Vol. 38, pp. 904–910, 1992.

[99] G.W. WORNELL, "A Karhunen-Loeve-like expansion for 1/f processes via wavelets," *IEEE Trans. Inf. Theory*, Vol. 36, No. 4, pp. 859–861, 1990.

[100] J. ZERUBIA and R. CHELLAPA, "Mean Field annealing using compound Gauss-Markov Random Fields for edge detection and image estimation," *IEEE Trans. Neural Networks*, Vol. 4, No. 4, pp. 703–709, 1993.

[101] J. ZERUBIA, Z. KATO and M. BERTHOD, "Multi-temperature annealing: a new approach for the energy minimization of hierarchical Markov random field models," In *Proc. Int. Conf. Pattern Recognition*, Jerusalem, Israel, Oct. 1994.