

Pairwise and Hidden Markov Random Fields in Image Segmentation

Jean-Baptiste Courbot
 IRIMAS UR 7499, Université de Haute Alsace,
 Mulhouse, France
 Email: jean-baptiste.courbot@uha.fr

Vincent Mazet
 ICube UMR 7357, Université de Strasbourg – CNRS,
 Illkirch, France
 Email: vincent.mazet@unistra.fr

Abstract—The purpose of this paper is to identify the similarities and differences between two image restoration approaches based on Markov field modeling. The first one is the well-known Bayesian approach which models the unknowns with a Markovian prior. In the second approach, as proposed by Pieczynski and Tebbache [1], the pair unknowns–observations as a whole is considered Markovian. The two approaches are compared based on their posterior distribution, synthetic results and real examples, when applied to the segmentation of degraded images.

Index Terms—Bayesian Image Restoration, Pairwise Markov Random Fields, Hidden Markov Random Fields

I. INTRODUCTION

Hidden Markov random fields (HMRF) are a widespread tool for image segmentation introduced in the 1980’s [2], [3]. This approach models the observation \mathbf{y} by a likelihood $p(\mathbf{y}|z)$ where z are the (hidden) labels, and the prior $p(z)$ of the latter is a Markov random field so as to model spatial interactions between pixels. HMRF have been successfully applied in many works [2]–[6].

Besides, Pairwise Markov random fields (PMRF) have been introduced by Pieczynski and Tebbache in 2000 [1], [7]. The PMRF model considers the Markovianity on the couple (z, \mathbf{y}) instead of the labels z only as in the HMRF model. Neither \mathbf{y} nor z are necessarily Markovian, but $z|\mathbf{y}$ and $\mathbf{y}|z$ are Markovian. PMRF models have been introduced to avoid approximations when estimating the labels from the observations in textured images.

We and colleagues have often employed the PMRF model in the past, see e.g. [8]–[10]. However the PMRF model has been rarely used in the community, contrary to the HMRF model. Nevertheless, both share strong similarities: this is quite intriguing and invites comparison of the two models. To the best of our knowledge, no comparison has been made in the literature, even in [1]. Therefore, our goal in this paper is to investigate the links and differences between the two aforementioned models, so as to identify the limits and advantages of the models.

The problem is formalized in section II. The two approaches are mathematically detailed in terms of their posterior distribution in section III and their results are compared on synthetic and real images in section IV. A discussion is finally given in section V.

II. SCOPE OF THE STUDY

We consider the problem of joint deconvolution–segmentation [4], [11]–[13], in which a blurry observation $\mathbf{y} \in \mathbb{R}^{S \times B}$ with S pixels and B bands has to be segmented in K classes. Introducing a blur within the segmentation task allows to consider more general problems.

The goal is to estimate the label image $z \in \mathbb{K}^{S \times B}$ that gathers the labels of each pixel in the observation, where $\mathbb{K} = \{1, \dots, K\}$ denotes the set of classes. The observation \mathbf{y} is modeled as a blurry and noisy observation of an unknown image $\mathbf{x} \in \mathbb{R}^{S \times B}$:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e} \quad (1)$$

where $\mathbf{H} \in \mathbb{R}^{S \times S}$ and $\mathbf{e} \in \mathbb{R}^{S \times B}$. \mathbf{H} is the Toeplitz matrix of the point spread function (PSF), which introduces the same spatial blur for each band, without blur between bands. The PSF is supposed to be known and symmetric, so \mathbf{H} is a known and symmetric matrix. The special case where \mathbf{H} is the identity matrix yields the classical segmentation problem. \mathbf{e} is an additive white Gaussian noise with zero mean and known variance σ^2 .

In addition to this, the unknown image \mathbf{x} is directly related to the label image z such that each pixel x_s with class z_s equals a signal $w_k \in \mathbb{R}^B$ where $k = z_s$:

$$\forall s, \quad x_s = w_k \quad \text{if and only if} \quad z_s = k. \quad (2)$$

The classes are considered to follow a Potts model [14] so as to favor the grouping of pixels of the same class:

$$p(z) \propto \exp \left(\beta \sum_{(s,t) \in \mathcal{C}} \delta(z_s, z_t) \right) \quad (3)$$

where β is an hyperparameter, a.k.a. the “granularity” coefficient and \mathcal{C} is the set of neighboring pair of sites in the images (or set of cliques).

III. DESCRIPTION OF THE TWO APPROACHES

The HMRF approach usually works with the posterior $p(z|\mathbf{y})$ whereas the PMRF approach describes the joint distribution $p(z_s, \mathbf{y}_s | z_{-s}, \mathbf{y}_{-s})$ in one site s (the notation $-s$ means the whole set of sites but s). To compare the two approaches we choose to write their conditional posteriors $p(z_s, \mathbf{y}_s | z_{-s}, \mathbf{y}_{-s})$.

A. HMRF Approach

The likelihood results from the model of an additive white Gaussian noise:

$$p(\mathbf{y}|\mathbf{z}) = \frac{1}{(2\pi\sigma^2)^{(SB)/2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_F^2}{2\sigma^2}\right), \quad (4)$$

where $\mathbf{x}_s = \mathbf{w}_k$ if and only if $z_s = k$, and $\|\cdot\|_F$ denotes the Frobenius norm.

Then, the joint distribution is:

$$p(\mathbf{z}, \mathbf{y}) \propto \exp\left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_F^2}{2\sigma^2} + \beta \sum_{(s,t) \in \mathcal{C}} \delta(z_s, z_t)\right). \quad (5)$$

To get the conditional posterior of (z_s, \mathbf{y}_s) , Eq. (5) has to be written by distinguishing the terms with z_s or \mathbf{y}_s from the others (which will be denoted ‘‘ct’’). So,

$$\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_F^2 = \|\mathbf{y}_s - (\mathbf{H}\mathbf{x})_s\|^2 + \sum_{i \neq s} \|\mathbf{y}_i - (\mathbf{H}\mathbf{x})_i\|^2 \quad (6)$$

where $\|\cdot\|$ denotes the Euclidean norm. Note that the vectors are row vectors since they correspond to a site in the images. For example, $(\mathbf{H}\mathbf{x})_i$ is the i th row of matrix $\mathbf{H}\mathbf{x}$. Because of the PSF \mathbf{H} , the term in the sum in Eq. (6) may depend of z_s (through \mathbf{x}). For all $i \neq s$, this terms expands as:

$$\|\mathbf{y}_i - (\mathbf{H}\mathbf{x})_i\|^2 = \|\mathbf{y}_i\|^2 - 2\mathbf{y}_i(\mathbf{H}\mathbf{x})_i^T + \|(\mathbf{H}\mathbf{x})_i\|^2 \quad (7)$$

where:

$$\begin{aligned} \|\mathbf{y}_i\|^2 &= \text{ct} \\ \mathbf{y}_i(\mathbf{H}\mathbf{x})_i^T &= \sum_b y_{ib}(\mathbf{H}\mathbf{x})_{ib} = \sum_b y_{ib}H_{is}x_{sb} + \text{ct} \\ \|(\mathbf{H}\mathbf{x})_i\|^2 &= \sum_b ((\mathbf{H}\mathbf{x})_{ib})^2 = \sum_b \left(\sum_n H_{in}x_{nb}\right)^2 \\ &= \sum_b \left(H_{is}x_{sb} + \sum_{n \neq s} H_{in}x_{nb}\right)^2 \\ &= \sum_b \left((H_{is}x_{sb})^2 + 2H_{is}x_{sb} \sum_{n \neq s} H_{in}x_{nb}\right) + \text{ct} \end{aligned}$$

So, Eq. (7) yields:

$$\begin{aligned} \|\mathbf{y}_i - (\mathbf{H}\mathbf{x})_i\|^2 &= \sum_b \left(-2y_{ib}H_{is}x_{sb} \right. \\ &\quad \left. + (H_{is}x_{sb})^2 + 2H_{is}x_{sb} \sum_{n \neq s} H_{in}x_{nb}\right) + \text{ct}. \end{aligned} \quad (8)$$

In addition, the prior on \mathbf{z} is defined in Eq. (3), and:

$$\sum_{(s,t) \in \mathcal{C}} \delta(z_s, z_t) = \sum_{t \in N_s} \delta(z_s, z_t) + \text{ct} \quad (9)$$

where N_s is the neighborhood of s , to be defined by the user. Putting Eqs. (6), (8) and (9) in Eq. (5), one can show that the

conditional posterior of the HMRF approach is the product of two Gaussians and a Potts distribution:

$$\begin{aligned} p(z_s, \mathbf{y}_s | \mathbf{z}_{-s}, \mathbf{y}_{-s}) &\propto \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y}_s - (\mathbf{H}\mathbf{x})_s\|^2\right) \\ &\times \exp\left(-\frac{1}{2\rho_s} \|\mathbf{x}_s - \mathbf{m}_s\|^2\right) \exp\left(\beta \sum_{t \in N_s} \delta(z_s, z_t)\right) \end{aligned} \quad (10)$$

where $\mathbf{m}_s \in \mathbb{R}^{S \times B}$ and $\rho_s \in \mathbb{R}$ such that

$$\begin{aligned} \mathbf{m}_s &= \frac{1}{\sum_{i \neq s} H_{is}^2} \sum_{i \neq s} \left(\mathbf{y}_i H_{is} - \sum_{n \neq s} H_{is} H_{in} \mathbf{x}_n\right), \\ \rho_s &= \frac{\sigma^2}{\sum_{i \neq s} H_{is}^2}. \end{aligned}$$

B. PMRF Approach

With the PRMF approach, the pair (\mathbf{z}, \mathbf{y}) is assumed Markovian with respect to a given neighborhood [1]. The assumption of white noise implies that the variables in \mathbf{y} are independent given \mathbf{z} , so the joint distribution in a site s can be written:

$$\begin{aligned} p(z_s, \mathbf{y}_s | \mathbf{z}_{-s}, \mathbf{y}_{-s}) &= p(\mathbf{y}_s | z_s, \mathbf{z}_{-s}, \mathbf{y}_{-s}) p(z_s | \mathbf{z}_{-s}, \mathbf{y}_{-s}) \\ &= p(\mathbf{y}_s | z_s, \mathbf{z}_{-s}) p(z_s | \mathbf{z}_{-s}, \mathbf{y}_{-s}) \end{aligned} \quad (11)$$

Assuming a white Gaussian noise,

$$p(\mathbf{y}_s | z_s, \mathbf{z}_{-s}) \propto \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y}_s - (\mathbf{H}\mathbf{x})_s\|^2\right). \quad (12)$$

We assume, furthermore, that z_s and \mathbf{z}_{-s} are independent given \mathbf{y}_{-s} , and that \mathbf{z} is Markovian. So, the second term of (11) is written as:

$$p(z_s | \mathbf{z}_{-s}, \mathbf{y}_{-s}) = p(z_s | \mathbf{z}_{N_s}). \quad (13)$$

Following the Potts model (3), we obtain that

$$p(z_s | \mathbf{z}_{N_s}) \propto \exp\left(\beta \sum_{t \in N_s} \delta(z_s, z_t)\right) \quad (14)$$

Using Eqs. (12) and (14) in (11), we have:

$$\begin{aligned} p(z_s, \mathbf{y}_s | \mathbf{z}_{-s}, \mathbf{y}_{-s}) &\propto \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y}_s - (\mathbf{H}\mathbf{x})_s\|^2\right) \\ &\times \exp\left(\beta \sum_{t \in N_s} \delta(z_s, z_t)\right) \end{aligned} \quad (15)$$

Note that there are two neighborhood systems embedding the Markovianity of the pair (\mathbf{z}, \mathbf{y}) : in the first term, the considered neighborhood covers the whole lattice, while in the second it covers the closest neighbors of s , denoted N_s . This choice ensures that the comparison between PMRF and HMRF does not depend on a particular neighboring system.

Let us also remark that since pairwise Markov fields are a general class of models, specific assumptions have been made to tailor it to the considered problem. Hence, it should be noted that Eq. (15) describes a *particular* PMRF model.

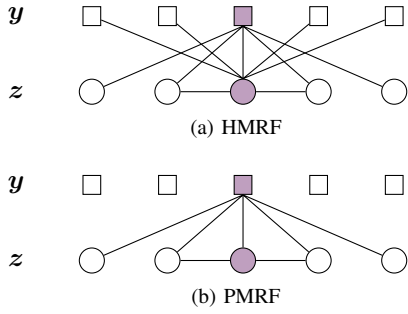


Fig. 1. Graphical representation of the statistical dependencies embedded by the two models, conditioned on a site s (colored). No link between two variables means that they are independent given the variables in the s site. For clarity, only one dimension of the lattice is depicted. The neighborhood considered for the Markov process contains only the two immediate neighbors, while the PSF covers 5 sites.

C. Comparison of the two models

The posterior distributions in a site for the two approaches are given in Eqs. (10) and (15). Identifying the three exponential terms in (10) as proportional to three densities f , g and h let us rewrite (10) as

$$p(z_s, \mathbf{y}_s | \mathbf{z}_{-s}, \mathbf{y}_{-s}) \propto f(\mathbf{y}_s, z_s, \mathbf{z}_{-s}) g(\mathbf{y}_{-s}, z_s, \mathbf{z}_{-s}) h(z_s, \mathbf{z}_{-s}) \quad (16)$$

and (15) as

$$p(z_s, \mathbf{y}_s | \mathbf{z}_{-s}, \mathbf{y}_{-s}) \propto f(\mathbf{y}_s, z_s, \mathbf{z}_{-s}) h(z_s, \mathbf{z}_{-s}). \quad (17)$$

Then, the only difference between the two models lies in the density g , which adds, with respect to Eq. (15), a novel dependency between \mathbf{y}_{-s} and z . The graphical representations of the dependencies, conditioned on a site s , are given in Figure 1.

Noteworthy, in the PMRF approach, the relation between two variables from z and \mathbf{y} depends on the conditioning site. Indeed, let s and t be two sites in the image: if we condition by the s site, the dependency between z_s and \mathbf{y}_t is embedded in the g function, whereas if t is the conditioning site, the dependency is contained in the f function.

Let us also remark that the dependency embedded in the g density is also related to the assumption, for PMRF, that z_{-s} and z_s are independent given \mathbf{y}_{-s} .

Finally, we can see that if the blur is missing, *i.e.* \mathbf{H} is the identity matrix, then $(\mathbf{y}_i - (\mathbf{H}\mathbf{x})_i)^2 = (\mathbf{y}_i - \mathbf{x}_i)^2$ so the corresponding term in (7) does not depend on s . In this case $g \propto 1$ and the two approaches are equivalent.

IV. NUMERICAL RESULTS

The theoretical difference between the two approaches highlighted in the previous section is hard to interpret in practice. So we present in this section a numerical comparison of the two methods for the retrieval of noisy, blurred images. Using the posterior distributions (10) and (15), the segmentation is obtained by estimating the Maximum A Posteriori (MAP) [3]:

$$\hat{\mathbf{z}}^{\text{MAP}} = \arg \max_{\mathbf{z} \in \mathbb{K}^S} p(\mathbf{z} | \mathbf{y}). \quad (18)$$

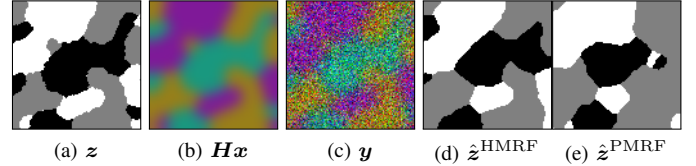


Fig. 2. Example of synthetic image with its MAP estimation using the HMRF model and the PMRF model. The parameters are: $\sigma = 0.2$, $\sigma_{\text{PSF}} = 5$ and $r = 1$ (cf. dotted line in Fig. 3).

The MAP is estimated by using a Gibbs sampler (50 iterations) with a simulated annealing, and obtained in a supervised fashion: the model parameters are assumed to be known.

We first consider the estimation of z from synthetic images \mathbf{y} with $S = 100 \times 100$ pixels, for a segmentation in 3 classes using $B = 3$ colors¹. The Potts parameter β , which is not involved in the differences between the two approaches, is fixed to $\beta = 0.6$ in all experiments.

We consider a Gaussian PSF blur, with its standard deviation σ_{PSF} . Since most of its energy is contained in its center, it is usual to truncate the PSF to its central values. We represent this truncation as a ratio r between the truncation radius and σ_{PSF} , such that $r = 0.5$ means that 48.5% of the PSF energy is kept, and $r = 3$ means 99.9% of the PSF energy is kept. Summing up, the parameters considered in the numerical comparison are σ , σ_{PSF} , and r , and we evaluate error rate between the ground truth z and its MAP estimates $\hat{\mathbf{z}}^{\text{HMRF}}$ and $\hat{\mathbf{z}}^{\text{PMRF}}$.

An example of synthetic image, together with its segmentations, is depicted in Fig. 2, and the complete results are given in Fig. 3. Several comments can be made regarding these results:

- The two models behave as expected: when the image is harder to process (higher σ or σ_{PSF}) the results worsen and are more scattered, while accounting for more PSF coefficient yields better results.
- In average the HMRF model yields better results than the PMRF model. Hence, the additional term g in Eq. 16 does play an efficient part in the segmentation.
- Unexpectedly, the main differences between the models are driven by changes in σ , as seen in Fig. 3. On the contrary, for a given σ , varying σ_{PSF} or r does not change the performance gap between the methods, as depicted in Figs. 3b and 3c. In consequence, the additional term is weakly influenced by the PSF parameters.
- When either σ or σ_{PSF} are low, the two models tend to provide similar results. This was theoretically expected for σ_{PSF} but it is also the case for σ .

To sum up, the difference between the two approaches is visible at low signal-to-noise ratio: clearly, the HMRF model yields better results than the PMRF model at high noise level.

We also tested the two models on real images extracted from the McGill calibrated colour image database [15] (<http://tabby.vision.mcgill.ca/>) in the “flower” category. That these 100×100 -pixel images are blurred with a Gaussian PSF ($\sigma_{\text{PSF}} = 5$),

¹Note that an extension to $B > 3$ colors could be considered to handle multi- and hyper-spectral images, together with cross-band noise correlations.

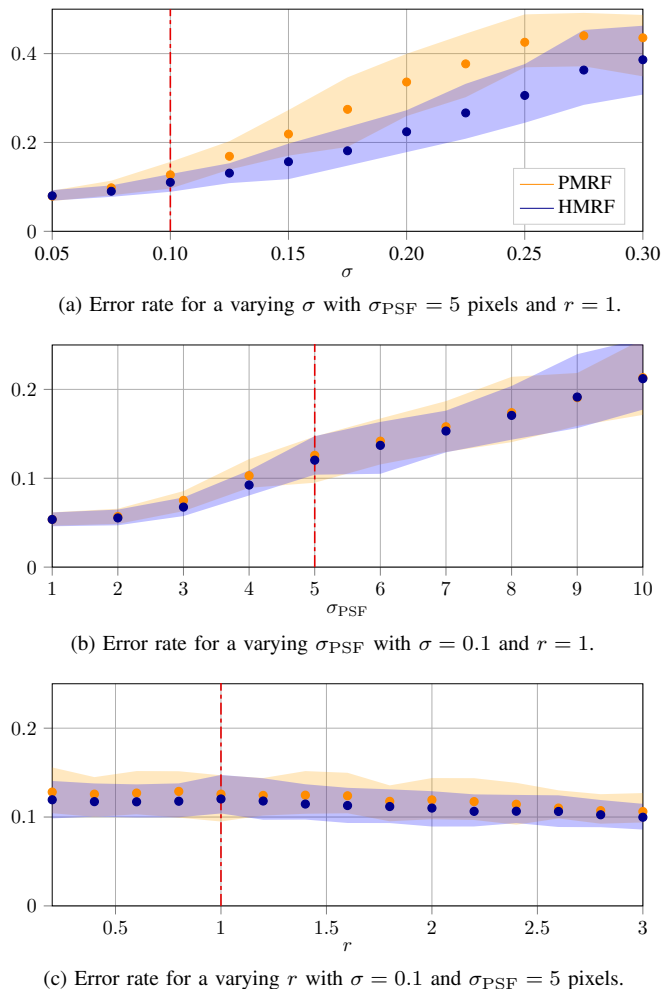


Fig. 3. Numerical comparison of the two models studied in this paper. The dots represent the average error rates over 30 trials, and the shaded contours depict the 1st and 9th deciles of the results. For each studied parameter, the two others are fixed, so that the dashed vertical line is common between all graphs.

add Gaussian white noise ($\sigma = 0.05$), and consider $r = 1$ and $\beta = 0.6$. The means in $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ defined in Eq. (2) were obtained by using the Kmeans algorithm [16] on $\mathbf{H}\mathbf{x}$. The results on five real images are depicted in Fig. 4:

- On the most homogeneous images (#1 and #2), there is no major differences between the two models. Hence, most of the classification is carried out in the f and h densities common to the two models.
- Clear differences appear on images #3, #4, and #5: the HMRF model follows more closely the observed \mathbf{y} than the PMRF model, and the latter tolerates gaps from \mathbf{y} that enforces spatial regularity.
- The parameter set is the same for the two methods, but may not be optimal, as we do not know the “true” parameters of the images. In addition, the optimal parameters may differ between the PMRF and HMRF models. Unsupervised parameter estimation should handle this point, but is beyond the scope of this paper.

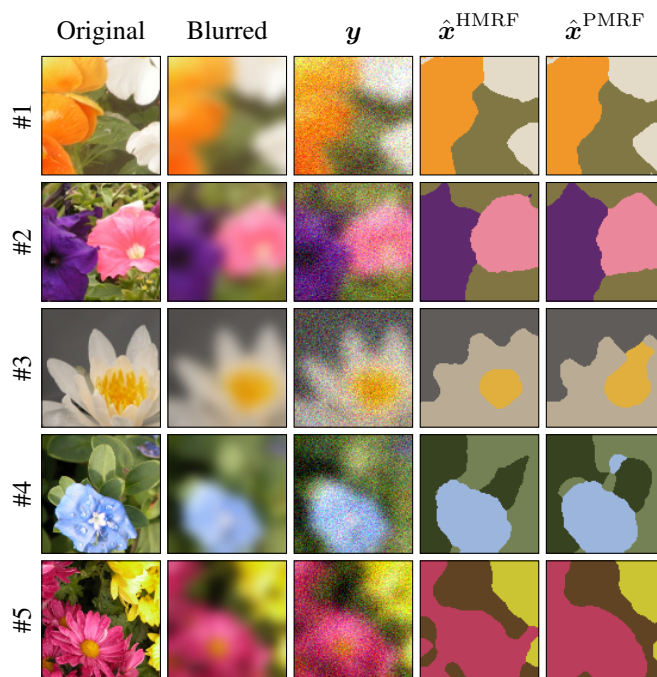


Fig. 4. Example of real image with the 3-class MAP estimations using the HMRF model and the PMRF models. The results are depicted for \mathbf{x} in order to indicate the mean color vectors used for each class. The parameters are $\beta = 0.6$, $\sigma = 0.1$, $\sigma_{\text{PSF}} = 5$ and $r = 1$.

V. DISCUSSION

In the context of image segmentation, we have compared the Markovian approach commonly used in the Bayesian framework with an instance of pairwise Markov field proposed by Pieczynski and Tebbache [1].

We highlighted that these approaches differ by a single term in their site-wise posterior distribution. This term is written as a Gaussian on \mathbf{x}_s , whose mean and variance depend through the PSF on the observations and classes in sites other than s . While it is difficult to grasp the theoretical impact of this term, numerical simulations show that it does affect the results, in favor of the HMRF model. Surprisingly, this clear difference is driven by the noise level instead of the PSF, although the main difference between the two models is a term modeling the spatial dependence through the PSF. The next step of this research will be dedicated to the interpretation and understanding of this observation.

Further works on this topic may compare further these models in the context of unsupervised parameter estimation, with non-stationary processes, or in the image restoration problem for instance.

REFERENCES

- [1] W. Pieczynski and A.-N. Tebbache, “Pairwise Markov random fields and segmentation of textured images,” *Machine Graphics and Vision*, vol. 9, no. 3, pp. 705–718, 2000.
- [2] J. Besag, “On the statistical analysis of dirty pictures,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 48, no. 3, pp. 259–279, 1986.

- [3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [4] H. Ayasso and A. Mohammad-Djafari, "Joint NDT image restoration and segmentation using Gauss–Markov–Potts prior models and variational Bayesian computation," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2265–2277, 2010.
- [5] C. Kevrann and F. Heitz, "A Markov random field model-based approach to unsupervised texture segmentation using local and global statistics," *IEEE Transactions on Image Processing*, vol. 4, no. 6, pp. 856–862, 1995.
- [6] S. Li, *Markov Random Field Modeling in Image Analysis*. Springer-Verlag, 2001.
- [7] W. Pieczynski and A.-N. Tebbache, "Pairwise Markov random fields and its application in textured images segmentation," in *4th IEEE Southwest Symposium on Image Analysis and Interpretation*, 2000.
- [8] D. Benboudjema and W. Pieczynski, "Parameter estimation in pairwise Markov fields," *Advanced Concepts for Intelligent Vision Systems (ACVIS 04)*, Aug, 2004.
- [9] M. E. Y. Boudaren, L. An, and W. Pieczynski, "Dempster–Shafer fusion of evidential pairwise Markov fields," *International Journal of Approximate Reasoning*, vol. 74, pp. 13–29, 2016.
- [10] J.-B. Courbot, V. Mazet, E. Monfrini, and C. Collet, "Pairwise Markov fields for segmentation in astronomical hyperspectral images," *Signal Processing*, vol. 163, pp. 41–48, 2019.
- [11] D. Kim and M. Comer, "Joint deconvolution/segmentation of microscope images of materials," in *IEEE Statistical Signal Processing Workshop*, 2012.
- [12] C. Vacar and J.-F. Giovannelli, "Unsupervised joint deconvolution and segmentation method for textured images: a Bayesian approach and an advanced sampling algorithm," *EURASIP Journal on Advances in Signal Processing*, 2019.
- [13] N. Zhao, A. Basarab, D. Kouamé, and J.-Y. Tourneret, "Joint segmentation and deconvolution of ultrasound images using a hierarchical Bayesian model based on generalized Gaussian priors," *IEEE Transactions on Image Processing*, 2016.
- [14] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Estimating the granularity coefficient of a Potts–Markov random field within a Markov chain Monte Carlo algorithm," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2385–2397, 2013.
- [15] A. Olmos and F. A. Kingdom, "A biologically inspired algorithm for the recovery of shading and reflectance images," *Perception*, vol. 33, no. 12, pp. 1463–1473, 2004.
- [16] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14, 1967, pp. 281–297.