Triplet markov trees for image segmentation

Jean-Baptiste Courbot, Emmanuel Monfrini, Vincent Mazet, Christophe Collet

TRIPLET MARKOV TREES FOR IMAGE SEGMENTATION

Jean-Baptiste Courbot[1,2]*, Emmanuel Monfrini*[3]*, Vincent Mazet*[1]*, Christophe Collet*[1]

[1]ICube, Université de Strasbourg - CNRS, 67412 Illkirch, France
[2]Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, CRAL UMR5574, F-69230, Saint-Genis-Laval, France
[3]Département CITI, CNRS, UMR 5157 SAMOVAR, 91011 Évry, France

ABSTRACT

This paper introduces a triplet Markov tree model designed to minimize the block effect that may be encountered while segmenting image using Hidden Markov Tree (HMT) modeling. We present the model specificities, the Bayesian Maximum Posterior Mode segmentation, and a parameter estimation strategy in the unsupervised context. Results on synthetic images show that the method greatly improves over HMT-based segmentation, and that the model is competitive with a hidden Markov field-based segmentation.

Index Terms— Triplet Markov Tree, Image Segmentation, Unsupervised segmentation

1. INTRODUCTION

Markov models have proven their efficiency in the context of Bayesian image segmentation, *i.e.* the estimation of the classification $\mathbf{X}$ from the observation $\mathbf{Y} = \mathbf{y}$. Among those models, one can mention hidden Markov chains, hidden Markov fields and hidden Markov trees (see for instance [1, 2, 3, 4] and the references therein). The Bayesian segmentation is obtained after choosing an estimator, such as the Maximum Posterior Mode (MPM) [5] or the Maximum A Posteriori (MAP) [2] in an appropriate context.

In general, Markov chains models allow the fastest segmentation and permit the exact computation of the *a posteriori* distribution.While this model is not fully intuitive, it often offers satisfactory robustness. In comparison, the Markov fields framework is richer. This richness is balanced by the impossibility to compute exactly the posterior densities, making it necessary to use sampling methods, such as the Gibbs sampler [2].

A compromise between the model richness and exact computation feasibility may be found with the Hidden Markov Tree (HMT) models, often used within an independent noise assumption [3]. This modeling allows the exploitation, within a segmentation context, of a hierarchy representing the classification to obtain. The main drawback of this hierarchy is the introduction, in the most difficult cases, either of "speckles" or block effects in the segmentation result. Several models were introduced to compensate these effects. For instance, let us mention the *evolutive* Markov tree models [6], in which the parent-child transition probabilities depend on the considered resolution. One can also mention the *hierarchical fields* modeling, in which the prior density is Markovian both spatially (over each scale) and hierarchically [7].

In a similar fashion as Markov chains and Markov fields models, Markov tree models have been enriched by the introduction of *triplets* models [8, 9, 10]. In this framework, in addition to the observed process $\mathbf{Y}$ and the classification $\mathbf{X}$, a third *auxiliary* process denoted $\mathbf{U}$ is adjoined. The objective of such model is the accurate handling of more complex phenomena, such as privileged orientation for instance [11].

In this paper, we introduce a Triplet Markov Tree (TMT) model, in which the auxiliary process modulates the parent-child transition probabilities, depending on the classification of the parent's neighbors. This model will be referred as Spatial Triplet Markov Tree (STMT) and is introduced in Section 2, along with the MPM segmentation computation. We apply this model in the case of a single (high-resolution) in Section 3. The numerical results presented in Section 4 show that our model improves the segmentation of very noisy images over a classical HMT method. We also show that this method is competitive with a combination of HMT and Hidden Markov Field (HMF) models.

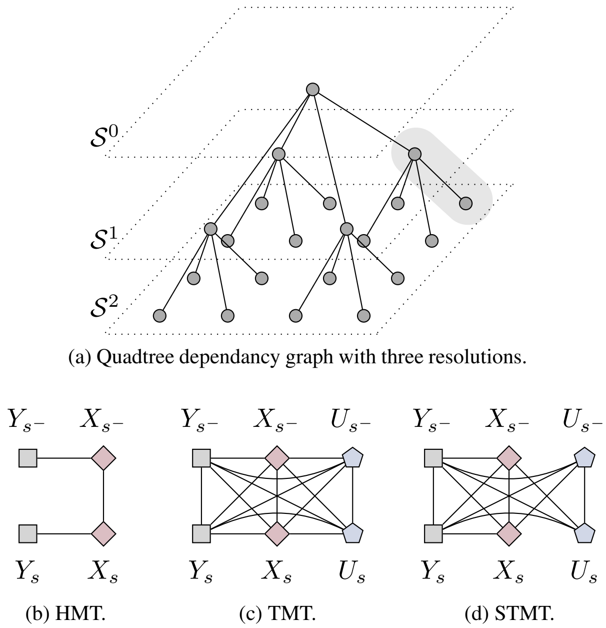
2. SPATIAL TRIPLET MARKOV TREES

2.1. Model

Let $\mathbf{T} = (T_s)_{s\in\mathcal{S}}$ be a stochastic process, where $\mathcal{S}$ is the set of resolutions of a quadtree: $\mathcal{S} = \{\mathcal{S}^0, \dots, \mathcal{S}^N\}$. Each $\mathcal{S}^n$ contains 4^n sites : $n = 0$ represents the tree root, and $n = N$ is the finest resolution. Besides, we set $\mathbf{T} = (\mathbf{Y}, \mathbf{X}, \mathbf{U})$, where $\mathbf{Y} \in \mathbb{R}^{|\mathcal{S}|}$ is an observation process, $\mathbf{X} \in \Omega_x^{|\mathcal{S}|}$ is a class process and $\mathbf{U} \in \Omega_u^{|\mathcal{S}|}$ is an auxiliary process.

J.-B. Courbot acknowledges support from the ERC grant 339659-MUSICOS and partial support from the ANR project DSIM number ANR-14-CE27-0005.

J.-B. Courbot is now with INRIA Paris, MOKAPLAN, rue Simone Iff, 75012, Paris, France and Laboratoire de Météorologie Dynamique, UMR 8539, PSL-ENS/CNRS/UMPC/Ecole Polytechnique, Paris, France



(a) Quadtree dependancy graph with three resolutions.

(b) HMT. (c) TMT. (d) STMT.

Fig. 1: Parent-child transition detail (gray region in (a)) for the different tree models.

Formally, $\mathbf{T}$ is a Markov tree [3] and verifies:

$$p(\mathbf{T}) = p(T_r) \prod_{s \in \mathcal{S} \setminus \mathcal{S}^0} p(T_s | T_{s^-}); \tag{1}$$

where s^- is the parent site of s, and T_r is the value of T_s in the root site $s \in \mathcal{S}^0$.

$\mathbf{U}$ is an auxiliary process tuning the distribution of $\mathbf{X}$ at each parent-child transition. Each U_s is a random vector ruled by the same distribution as $(X_v)_{v \in V_s}$, where V_s is a neighborhood of s to be defined. In the remaining of this paper, we assume that X_s and U_s are independent given T_{s^-}. This yields $\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^N\}$:

$$p(T_s | T_{s^-}) = p(Y_s | X_s, U_s, T_{s^-}) p(X_s | T_{s^-}) p(U_s | T_{s^-}). \tag{2}$$

The three distributions appearing in this equation are specified in Section 3. Figure 1 represents the quadtree dependency graph and the dependencies involved in the HMT, the TMT and the STMT models.

2.2. MPM segmentation

We choose to use the MPM estimator for the segmentation. For concision, we note $p(A = a)$ as $p(a)$ when possible. The MPM estimator requires the computation of $p(X_s = \omega_i, U_s = \nu_j | X_{s^-} = \omega_k, U_{s^-} = \nu_l, \mathbf{y})$ in each $s \in \mathcal{S}$ and for all $(\omega_i, \nu_j, \omega_k, \nu_l) \in (\Omega_x \times \Omega_u)^2$.

An algorithm for computing the MPM in the HMT model can be found in [12]. We adapted this method for the more general case of TMT used here. It consists of the five steps described below, and requires the computation of auxiliary functions denoted A_n for $n \in \{0, 1, \ldots, N\}$ in the following *fine-to-coarse pass*:

1. At the finest resolution, $\forall s \in \mathcal{S}^N$:
$$A_N(s; \omega_i, \nu_j, \omega_k, \nu_l) = p(y_s, \omega_i, \nu_j | y_{s^-}, \omega_k, \nu_l). \tag{3}$$
2. On the intermediate resolutions, $\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^{N-1}\}$:
$$A_n(s; \omega_i, \nu_j, \omega_k, \nu_l) = p(y_s, \omega_i, \nu_j | y_{s^-}, \omega_k, \nu_l) \times \prod_{s^+ \in \mathcal{E}_s} \left(\sum_{(\omega,\nu) \in \Omega_x \times \Omega_u} A_{n+1}(s^+; \omega, \nu, \omega_i, \nu_j) \right); \tag{4}$$
where $\mathcal{E}_s$ is the children set of s.
3. On the root, $r \in \mathcal{S}^0$:
$$A_0(\omega_i, \nu_j) = p(y_0, \omega_i, \nu_j) \times \prod_{s^+ \in \mathcal{E}_0} \left(\sum_{(\omega,\nu) \in \Omega_x \times \Omega_u} A_1(s^+; \omega, \nu, \omega_i, \nu_j) \right); \tag{5}$$
where $\mathcal{E}_0 = \mathcal{S}^1$.

A triplet Markov tree benefits from the posterior Markovianity: $p(\mathbf{X}, \mathbf{U} | \mathbf{Y} = \mathbf{y})$ is a Markov distribution[1]. Hence, it is feasible to compute $p(X_s = \omega_i, U_s = \nu_j | \mathbf{y})$ for $s \in \mathcal{S}^N$ and $(\omega_i, \nu_j) \in \Omega_x \times \Omega_u$ in the following *coarse-to-fine pass*:

4. Computation of the posterior distribution in the root r:
$$p(X_r = \omega_i, U_r = \nu_j | \mathbf{y}) = \frac{A_0(\omega_i, \nu_j)}{\sum_{(\omega,\nu) \in \Omega_x \times \Omega_u} A_0(\omega, \nu)}. \tag{6}$$
5. Computation of the posterior $\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^N\}$:
$$p(X_s = \omega_i, U_s = \nu_j | X_{s^-} = \omega_k, U_{s^-} = \nu_l, \mathbf{y}) = \frac{A_n(s; \omega_i, \nu_j, \omega_k, \nu_l)}{\sum_{(\omega,\nu) \in \Omega_x \times \Omega_u} A_n(s; \omega, \nu, \omega_k, \nu_l)}. \tag{7}$$

Once the cascading posterior density computations are performed, one can obtain the MPM estimation for the root:

$$(\hat{x}_r, \hat{u}_r)^{\mathrm{MPM}} = \operatorname*{arg\,max}_{(\omega_i,\nu_j) \in \Omega_x \times \Omega_u} p(X_r = \omega_i, U_r = \nu_j | \mathbf{y}); \tag{8}$$

and $\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^N\}$:

$$(\hat{x}_s, \hat{u}_s)^{\mathrm{MPM}} = \operatorname*{arg\,max}_{(\omega_i,\nu_j) \in \Omega_x \times \Omega_u} p\left((X_s, U_s) = (\omega_i, \nu_j) | (\hat{x}_{s^-}, \hat{u}_{s^-})^{\mathrm{MPM}}, \mathbf{y} \right); \tag{9}$$

so that the MPM estimation is the set $\left((\hat{x}_s, \hat{u}_s)^{\mathrm{MPM}} \right)_{s \in \mathcal{S}}$.

[1] Let us add that $p(\mathbf{X} | \mathbf{Y} = \mathbf{y})$ is not necessarily a Markov distribution.

3. APPLICATION TO IMAGE SEGMENTATION

3.1. Model specificities

An observed image is represented by $(Y_s)_{s\in\mathcal{S}^N} \in \mathbb{R}^{|\mathcal{S}^N|}$, and we assume that:

$$p(Y_s|X_s, U_s, T_{s^-}) = p(Y_s|X_s). \tag{10}$$

Besides, $\forall s \in \{\mathcal{S}^1, \dots, \mathcal{S}^{N-1}\}$, we formally consider that $p(Y_s|X_s, U_s, T_{s^-}) \propto 1$ and $p(Y_r|X_r, U_r) \propto 1$.

Concerning the auxiliary process $\mathbf{U}$, the considered neighborhood is the set of 8 closest neighbors at the same resolution. Hence U_s is $8-$valued and is noted: $U_s = (U_{s,1}, \dots, U_{s,8})$. Since U_s has the same distribution as $(X_v)_{v\in V_s}$ and since the $(X_v)_{v\in V_s}$ are independent given T_{s-}, one can write:

$$p(U_s|T_{s^-}) = p\left((X_v)_{v\in V_s}|T_{s^-}\right) = \prod_{v\in V_s} p(X_v|T_{s^-}). \tag{11}$$

X_v may belong to the children of T_{s-} or not (see Figure 2). In the first case, we define $p(X_v|T_{s^-})$ along the lines of a Potts model:

$$p(X_s|T_{s^-}) \propto \exp\left(\beta\delta_{X_{s^-}}^{X_s} + \gamma \sum_{1\le k\le 8} \delta_{U_{s^-,k}}^{X_s}\right); \tag{12}$$

where δ_a^b is the Kronecker product, which equals to 1 when $a = b$ and 0 otherwise, and β and γ are model parameters.

Remark. When $\gamma = 0$, the transition probabilities correspond to a classical HMT model.

In the second case, when X_v does not belong to the children of T_{s^-}, we know that $X_{v^-} \in (X_w)_{w\in V_{s^-}}$. Let us assume that $U_{s^-,k}$ models the same site as the parent of v.

Then, we define:

$$p(X_v|T_{s^-}) \propto \exp(\beta\delta_{U_{s^-,k}}^{X_v} + \gamma\delta_{X_{s^-}}^{X_v} + \gamma \sum_{\substack{1\le k'\le 8\\ k'\neq k}} \delta_{U_{s^-,k'}}^{X_v}). \tag{13}$$

Besides, in the root and $\forall(\omega_i, \nu_j) \in \Omega_x \times \Omega_u$, we have $p(T_0 = (y_0, \omega_i, \nu_j)) = \pi_i$, where the π_i are model parameters representing the prior on the tree root.

Remark. In practice, the $8-$neighborhood is not defined on the image border. In the model implementation, borders are replicated to bypass this limit.

In the remaining of this paper, we assume a Gaussian noise model parametrized by μ_i, σ_i and consider the segmentation with $|\Omega_x| = 2$ classes.

3.2. Unsupervised Parameter Estimation

The model parameters are:

$$\theta = \{\mu_0, \mu_1, \sigma_0, \sigma_1, \pi_0, \pi_1, \gamma, \beta\}. \tag{14}$$

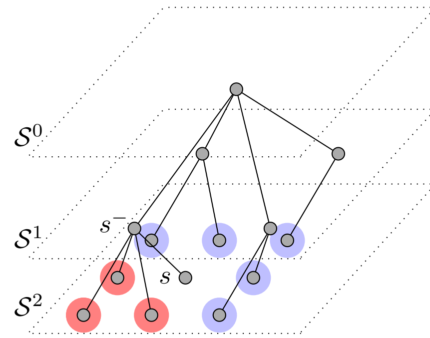


Fig. 2: Restriction of the quadtree dependency graph from Figure 1a to the 8 neighbors of s and their parents. The light red disks indicate the neighbor of X_s which are children of X_{s^-}, the distribution of which is given by (12). The light blue disks represent the other neighbor of X_s, which are ruled by distribution (13).

In an unsupervised framework, θ must be estimated from $\mathbf{Y} = \mathbf{y}$ only. At first, we present the estimators when a "complete" realization $(\mathbf{y}, \mathbf{x}, \mathbf{u})$ is available.

The estimation of μ_i et σ_i is performed with the maximum likelihood estimator using $(\mathbf{x}, \mathbf{y})$. The π_i are directly estimated from the posterior density computation (6) :

$$\begin{aligned} \hat{\pi}_i &= p(X_r = \omega_k|\mathbf{Y} = \mathbf{y}) \\ &= \sum_{\nu\in\Omega_u} p(X_r = \omega_k, U_r = \nu|\mathbf{Y} = \mathbf{y}). \end{aligned} \tag{15}$$

We now detail the estimators for β and γ, inspired by the least-squares estimator from [13], initially proposed for Markov field distributions. We have $\forall s \in \{\mathcal{S}^1, \dots, \mathcal{S}^N\}$ and for $\omega_i \neq \omega_j$:

$$\frac{p(X_s = \omega_i|T_{s^-})}{p(X_s = \omega_j|T_{s^-})} = \exp\left[\beta\left(\delta_{X_{s^-}}^{\omega_i} - \delta_{X_{s^-}}^{\omega_j}\right) + \gamma \sum_{1\le k\le 8} (\delta_{U_{s^-,k}}^{\omega_i} - \delta_{U_{s^-,k}}^{\omega_j})\right]. \tag{16}$$

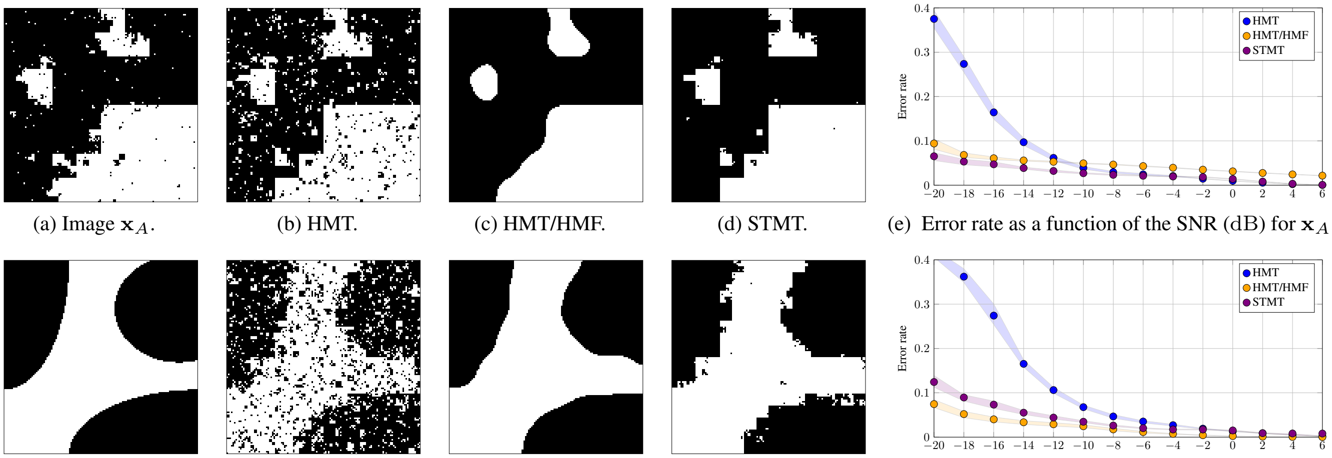
The left-hand term of this equation is computed thanks to the posterior distribution (7)[2]. We obtain the "partial" estimation of β for all pairs (x_s, t_{s^-}):

$$\hat{\beta}_{s,s^-} = \frac{\log\left[\frac{p(x_s=\omega_i|t_{s^-},\mathbf{Y}=\mathbf{y})}{p(x_s=\omega_j|t_{s^-},\mathbf{Y}=\mathbf{y})}\right] - \gamma \sum_{1\le k\le 8} (\delta_{u_{s^-,k}}^{\omega_i} - \delta_{u_{s^-,k}}^{\omega_j})}{\delta_{x_{s^-}}^{\omega_i} - \delta_{x_{s^-}}^{\omega_j}}. \tag{17}$$

Hence the least-squares estimation of β is:

$$\hat{\beta} = \frac{1}{|\mathcal{S} \setminus \mathcal{S}^0|} \sum_{s\in\mathcal{S}\setminus\mathcal{S}^0} \hat{\beta}_{s,s^-}. \tag{18}$$

[2] In practice, this estimator is more robust than the histogram estimation.



(a) Image $\mathbf{x}_A$. (b) HMT. (c) HMT/HMF. (d) STMT. (e) Error rate as a function of the SNR (dB) for $\mathbf{x}_A$.

(f) Image $\mathbf{x}_B$. (g) HMT. (h) HMT/HMF. (i) STMT. (j) Error rate as a function of the SNR (dB) for $\mathbf{x}_B$.

Fig. 3: Numerical results. Left column: original images. Columns 2, 3 and 4: results instances with $\mathrm{SNR} = -14$ dB. The fifth column represents the average results with the first and third quartile variations.

Based on the same reasoning, we obtain with (7) and for all pairs (x_s, t_{s^-}):

$$\hat{\gamma}_{s,s^-} = \frac{\log\left[\frac{p(x_s=\omega_i|t_{s^-},\mathbf{Y}=\mathbf{y})}{p(x_s=\omega_j|t_{s^-},\mathbf{Y}=\mathbf{y})}\right] - \beta\left(\delta^{\omega_i}_{x_{s^-}} - \delta^{\omega_j}_{x_{s^-}}\right)}{\sum\limits_{1\leq k\leq 8}(\delta^{\omega_i}_{u_{s^-,k}} - \delta^{\omega_j}_{u_{s^-,k}})}. \tag{19}$$

When $\sum_{1\leq k\leq 8}(\delta^{\omega_i}_{u_{s^-,k}} - \delta^{\omega_j}_{u_{s^-,k}}) = 0$, this "partial" estimation is not defined. Denoting $C(s^-, i, j)$ this value, the least-squares estimation of γ is:

$$\hat{\gamma} = \frac{\sum\limits_{s\in\mathcal{S}\setminus\mathcal{S}^0} \mathbb{1}_{\{C(s^-,i,j)\neq 0\}}\hat{\gamma}_{s,s^-}}{\sum\limits_{s\in\mathcal{S}\setminus\mathcal{S}^0} \mathbb{1}_{\{C(s^-,i,j)\neq 0\}}}. \tag{20}$$

The parameters θ are estimated with the Stochastic Expectation-Maximization (SEM) method [14]. This stochastic variant of EM [15] iteratively generates realizations of $(\mathbf{X}, \mathbf{U})$ thanks to the previous value of θ, and re-estimate θ thanks to the simulated realizations with the estimators presented in this section.

4. NUMERICAL RESULTS

We use two synthetic images of size 128^2 pixels, denoted $\mathbf{x}_A$ and $\mathbf{x}_B$. The former is a realization of a STMT using $\beta = 2.25$ et $\gamma = 0.25$ (Figure 3a), and the latter present wide homogeneous regions with smooth boundaries (Figure 3f). Besides, we set $\mu_0 = 0$, $\mu_1 = 1$ and $\sigma_0 = \sigma_1 = \sigma$ so that σ tunes the Signal-to-Noise Ratio (SNR) defined as $20\log_{10}(\overline{\mathbf{x}}/\sigma)$, $\overline{\mathbf{x}}$ being the averaged value of $\mathbf{x}$.In this setting, we evaluate three methods:

- the HMT-based segmentation with the MPM criterion;
- a "mixed" HMT/HMF method: the segmentation is made within a classical HMF model, whose Gibbs sampling are initialized using the HMT result. This model is expected to handle well both spatial and hierarchical features in images;
- the STMT-based segmentation introduced in this paper.

Figures 3b-3d and 3g-3i illustrates the segmentation results, and the error rate evaluated on 100 simulations are displayed in Figures 3e and 3j.

These results first show that, in any cases, the HMT-based segmentation is outperformed by its alternatives. Besides, the HMT/HMF and STMT model yields close error rates, and the achievement of the best average error rate depends on the image to process. This can be visually interpreted as a consequence of the "smoothness" of the image $\mathbf{x}_B$, which favors a Markov field-based method. As a final remark, let us add that our numerical experiments showed that the computation of the STMT-based segmentation method is always faster than the HMT/HMF-based segmentation (up to 10 times faster on 512^2 images) due to the ability to compute exactly the posterior densities.

5. CONCLUSION

This paper introduced the STMT model, as well as the computation of the MPM criterion and the segmentation of images. Results showed that the model is robust and competitive with a mixed HMT/HMF model, while providing the ability to compute exactly the posterior densities.

6. REFERENCES

[1] Nathalie Giordana and Wojciech Pieczynski, "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 465–475, 1997.

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

[3] J.-M. Laferté, P. Pérez, and F. Heitz, "Discrete Markov image modeling and inference on the quadtree," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 390–404, 2000.

[4] J-N Provost, Christophe Collet, Philippe Rostaing, Patrick Pérez, and Patrick Bouthemy, "Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps," *Computer Vision and Image Understanding*, vol. 93, no. 2, pp. 155–174, 2004.

[5] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Am. Stat. Assoc.*, vol. 82, no. 397, pp. 76–89, 1987.

[6] E. Monfrini and W. Pieczynski, "Estimation de mélanges généralisés dans les arbres de Markov cachés, application à la segmentation des images de cartons d'orgue de barbarie.," *Traitement du Signal*, vol. 22, no. 2, 2005.

[7] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, 2000.

[8] D. Benboudjema and W. Pieczynski, "Unsupervised statistical segmentation of nonstationary images using triplet Markov fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1367–1378, 2007.

[9] W. Pieczynski, "Arbres de Markov triplet et fusion de Dempster–Shafer," *Comptes Rendus Mathématique*, vol. 336, no. 10, pp. 869–872, 2003.

[10] W. Pieczynski, C. Hulard, and T. Veit, "Triplet Markov chains in hidden signal restoration," in *International Symposium on Remote Sensing*. International Society for Optics and Photonics, 2003, pp. 58–68.

[11] J.-B Courbot, E. Monfrini, V. Mazet, and C. Collet, "Oriented triplet Markov fields," *to appear in Pattern Recognition Letters*, 2018.

[12] E. Monfrini, T. Ledru, E. Vaie, and W. Pieczynski, "Segmentation non supervisée d'images par arbres de Markov cachés," in *17ème Colloque GRETSI*, 1999.

[13] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, , no. 1, pp. 39–55, 1987.

[14] G. Celeux and J. Diebolt, "L'algorithme SEM: un algorithme d'apprentissage probabiliste pour la reconnaissance de mélange de densités," *Revue de statistique appliquée*, vol. 34, no. 2, pp. 35–52, 1986.

[15] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.. Series B*, pp. 1–38, 1977.