UNSUPERVISED TEXTURE SEGMENTATION USING A STATISTICAL WAVELET-BASED HIERARCHICAL MULTI DATA MODEL

F. Destrempes ‡ *M. Mignotte* ‡

‡ DIRO, Département d'Informatique et de Recherche Opérationnelle, C.P. 6128,
Succ. Centre-Ville, Montréal, Canada (Quebec), H3C 3J7.
E-MAIL : DESTREMP@IRO.UMONTREAL.CA

ABSTRACT

In this paper, we describe a new hidden Markov random field model, which we call hierarchical multi-data model, and which is based on a triplet of random fields (two hidden random fields and one observed field) in order to capture inter-scale and within-scale dependencies between various scales of resolution of wavelet-based texture features. We present a variation of the Iterated Conditional Modes (ICM) algorithm for the segmentation, and an adaptation of the Iterative Conditional Estimation (ICE) procedure for the estimation of the statistical parameters of the model. Results of tests performed on 75 mosaics of Brodatz textures are reported.

1. INTRODUCTION

Hidden Markov random field (HMRF) models have shown to be fundamental in understanding the problem of image segmentation. After the pioneer work of [1], [2], various hierarchical versions have been presented in order to speed up convergence. Multiresolution (MR) models (for instance [3]) consider a same HMRF but at various levels of resolutions, whereas in multi-model (MM) approaches [4], the HMRF varies with the level of resolution.

One can also consider only the full resolution but analyzed at various scales, such as in the hierarchical multi-grid (HMG) model of [5] (where no inter-scale relation is modeled). In [6], a hierarchical multi-scale (HMS) model is introduced in order to take into account the inter-scale dependency. In the HMS models of [7] and [8], both the inter-scale dependencies and the spatial within-scale relations are considered.

One can also consider different data at each scale, such as in the HMS model of [9] (where no spatial within-scale relation is modeled). In the HMS of [10], the spatial within-scale and inter-scale dependencies are captured in a contextual model, and the data varies according to the scale of wavelet coefficients. In [11], the previous model is extended to a joint multi-context and multi-scale (JMCMS) model.

We present in this paper a new HRMF model, which we call hierarchical multi-data (HMD) model (see Section 2). The model is based on a triplet of random fields : the observed field in which the data can vary from one scale to another ; one hidden discrete field that represents the texture classes ; and one auxiliary hidden discrete field that is used in modeling mixtures of distributions. When a Sequential Maximum A Posteriori (SMAP) recursive algorithm is adopted, the data gets truncated when passing from one scale to the next finer scale, whereas the parent label is inherited. Thus, fusion of decisions is performed. In this paper, we consider the data as a whole, as is done in [7]. Thus, we use the HMD model to perform fusion of data, the decision being taken at the bottom level.

The data features represent a wavelet-based description of the gray levels contained in windows of various sizes, in a manner closely related to [12], and are modeled using mixtures of Probabilistic Principal Component Analyzers (PPCA) [13] (Section 4). In contrast, [10] and [11] use the wavelet coefficients as modeled by [14]. We present a variation of the Iterated Conditional Modes (ICM) algorithm [2] for the segmentation (Section 3), and an adaptation of the Iterative Conditional Estimation (ICE) procedure [15] for the estimation of the statistical parameters of the model (Section 5). We report in Section 6, the results of tests performed on 75 mosaics of Brodatz textures.

2. A HIERARCHICAL MULTI-DATA (HMD) MODEL

2.1. The triplet (X, Y, C)

We consider a standard segmentation problem in which one wants an optimal realization (in the sense of some statistical criterion) of a discrete hidden random field X given an observed random field Y. Both X and Y are defined on a graph G whose nodes form a set S ; i.e., the components of the vectors X and Y are attached to the nodes s of the graph : $X = (X_s)_{s \in S}$ and $Y = (Y_s)_{s \in S}$. We assume that the graph G has a hierarchical structure organized in J levels (or scales) : the set of nodes S of G is partitioned into J subsets S^n, $1 \leq n \leq J$, such that any node $s \in S^n$ at level $l(s) = n$ is connected only to nodes belonging to levels $n-1$, n or $n+1$. Furthermore, we assume that for each node s at the top level J, the subgraph $G(s)$ of G obtained by gathering the inter-level descendants of s (i.e. the subset $\{t \ : \ d(s,t) = l(s) - l(t)\}$, where $d(s,t)$ denotes the distance between the nodes s and t in the graph G) forms a tree of depth $J-1$. For simplicity, we will consider in what follows a quad-tree graph structure on each $G(s)$, $s \in S^J$, so that S^1 corresponds to the pixels of the image, and for each level $n = 2, ..., J$, S^n corresponds to blocks of $2^{n-1} \times 2^{n-1}$ pixels. Moreover, at level n, each block of $2^{n-1} \times 2^{n-1}$ pixels is connected to its usual 8 within-level neighboring blocks of pixels. For $n = 1, ..., J$, the restrictions of X and Y to the level S^n are denoted X^n and Y^n, respectively. We assume that each random variable X_s takes its value in a fixed finite set of classes $\Lambda = \{e_1, ..., e_K\}$, no matter the level of s. Moreover, for a given level n, we assume that each random variable Y_s, $s \in S^n$, takes its values in a fixed state space Υ^n ; the state spaces Υ^n are allowed to vary from one level to another. In this setting, each level corresponds to a possibly distinct set of data (or features) y^n attached to the image, that allows making a decision

as to what should be the realization of X^n ; moreover, these realizations of X^1, X^2, ..., X^J are of same nature, in the sense that they correspond to the same set of classifying hidden labels. The prior distribution $P(x)$ is defined up to a constant by

$$\prod_{s\in S} P(x_s \mid x_{N(s)}) \prod_{s\notin S^J} P(x_s \mid x_{N(\mathcal{P}(s))}),$$

where $x_{N(s)}$ is the restriction of x to the set of 8 within-level neighbors of s, and $\mathcal{P}(s)$ is the inter-level parent of s. So, the prior distribution captures within-level as well as inter-scale relations of the labeling field X.

We also consider an auxiliary hidden random field C on the graph G, where at each level n, C_s takes its values in a finite set of auxiliary labels $\Lambda^n = \{g_1^n, ..., g_{K_n}^n\}$, possibly different from one level to another. We assume that the likelihood $P(y \mid c, x)$ admits a site-wise decomposition

$$\prod_{s\in S} P(y_s \mid c_s, x_s);$$

i.e., the components of Y are mutually independent given X and C, and $P(y_s \mid c, x) = P(y_s \mid c_s, x_s)$. Moreover, the two hidden fields are related by a likelihood $P(c \mid x)$ of the form (up to a constant)

$$\prod_{s\notin S^J} P(c_s \mid c_{\mathcal{P}(s)}, x_s) \prod_{s\in S} P(c_s \mid x_s).$$

Altogether, the joint distribution of the triplet of random fields (X, Y, C) is given by $P(x, y, c) = P(y \mid c, x)P(c \mid x)P(x)$.

2.2. Examples of HMD models in the literature

In the HMS model presented in [9], one takes $\Lambda^n = \Lambda$, for $n = 1$ to J, and $C = X$. Furthermore, the inter-scale distributions $P(x_s \mid x_{N(\mathcal{P}(s))})$ reduce to $P(x_s \mid x_{\mathcal{P}(s)})$, and the within-scale distributions $P(x_s \mid x_{N(s)})$ are the uniform distributions, except at the top level J, in which case it is of the form $P(x_s)$.

In the HMS model of [10] and [11], the auxiliary hidden variables C_s take their values in the label sets $\Lambda^n = \{H, L\}^3$ for each n (corresponding to the coefficients of the three wavelet subbands HL, HH, and LH at scale n), whereas the classifying set Λ represents the texture classes in the image. For each site s, the vector $x_{N(s)\setminus\{s\}} \cup x_{N(\mathcal{P}(s))}$ is called there the context vector of s.

2.3. The couple (X, Y)

In the framework of HMD models, the distribution $P(y \mid x)$ is equal to the sum $\sum_c P(y, c \mid x)$. This sum can be expressed as a product of the form $\prod_{s\in S^J} f_{x,y}(s)$. Each factor $f_{x,y}(s)$ depends on all the nodes of the quad-tree based at the root s and is given by

$$\sum_{c_s} P(y_s \mid c_s, x_s)P(c_s \mid x_s) \prod_{t\in d^-(s)} \mathcal{F}_{x,y}(t, c_s),$$

where $d^-(s)$ is the set of inter-level children of s. For $t \notin S^J$, $\mathcal{F}_{x,y}(t, c_{\mathcal{P}(t)})$ is defined recursively by

$$\sum_{c_t} P(y_t \mid c_t, x_t)P(c_t \mid c_{\mathcal{P}(t)}, x_t)P(c_t \mid x_t) \prod_{r\in d^-(t)} \mathcal{F}_{x,y}(r, c_t),$$

and at the leaves, we have

$$\mathcal{F}_{x,y}(t, c_{\mathcal{P}(t)}) = \sum_{c_t} P(y_t \mid c_t, x_t)P(c_t \mid c_{\mathcal{P}(t)}, x_t)P(c_t \mid x_t).$$

3. MAP SEGMENTATION USING HMD MODELS

One could use the SMAP recursive algorithm on the couple (X, Y) of the HMD model, in order to do a segmentation of the image as a fusion of decisions. In this paper, a different point of view is adopted : fusion of the data located at the various levels is performed, leaving the decision at the bottom level. In order to do so, we restrict the realizations of X to those of the form $(x^1, x^2, ..., x^J)$, where each x_s at level $n > 1$ is obtained from the realization x^1 at the lower level, by voting over the corresponding block of size $2^{n-1} \times 2^{n-1}$ (taking the label of minimal index in case of a tie). In that setting, we denote x by $\tilde{x}^1$. We then consider the segmentation of an image in the sense of the MAP.

Thus, we want to minimize the average cost $E[C(X^1, x^1)|Y = y]$, where the cost function $C(X^1, x^1)$ is defined by $1 - \delta(X^1 - x^1) = 1 - \delta(\tilde{X}^1 - \tilde{x}^1)$. The solution is given by

$$\begin{aligned}\hat{x}^1 &= \arg\max_{x^1} P(\tilde{x}^1 \mid y)\\ &= \arg\min_{x^1} -\ln\{P(y \mid \tilde{x}^1)\} - \ln\{P(\tilde{x}^1)\}.\end{aligned}$$

For a given pixel $s \in S^1$, the corresponding local energy term is given by

$$\begin{aligned}-\ln\{f_{\tilde{x}^1,y}(s^J)\} &- \sum_{n=1}^{J} \ln\{P(\tilde{x}^1_{s^n} \mid \tilde{x}^1_{N(s^n)})\}\\ &- \sum_{n=1}^{J-1} \ln\{P(\tilde{x}^1_{s^n} \mid \tilde{x}^1_{N(s^{n+1})})\}\end{aligned}$$

where s^n denotes the ancestor of s located at the nth level.

One could use the Simulated Annealing (SA) algorithm [1] or the Iterated Conditional Modes (ICM) algorithm [2] directly at the bottom level. Instead, we use what we call a coarse-to-fine ICM algorithm. This approach consists in taking at each intermediate level n_0, starting with the top level, the partial energy term

$$\begin{aligned}-\ln\{f_{\tilde{x}^{n_0},y}(s^J)\} &- \sum_{n=n_0}^{J} \ln\{P(\tilde{x}^{n_0}_{s^n} \mid \tilde{x}^{n_0}_{N(s^n)})\}\\ &- \sum_{n=n_0}^{J-1} \ln\{P(\tilde{x}^{n_0}_{s^n} \mid \tilde{x}^{n_0}_{N(s^{n+1})})\}\end{aligned}$$

where $\tilde{x}^{n_0}$ is obtained by voting at level n_0, and $f_{\tilde{x}^{n_0},y}(s^J)$ is computed as $f_{\tilde{x}^1,y}(s^J)$, but ending the recursion at level n_0 rather than 1 (i.e., as if levels $n < n_0$ were removed from G).

Thus, taking $n_0 = J$, we start with a segmentation $\hat{x}^J$ of S^J in the sense of the Maximum Likelihood (ML) :

$$\hat{x}^J_s = \arg\min_e -\ln\{\sum_{c_s} P(y_s \mid c_s, e)P(c_s \mid e)\}.$$

We then update dynamically the resulting realization $\hat{x}^J$ by sweeping in a fixed order the sites s of S^J recursively, choosing at each site s the label

$$\hat{x}^J_s = \arg\min_e -\ln\{f_{\hat{x}^J(s,e),y}(s)\} - \ln\{P(e \mid \hat{x}^J(s, e)_{N(s)})\}$$

where $\hat{x}^J(s, e)$ is obtained from the current realization $\hat{x}^J$ by setting the label to e at the site s. The recursion comes to an end when no more site is modified through an entire sweep.

A segmentation $\hat{x}^{n+1}$ of S^{n+1} obtained at level $n+1$ ($n < J$) is passed on to the next lower level by projection onto S^n. We then update the resulting realization $\hat{x}^n$ by scanning S^n, using the partial energy term presented above to update the label at a site s. At the bottom level ($n_0 = 1$), the complete energy term is used.

4. A HMD MODEL BASED ON WAVELET TEXTURE FEATURES

4.1. Assumptions on the HMD model

In what follows, we make the assumption that $C^J = X^J$ and $|\Lambda^n| = |\Lambda|$, $1 \leq n \leq J$, and that $P(c_s \,|\, x_s) = \delta(c_s - x_s)$ whenever $l(s) = J$. Also, we assume that the likelihoods $P(y_s \mid c_s, x_s)$ are of the simpler form $P(y_s \mid c_s)$. We model the distributions $P(c_s | c_{\mathcal{P}(s)}, x_s)$ and $P(x_s | x_{N(\mathcal{P}(s))})$ by uniform distributions. Finally, we take an isotropic Potts model for $P(x_s \,|\, x_{N(s)})$. Note that these assumptions simplify greatly the computation of the MAP segmentation (the details are omitted here).

4.2. Description of the texture features

For each level $n = 1, 2, ..., J$ in the pyramid, we consider at every block of $2^{n-1} \times 2^{n-1}$ pixels in the image, a window of size $N \times N = 2^n \times 2^n$ centered at that block. The spatial configuration of the gray levels of the image contained in this window, up to translations, is viewed as the raw data y'_s associated to the corresponding abstract site $s \in S^n$ at level n. We perform a transformation on the raw data y'_s by using a wavelet-based filter bank, as follows.

Fixing N, let h be a low-pass filter defined on the discrete interval $[0, ..., N-1]$ (we consider periodic discrete signals of period N). The complementary high-pass filter is defined by $g(k) = (-1)^{k+1} h(-k+1)$. We assume the low-pass condition $\sum_k h(k) = 1$, and the standard condition $|\hat{h}(k)|^2 + |\hat{g}(k)|^2 = 1$, where $\hat{h}$ denotes the discrete Fourier transform of h. For $i \geq 2$, low-pass filters h_i and band-pass filters g_i are defined recursively by the relations

$$\begin{aligned} \hat{h}_0(l) &= 1, \\ \hat{h}_{i+1}(l) &= \hat{h}(2^i l)\hat{h}_i(l), \quad i \geq 0 \\ \hat{g}_{i+1}(l) &= \hat{g}(2^i l)\hat{h}_i(l), \quad i \geq 0. \end{aligned}$$

One can show that the identity $|\hat{h}_I(l)|^2 + \sum_{i=1}^{I} |\hat{g}_i(l)|^2 = 1$ holds for any $I \geq 1$, so that a full coverage of the frequency domain is provided. In particular, any signal f admits a decomposition of the form $f = h_I * (h_I^T * f) + \sum_{i=1}^{I} g_i * (g_i^T * f)$, where $I \geq 1$, and $f^T(k) = f(-k)$. In this paper, we take $I = 1$ at the bottom level, and $I = n - 1$ at the other levels n. We then obtain a filter bank for periodic discrete signals of dimension 2 and period $N \times N$, by considering for $1 \leq i \leq I$, the standard filters [16] $g_i * h_i$, $g_i * g_i$, $h_i * g_i$, as well as the low-pass filter $h_I * h_I$.

Now, given a vector of raw features y'_s (viewed as a two-dimensional signal f), we consider the square of the mean of the filter response $h_I * h_I * f$, and for all other filters, the energy of the filter response. This yields a vector of transformed data y_s of dimension $3I + 1$. Note that y_s is invariant under translations of the raw data y'_s, which is a desirable property for texture segmentation. See [12] for the exact analogue in the case of non-periodic signals, using the z-transform. Here, we normalize each feature independently between 0 and 1 on the whole image.

4.3. Statistical model for the texture features

For each hidden auxiliary class c_s, we consider a Probabilistic Principal Component Analyzer (PPCA) model [13] defined by a Gaussian distribution of the form

$$P(y_s \mid c_s) = \mathcal{N}(y_s; \nu, \sigma^2 I_d + WW^t)$$

where $\sigma^2 > 0$ represents the variance of the noise in the data, ν is the average d-dimensional vector of features (with d the dimension of y_s), and W is the reconstruction $d \times q$ matrix (where q is the reduced dimension). Here, the statistical parameters σ^2, ν, W depend on the class c_s. In [13], it is proved that the ML estimators on a sample of i.i.d. observations $t_1, ..., t_m$ issued from that model, are given by the closed formulas

$$\begin{aligned} \sigma^2 &= \frac{1}{d-q} \sum_{i=q+1}^{d} \lambda_i, \\ \nu &= \bar{t} = \frac{1}{m} \sum_{i=1}^{m} t_i, \\ W &= U_q (\Lambda_q - \sigma^2 I_q)^{1/2}, \end{aligned}$$

where $\lambda_1, ..., \lambda_d$ are the eigenvalues of the sample covariance matrix, in decreasing order, Λ_q is the diagonal matrix with entries $\lambda_1, ..., \lambda_q$, and U_q is the $d \times q$ matrix with columns equal to the corresponding eigenvectors, normalized so that they have euclidean norm equal to 1 (i.e., the columns of U_q span the principal subspace of the sample covariance matrix).

5. ESTIMATION OF THE HMD MODEL OF SECTION 4

In order to estimate the HMD model (X, Y, C) described in Section 4, we use an adaptation of the Iterative Conditional Estimation (ICE) procedure [15], [17], [18].

Each distribution $P(y_s \mid g_i^n)$ depends on a vector of statistical parameters $\Phi_i^n = (\sigma_{i,n}^2, \nu_{i,n}, W_{i,n})$, where n is the level of s. Together with the probabilities $P(g_i^n | e_j)$, we obtain a vector of parameters Φ which describes completely the model (X, Y, C). The ICE procedure relies on ML estimators $\hat{\Phi}(X, Y, C)$ of the vector Φ on the "complete data" (x, y, c) ; i.e, given a realization c, we have an ML estimator $\hat{\Phi}_i^n(y, c)$ of Φ_i^n on the subset $\{y_s : s \in S^n,\ c_s = g_i^n\}$; the probabilities $P(c_s \mid x_s)$ can simply be estimated by proportions of occurrences. The goal of the ICE procedure is to minimize the function $||\Phi - E[\hat{\Phi}(X, Y, C) \mid Y, \Phi]||^2$, where $E[\hat{\Phi}(X, Y, C) \mid Y, \Phi] = \sum_{(c,x)} \hat{\Phi}(x, y, c)\, P(c, x \mid y, \Phi)$. The estimation of the HMD parameters for the "incomplete data" y using the ICE procedure is outlined as follows :

I) Estimation of $P(y_s \mid g_i^n)$ for all $n \leq J$

Parameter initialization : for each level n in the HMD model (independently of the other levels), use the K-means algorithm described in [19] to obtain a raw segmentation c^n of S^n. The first estimate $\Phi^{[0]}$ of Φ is then obtained using the ML estimators $\hat{\Phi}_i^n(y, c)$.

Then, $\Phi^{[p+1]}$ is computed recursively from $\Phi^{[p]}$ until convergence is achieved (i.e., $\Phi^{[p+1]} \approx \Phi^{[p]}$), in the following way :

1. **Simulation** : For each level n (independently), simulate one realization of c_s at each site $s \in S^n$ according to the weights $P(y_s \mid c_s)$, with parameter vector $\Phi^{[p]}$.

2. **Estimation** : the parameter vector $\Phi^{[p+1]}$ is estimated with the ML estimator of the "complete data" corresponding to each class, as described above.

II) Estimation of $P(g_i^n \mid e_j)$ for all $n < J$

Parameter initialization : Set $x^J = c^J$ and project x^J to each lower level. The next $\Phi^{[0]}$ of Φ is then obtained using the proportions of occurrences of labels for $P(g_i^n \mid e_j)$.

Then, $\Phi^{[p+1]}$ is computed recursively from $\Phi^{[p]}$ until convergence is achieved (i.e., $\Phi^{[p+1]} \approx \Phi^{[p]}$), in the following way :

1. **Simulation** : Simulate one realization of x^1 using the Gibbs sampler according to the distribution $P(x \mid y)$ with parameter vector $\Phi^{[p]}$.

2. **Estimation** : $P(g_i^n \mid e_j)$ is estimated using the proportions of occurrences of labels.

6. EXPERIMENTAL RESULTS

For our tests, we took 110 Brodatz textures located at the URL "http ://www.ux.his.no/~tranden/brodatz.html", with $\#111$ omitted, and $\#14$ (absent from the site) replaced by $\#112$. We removed the highly non-periodic textures # 2,7,12,13,15,23,30,31,39-45, 58,59,61,62,69-72,88-91,97,99, as well as 48 (the period is too large), 79 (almost equal to 78) and 106 (almost equal to 105). We split the remaining list into 15 groups of 5 consecutive textures (dropping 109-111). We then considered all mosaics of 4 textures taken among the 5 textures of each group. This gave us a collection of 75 mosaics, each of size 640×640, on which we applied a coarse-to-fine ICM segmentation using the Haar wavelet, with $J = 5$ and $|\Lambda| = 4$. We obtained 62.7% (74.7%, 80%, and 88%) of mosaics with less than 5% (10%, 15%, and 20%, respectively) of misclassified pixels. Next, we removed the textures # 27 and 28 (which are somewhat similar to 29), 50 (similar to 51), 74 (similar to 75), 80 (similar to 81), as well as 60, 67, 86 (which are not regular enough). We then obtain 49 mosaics with 83.7% (95.9%, and 100%) of mosaics with less than 5% (10%, and 15%, respectively) of misclassified pixels.

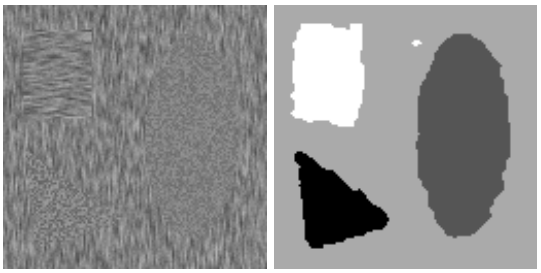

FIG. 1 – 128×128 *mosaic of four synthetic textures and its segmentation using the SA.* $J = 3$ *and* $|\Lambda| = 4$.

7. CONCLUSION

In this paper, we have presented a new hidden Markov random field model that seems well adapted to texture segmentation. Unlike multi-scale approaches, we consider the data as a whole, and hence aim at performing fusion of data, rather than fusion of decisions. The experimental results reported here seem very promising.

8. REFERENCES

[1] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6) :721–741, 1984.

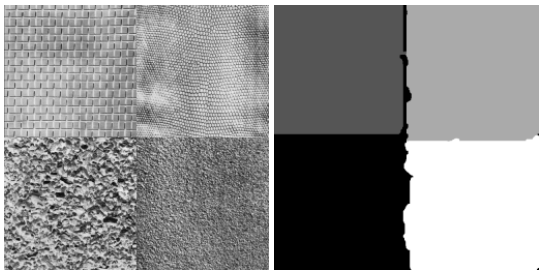

FIG. 2 – 640×640 *mosaic of four Brodatz textures and its segmentation using the coarse-to-fine ICM (*2.4% *of misclassifed pixels).* $J = 5$ *et* $|\Lambda| = 4$.

[2] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, B-48 :259–302, 1986.

[3] M. G. Bello. A combined markov random field and wave-packet transform-based approach for image segmentation. *IEEE Trans. on Geoscience and Remote Sensing*, 31(3) :618–633, 1993.

[4] C. Regazzoni, F. Arduini, and G. Vernazza. A multilevl gmrf-based approach to image segmentation and restoration. *Signal Processing*, 34(1) :43–67, 1993.

[5] P. Pérez. *Champs markoviens et anaylse multirésolution de l'image : application à l'analyse du mouvement.* Thèse de doctorat, Université de Rennes 1, IRISA, 1993.

[6] C. Bouman and M. Shapiro. A multiscale image model for bayesian image segmentation. *IEEE Trans. on Image Processing*, 3(2) :162–177, 1994.

[7] Z. Kato, M. Berthod, and Z. Zerubia. A hierarchical markov random field model and multi-temperature annealing for parallel image classification. *Graph. Mod. Image Process.*, 58(1) :18–37, 1996.

[8] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Sonar image segmentation using an unsupervised hierarchical MRF model. *IEEE Trans. on Image Processing*, 9(7) :1216–1231, 2000.

[9] J.-M. Laferté, P. Pérez, and F. Heitz. Discrete markov image modeling and inference on the quadtree. *IEEE Trans. on Image Processing*, 9(3) :390–404, 2000.

[10] H. Choi and R. Baraniuk. Multiscale image segmentation using wavelet-domain hidden markov models. *IEEE Trans. on Image Processing*, 10(9) :1309–1321, 2001.

[11] G. Fan and X.-G. Xia. A joint multicontext and multiscale approach to Bayesian segmentation. *IEEE Trans. on Geoscience and Remote Sensing*, 39(12) :2680–2688, 2001.

[12] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Trans. on Image Processing*, 4(11) :1549–1560, 1995.

[13] M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal pomponent analyzers. *Neural Computation*, 11(2) :443–482, 1999.

[14] M. S. Cruise, R. D. Nowak, and R. G. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. on Signal Processing*, 46(4) :886–902, 1998.

[15] W. Pieczynski. Champs de Markov cachés et estimation conditionnelle itérative. *Revue Traitement Du Signal*, 11(2) :141–153, 1994.

[16] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[17] B. Braathen, P. Masson, and W. Pieczynski. Global and local methods of unsupervised Bayesian segmentation of images. *GRAPHICS and VISION*, 2(1) :39–52, 1993.

[18] F. Salzenstein and W. Pieczynski. Parameter estimation in hidden fuzzy markov random fields and image segmentation. *Graphical Models and Image Processing*, 59(4) :205–220, 1997.

[19] S. Banks. *Signal processing, image processing and pattern recognition.* Prentice Hall, 1990.