



**HAL**  
open science

## Bayesian Texture Classification From Indirect Observations Using Fast Sampling

Cornelia Vacar, Jean-François Giovannelli, Yannick Berthoumieu

► **To cite this version:**

Cornelia Vacar, Jean-François Giovannelli, Yannick Berthoumieu. Bayesian Texture Classification From Indirect Observations Using Fast Sampling. *IEEE Transactions on Signal Processing*, 2016, 64 (1), pp.146 - 159. 10.1109/TSP.2015.2480040 . hal-01719190

**HAL Id: hal-01719190**

**<https://hal.science/hal-01719190>**

Submitted on 28 Feb 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Bayesian Texture Classification From Indirect Observations Using Fast Sampling

Cornelia Vacar, Jean-François Giovannelli, and Yannick Berthoumieu

**Abstract**—A Bayesian method for texture model choice from blurred and noisy (i.e., indirect) observations is presented. The textures are modeled by stationary Random Fields, with various distribution laws, either Gaussian or Scale Mixtures of Gaussians. The power spectral densities of the fields are modeled by parametric functions and the aim is to select the most appropriate model among a set of candidates. This is achieved by computing the *a posteriori* model probabilities through parameter marginalization. The marginalization is done by sampling and harmonic mean approach, considering separately each model, in a within-model sampling strategy. The highly nonlinear dependency with respect to the parameters imposes the use of the Metropolis-Hastings sampler. Moreover, to achieve efficient sampling, the paper proposes a new fast algorithm based on the Fisher information matrix, the Fisher Metropolis-Hastings.

**Index Terms**—Bayes, harmonic mean, Metropolis-Hastings, model choice, texture modeling and analysis.

## I. TEXTURE AND MODEL CHOICE

THE central problem in this work is selecting a texture model from a blurred and noisy image. It is solved in a Bayesian framework, where the posterior probability for each model is determined from the model evidences. The classification method is optimal from the risk point of view, however, these evidences are intractable and thus are numerically computed by Monte Carlo Markov Chain (MCMC) methods.

Texture represents a central aspect in image processing. A desirable feature of a texture model is to provide a large capability of description, while being relatively easy to handle and numerically efficient. The literature devoted to texture modeling is fairly rich, covering various classes of models that can be analyzed from different points of view. The representation domain can be spatial [1], [2], Fourier [3], [4], wavelet [5], [6] and the model can be deterministic (pseudo-periodic, based on structural elements), or probabilistic (stochastic-like), as for instance the random field-based model [4], [7]–[11].

In this work, the textured images are modeled by Random Fields (RF), with different parametric forms for the pixel inter-

action and probability distributions. This type of model is very tractable, equally adapted to synthesis and analysis tasks and known to exhibit good performances for stochastic-like texture processing. Nevertheless, it can also be used in the case of more deterministic or pseudo-periodic textures, through specific pixel interactions. From the texture modeling point of view, the contribution of this paper is to introduce a new non-Gaussian model, fully adapted to analysis and synthesis tasks. The particularities of this model are the following:

- Independent Fourier coefficients for the textured image. These coefficients are marginally non-Gaussian, but conditionally Gaussian, to ensure model tractability. This is achieved using Scale Mixtures of GRFs (SMGRF).
- Formulation in the Fourier domain in order to exploit the multiplicative form of the blur in the frequency domain.

Model choice has applications in a wide range of fields, for instance microbiology, proteomics, genomics, economics, statistics, signal and image analysis. The Generalized Maximum Likelihood Estimator (GMLE) method for model selection consists in determining the model with the highest likelihood for the MLEs of the parameters. However, this estimator is not applicable for indirect observations or the non-Gaussian texture model. Among the most frequently used model selection methods are the well-known Akaike Information Criterion (AIC) [12], Deviance Information Criterion (DIC) [13] and the Bayesian Information Criterion (BIC) [14]. Nevertheless, all these criteria are based on two levels of approximations. First, a second order Taylor approximation of the likelihood around its maximum is done in order to derive the form of the criteria. On a second level, except for very simple cases, the maximum likelihood value cannot be computed analytically and, thus, numerical methods (such as optimization) are employed for its computation. In this context, the evidence based methods represent an alternative Bayesian approach [15] that does not rely on an approximation at the first level. Nevertheless, since the evidences are intractable, numerical methods are employed for their computation, this corresponding to an approximation on the aforementioned second level, the numerical one. The evidence-based classification can take the form of Bayes factors or the posterior model probabilities computation method presented in this work.

The latter relies on the formulation of an optimal decision function from the risk viewpoint. Practically, it consists in computing the *a posteriori* model probabilities, based on (i) the joint law for the model, the data, the model parameters and on (ii) computing the evidence for each model. This computation is done by marginalizing the parameters from the joint law. Since

Manuscript received April 30, 2014; revised August 27, 2014, June 20, 2015, and August 26, 2015; accepted August 26, 2015. Date of publication September 17, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Animashree Anandkumar.

The authors are with IMS (Univ. Bordeaux, CNRS, B-INP), UMR 5218, F-33400 Talence, France.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2015.2480040

this law is complicated, the integral is intractable and it is approximated either using MCMC sampling, or the Laplace approximation [16].

In the context of evidence computation using MCMC, some of the parameters have very complicated laws and their sampling can be costly, thus requiring the use of efficient solutions. The recent literature regarding efficient sampling is mainly focused on two types of algorithms: optimal tuning of the standard Metropolis Hastings (MH) samplers [17], or adapted proposition laws [18]. The adapted proposals exploit the information encoded in the target law in order to achieve an efficient exploration. They can be based on first order derivatives of the target, this being the case of the Metropolis Adjusted Langevin Algorithm (MALA) [19], [20] and of the Hamiltonian methods [21]. Another class of algorithms takes advantage of the target curvature through Hessian-based (Newton-like) proposals. Such terms have first been employed in optimization algorithms [22], but they have recently been successfully adapted to sampling [23]–[25]. From this stand point, our contribution is algorithmic and consists in integrating an improved Newton-like proposal [26] using the Fisher matrix instead of the inverse Hessian, i.e., the Fisher Random Walk MH (FRWMH), presented in [26]. In this manner, an important performance increase is achieved, due to the specific nature of our problem: there is no need for computing second order derivatives and thus a Newton type proposal is built using only first order derivatives.

To summarize, the main contributions are: (i) addressing the problem of texture model choice from indirect observations, (ii) tackling this problem in an optimal manner, (iii) introducing a new non-Gaussian model for the texture, tractable and efficient and (iv) enhancing the speed performances of the MCMC sampling phase, by using the FRWMH algorithm.

The paper is structured as follows: Section II presents the problem and the evidence-based model choice, Section III details the texture models and Section IV gives a Fisher information analysis of the problem. Section V is devoted to evidence computation and sampling, while Section VI presents the results. Section VII concludes the paper with comments on the method and perspectives for future work.

## II. PROBLEM STATEMENT AND PRELIMINARIES

Let us now present the method to choose the texture model, say  $\mathcal{M} = k$ , among  $K$  possible models. This is carried out starting from indirect observations, i.e., blurred and noisy data. The observation model is:

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

where  $\mathbf{y}$ ,  $\mathbf{x}$  and  $\mathbf{n} \in \mathbb{C}$  are the lexicographically ordered observations, unobserved textured image and noise, respectively. The images are of size  $N \times N$  and  $P = N^2$  denotes the number of pixels. The point spread function (PSF) is represented by the convolution matrix  $H$ , of size  $P \times P$ . The Fourier transform of the PSF is referred to as the Transfer Function (TF) and its coefficients are denoted  $\hat{h}_p$ .

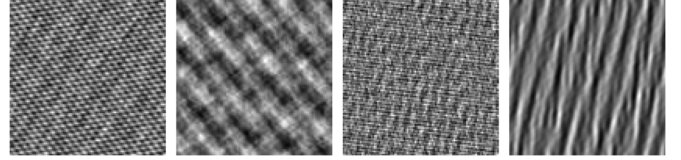


Fig. 1. Texture realizations of the employed image model.

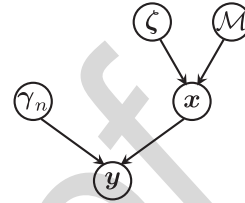


Fig. 2. Observation model graph and variable dependency.

### A. Distributions

The noise  $\mathbf{n}$  is zero-mean stationary Gaussian, with covariance matrix  $R_n(\boldsymbol{\gamma})$ , driven by the unknown parameter set  $\boldsymbol{\gamma}$ . From this model, the conditional law for  $\mathbf{y}$  writes:

$$f(\mathbf{y}|\mathbf{x}, \boldsymbol{\gamma}) \propto \det [R_n(\boldsymbol{\gamma})]^{-1} \exp \left[ -\|\mathbf{y} - H\mathbf{x}\|_{R_n(\boldsymbol{\gamma})}^2 \right] \quad (2)$$

The method is equally adapted to any type of noise correlation, however, our numerical study focuses on the white noise case, i.e.,  $R_n(\boldsymbol{\gamma}) = \gamma_n^{-1} \mathbf{I}$ , with  $\gamma_n$  the precision (inverse variance).

The texture models are parametric and driven by the parameter set  $\boldsymbol{\zeta}$ . Thus, for each model,  $k$ , we have the corresponding law  $f(\mathbf{x}|\boldsymbol{\zeta}_k, \mathcal{M} = k)$  and its associated parameter set  $\boldsymbol{\zeta}_k$ . The exact form of these laws and the structure of the parameter sets will be specified in Section III. Typical realizations of such textures are shown in Fig. 1.

In this context, our model selection method relies on a Bayesian hierarchical framework and Fig. 2 illustrates the variable interdependency. We are considering a case where the prior information about the parameters is very reduced, consequently, uninformative priors [27] are used. For  $\gamma_n$ , the conjugate form with respect to the likelihood (2) is a Gamma law:

$$\pi(\gamma_n) = \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \gamma_n^{\alpha_n - 1} \exp[-\beta_n \gamma_n] = \mathcal{G}(\alpha_n, \beta_n) \quad (3)$$

This law becomes an uninformative Jeffreys prior in the limit case ( $\alpha_n \rightarrow 0, \beta_n \rightarrow 0$ ).

The *a priori* distribution for the model is fully described by the  $p_k = \Pr(\mathcal{M} = k)$  probabilities. Our numerical study relies on an uninformative prior, i.e., equiprobable models:  $p_k = 1/K, k = 1 \dots K$ .

The priors for the texture parameters  $\boldsymbol{\zeta}_k$  will be defined in accordance with the employed texture models, in Section III.

### B. Decision Function

A decision function, denoted  $\Delta$ , associates a model  $k$  to the data  $\mathbf{y}$ , i.e.,  $\Delta(\mathbf{y}) = k$ . Moreover, a cost function quantifies the decision error. Let this function be  $C(k, k^*)$ , where  $k$  is the chosen model and  $k^*$  represents the true model. This cost has

the properties  $C(k, k^*) \geq 0$  and  $C(k^*, k^*) < C(k, k^*)$ , when  $k \neq k^*$ . Based on the cost, the risk is defined as:

$$\rho(\Delta) = \mathbb{E}_{\mathcal{M}, \mathbf{y}} [C(\Delta(\mathbf{y}), k)]$$

and the optimal decision function is  $\Delta_c = \arg \min_{\Delta} \rho(\Delta)$ . The optimality is achieved because the cost is averaged over all possible data, models and parameter values. Moreover, for a binary cost function  $C(k, k^*) = 1 - \delta(k, k^*)$ , the method implicitly selects the model with the Maximum *a Posteriori* (MAP) probability:

$$\hat{\mathcal{M}} = \arg \max_k \Pr(\mathcal{M} = k | \mathbf{y}) \quad (4)$$

Using Bayes' rule, the posterior model probabilities are:

$$\Pr(\mathcal{M} = k | \mathbf{y}) = \frac{f(\mathbf{y} | \mathcal{M} = k) \cdot p_k}{f(\mathbf{y})} \quad (5)$$

and require the two following quantities.

- 1) The probability distribution of the data,  $f(\mathbf{y})$ . Fortunately, it does not depend on the model, thus can be calculated by normalization.
- 2) The **evidence**/marginal likelihood,  $e_k = f(\mathbf{y} | \mathcal{M} = k)$ , obtained from the joint law of data and unknowns, given the model, by marginalizing the unobserved texture, the noise and the texture parameters:

$$e_k = \int_{\Psi} f(\mathbf{y}, \Psi | \mathcal{M} = k) d\Psi \quad (6)$$

where  $\Psi = \{\gamma_n, \mathbf{x}, \zeta_k\}$  gathers all the unknowns.

This law is written using the conditioning rule and the hierarchy shown in Fig. 2:

$$f(\mathbf{y}, \Psi | \mathcal{M} = k) = f(\mathbf{y} | \mathbf{x}, \gamma_n) \cdot f(\mathbf{x} | \zeta_k, \mathcal{M} = k) \cdot \pi(\zeta_k | \mathcal{M} = k) \cdot \pi(\gamma_n) \quad (7)$$

It now becomes clear that (6) is intractable due to the numerous layers of non-linearity and must be computed numerically. In this work, this is done by an MCMC algorithm.

A similar problem has been addressed in our previous work [28], in the context of direct observations,  $\mathbf{y} = \mathbf{x}$ , and for Gaussian textures. In that case, (6) reduced to a single integral with respect to the texture parameters.

### III. TEXTURE MODELING

From a theoretical standpoint, the proposed method can handle any texture model driven by the parameter set  $\zeta$ . In this context, the Gaussian RFs (GRF) model is simple and tractable. Its extension to a scale mixture of Gaussians (SMG) expands the representation capabilities without particularly affecting the algorithm efficiency. This model is based on a set of auxiliary variables  $\mathbf{s}$  such that, conditionally on these variables, the field  $\mathbf{x} | \mathbf{s}$  is Gaussian, but marginally  $\mathbf{x}$  is no longer Gaussian.

This idea, originating from statistics, has been exploited in problems of deconvolution and denoising [29]–[33] to formulate heavier tailed regularization terms and for modeling natural images [34]–[36]. The SMGs have been used so far to model the sparse character of the differences between neighboring pixels

(in the spatial domain) and of the wavelet coefficients (in the wavelet domain).

This is where our work is different. Here, we take into account that the blur introduced by the observation model is easily written as a multiplication in the Fourier domain. Moreover, the Fourier domain is well adapted for representing textural characteristics. Consequently, our SMG model is defined in the Fourier domain.

Hence the originality since, although the SMG model has been previously employed in image analysis, to the best of our knowledge, it has not been used so far to model the Fourier coefficients of a textured image.

For the sake of computational efficiency, we have made the assumption that the textures are zero-mean and stationary, thus the covariance matrix has a Toeplitz-block-Toeplitz structure. For general models,<sup>1</sup> the exact likelihood cannot be computed due to the huge dimension of the covariance to be inverted. But, by Whittle's approximation, this matrix has a Circulant-block-Circulant form, therefore is diagonalizable by discrete Fourier transform. The reader is invited to refer to [37], [38] for approximation properties and asymptotic behavior. Consequently, the Fourier coefficients, denoted by  $\hat{x}_p$ ,  $p = 1 \dots P$ , are decorrelated, conditionally on the power spectral density (PSD). The conditional law of the image is separable in the Fourier domain and writes:

$$f(\mathbf{x} | \zeta) \propto \prod_{p=1}^P \zeta_p \cdot \exp \left[ - \sum_{p=1}^P \zeta_p |\hat{x}_p|^2 \right] \quad (8)$$

Each Fourier coefficient follows a zero-mean Gaussian distribution, of variance given by the corresponding PSD element:

$$\hat{x}_p | \zeta_p \sim \mathcal{N}(0, \zeta_p^{-1}) \quad (9)$$

Due to the fact that the law for  $\mathbf{x} | \zeta$  is Gaussian, the Fourier coefficients are independent, conditionally on  $\zeta$ .

The model complexity and thus its representation capabilities can be enhanced, while keeping the limitation to independent Fourier coefficients, by changing i) the law of the Fourier coefficients and ii) the form of the PSD.

#### A. Fourier Coefficient Distributions

This work focuses on two models for the Fourier coefficient law, between which only the form of  $\zeta_p$  changes:

- Gaussian texture (GRF) -  $\zeta_p = \gamma_x \lambda_p(\boldsymbol{\theta})$
- non-Gaussian texture (SMGRF) -  $\zeta_p = \gamma_x s_p \lambda_p(\boldsymbol{\theta})$

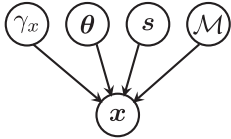
where  $\lambda_p(\boldsymbol{\theta})$  are the shape elements of the parametric PSD, driven by the parameter set  $\boldsymbol{\theta}$ . We will denote by  $\boldsymbol{\lambda}$  the parametric component of the PSD, i.e., the collection of all  $\lambda_p(\boldsymbol{\theta})$ .  $\gamma_x$  is a global scale parameter for the PSD and  $s_p$  is a local scale parameter.

The scale parameter pdf determines whether the textures follow a GRF or a SMGRF model and these parameters allow us to switch between the two laws, without changing the parametric part of the PSD,  $\boldsymbol{\lambda}$ .

In this context, Fig. 3 completes the problem description from Fig. 2 by showing the variable hierarchy for the two types of texture models.

<sup>1</sup>This is not the case if the field is defined through its precision matrix (inverse covariance), e.g. for Markov field.





- Non-Gaussian texture:

$$s = \{s_p\}_{p=1\dots P}$$

$$s_p \sim \mathcal{G}(\alpha_s, \beta_s)$$

- Gaussian texture:

$$s_p = 1, p = 1\dots P$$

Fig. 3. SMGRF texture model.

1) *Gaussian Model*: In the GRF case, the scales are identical:  $s_p = s = 1$ , for  $p = 1 \dots P$ , thus (8) and (9) become:

$$f(\mathbf{x}|\gamma_x, \boldsymbol{\theta}, s) \propto \gamma_x^P \left[ \prod_{p=1}^P \lambda_p(\boldsymbol{\theta}) \right] \cdot \exp \left[ - \sum_{p=1}^P \gamma_x \lambda_p(\boldsymbol{\theta}) |\hat{x}_p|^2 \right] \hat{x}_p | \gamma_x, \boldsymbol{\theta} \sim \mathcal{N} \left( 0, [\gamma_x \lambda_p(\boldsymbol{\theta})]^{-1} \right) \quad (10)$$

This is the model used in our previous papers [26], [28] for fast parameter sampling and pixel interaction model choice.

2) *Non-Gaussian Model*: In this case, a prior is assigned to  $s_p$  auxiliary variables and they are included in the estimation framework. Theoretically, any prior with positive support can be used. Some immediate examples are Markov or Wishart. Nevertheless, for numerical efficiency, it is essential for this prior to allow for parallel posterior processing (i.e., the  $s_p$  should be separable *a posteriori*) and to have some conjugacy property with respect to (8). For this reason, the paper is limited to *a priori* independent auxiliary variables, following a Gamma distribution  $\mathcal{G}(\alpha_s, \beta_s)$ :

$$f(s_p | \alpha_s, \beta_s) \propto s_p^{\alpha_s - 1} \exp[-\beta_s s_p] \quad (11)$$

Due to this separability, although  $f(\mathbf{x}|\gamma_x, \boldsymbol{\theta})$  (the marginal law with respect to  $\mathbf{s}$ ) is non-Gaussian, it remains separable. Here it has a Student's t form, hence (8) becomes:

$$f(\mathbf{x}|\gamma_x, \boldsymbol{\theta}, \alpha_s, \beta_s) \propto \prod_{p=1}^P \left[ 1 + \frac{\gamma_x \lambda_p(\boldsymbol{\theta})}{\beta_s} |\hat{x}_p|^2 \right]^{-\alpha_s - 1} \quad (12)$$

3) *Variance*: To individually evaluate the impact of each feature ( $\boldsymbol{\lambda}$ , scale parameter, Fourier coefficient pdf), we chose to keep the same marginal variance of the Fourier coefficients for the two models. In this manner, it is possible to compare textures with the same  $\boldsymbol{\lambda}$ , but different pdfs, or reversely, with the same pdf, but different forms for  $\boldsymbol{\lambda}$ .

The variances of the Fourier coefficients, conditionally on  $\boldsymbol{\theta}$  and  $\gamma_x$ , are given by:

$$\text{var}_{SMGRF}[\hat{x}_p | \gamma_x, \boldsymbol{\theta}] = \frac{\beta_s}{2\alpha_s - 1} \cdot \frac{1}{\gamma_x \lambda_p(\boldsymbol{\theta})}$$

$$\text{var}_{GRF}[\hat{x}_p | \gamma_x, \boldsymbol{\theta}] = \frac{1}{\gamma_x \lambda_p(\boldsymbol{\theta})}$$

Thus, the equality of these variances imposes a constraint on the parameters of the prior for  $s_p$ :  $\beta_s = 2\alpha_s - 1$ , with  $\alpha_s > 0.5$ . Then, the Fourier coefficients for the two types of textures have the same  $\boldsymbol{\lambda}$ , but different pdfs.

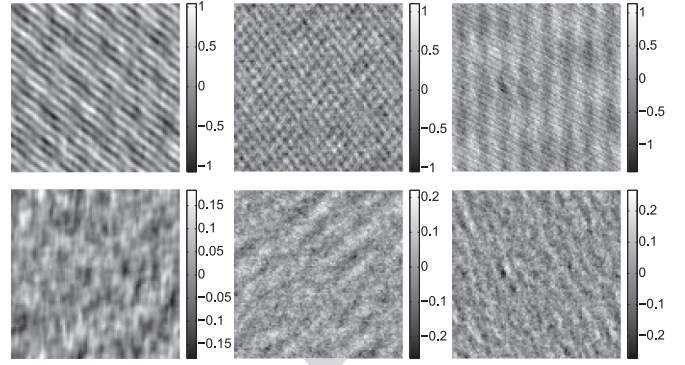


Fig. 4. Examples of texture pairs with the same PSD, but different law for the Fourier coefficients. On the first row, the SMGRF textures and on the second row the corresponding GRF textures.



Fig. 5. Examples of shapes for  $\boldsymbol{\lambda}$  that can be used with our GRF and SMGRF texture models. The representation is done in the Fourier domain on the reduced frequency domain  $[-0.5, 0.5] \times [-0.5, 0.5]$ .

Due to these common aspects, both models yield stochastic textures, however, the SMGRF is able to generate more complex patterns. Fig. 4 shows pairs of textures with the same  $\boldsymbol{\lambda}$ , but different pdfs. The  $\boldsymbol{\lambda}$  that were used to generate these textures have parametric forms, such as those depicted in Fig. 5, and will be analytically described in the next section (see Table I). The spectral content that is modeled using these forms consists in highlighting a cluster of spatially connected frequency components. For this reason, we can notice in Fig. 4 the strongly stochastic appearance of the GRF textures, with a dominant type of spectral components. The SMGRF textures present a spectrum (corresponding to the same parametric form of  $\boldsymbol{\lambda}$ ) with a more diverse range of frequencies and thus can represent more complex textures than the GRF. This is due to the presence of the auxiliary variables which increase the spectral complexity by increasing the amplitude of certain frequencies in the  $\boldsymbol{\lambda}$ . These supplementary components in addition to the grouped together high amplitude frequency components of the  $\boldsymbol{\lambda}$  will represent the new, richer PSD.

*Remark*: The SMGRF textures could also be obtained via the GRF model, however, for a non-parametric  $\boldsymbol{\lambda}$ . The  $\boldsymbol{\lambda}_{GRF} = \mathbf{s} \cdot \boldsymbol{\lambda}$ , which does not have a form that can be easily represented using a small number of parameters (is not compressible).

## B. Power Spectral Density Models

The form of the PSD actually encodes an important part of the textural content information. It must be stressed that there is virtually no constraint on the form of the PSD: it can be constant, corresponding to a white noise, or contain a single frequency. It can have a pulse-like form, this resulting in deterministic-like textures, or a more blobby shape, with one or several components, which would correspond to stochastic-like textures.

TABLE I  
EXAMPLES OF PSD MODELS USED IN THE SEQUEL OF THE PAPER

$\mathcal{M}$	Model $k$	Expression of $\lambda^k(\nu_x, \nu_y, \theta_k)$	$\theta_k$
1	Lorentz	$\left[1 + \frac{(\nu_x - \nu_{x_0})^2}{\sigma_x^2} + \frac{(\nu_y - \nu_{y_0})^2}{\sigma_y^2}\right]$	$\nu_{x_0}, \nu_{y_0}, \sigma_x, \sigma_y$
2	Generalized Lorentz	$\left[1 + \frac{(\nu_x - \nu_{x_0})^2}{\sigma_x^2} + \frac{(\nu_y - \nu_{y_0})^2}{\sigma_y^2}\right]^q$	$\nu_{x_0}, \nu_{y_0}, \sigma_x, \sigma_y, q$
3	Exponential	$\exp \frac{1}{2} \left[ \frac{ \nu_x - \nu_{x_0} }{\sigma_x} + \frac{ \nu_y - \nu_{y_0} }{\sigma_y} \right]$	$\nu_{x_0}, \nu_{y_0}, \sigma_x, \sigma_y$
4	Generalized Gauss	$\exp \frac{1}{2} \left[ \frac{ \nu_x - \nu_{x_0} ^q}{\sigma_x^q} + \frac{ \nu_y - \nu_{y_0} ^q}{\sigma_y^q} \right]$	$\nu_{x_0}, \nu_{y_0}, \sigma_x, \sigma_y, q$

From an algorithmic point of view, it is advantageous to use shapes encoded through a small number of parameters. For this reason, this paper focuses on parametric, unimodal functions, with a relatively reduced number of parameters: Lorentz ( $\mathcal{M} = 1$ ), Generalized Lorentz ( $\mathcal{M} = 2$ ), Exponential ( $\mathcal{M} = 3$ ) and Generalized Gaussian ( $\mathcal{M} = 4$ ). Each model  $k$  is driven by the corresponding  $\theta_k$  and Table I explicitly shows the  $\lambda^k(\theta_k)$ .

The parameters of these functions are the central frequencies  $\nu_{x_0}, \nu_{y_0}$ , the widths  $\sigma_x, \sigma_y$  and the positive shape parameter  $q$ , specific to models  $\mathcal{M} = 2$  and  $\mathcal{M} = 4$ .

We have chosen to include embedded models, for instance  $\mathcal{M} = 1$  and  $\mathcal{M} = 3$  are nested in  $\mathcal{M} = 2$  and  $\mathcal{M} = 4$ , respectively, for  $q = 1$ . This enables an analysis of the method capacity to penalize model complexity when the extra parameters do not trigger a significant model fit increase.

The textured images are spatially discrete, thus the PSD is defined on the reduced frequency domain [39], [40], i.e., the variables  $(\nu_x, \nu_y) \in [-0.5, 0.5]^2$ . Furthermore, let us consider that the Fourier coefficient  $p$  has the  $(\nu_m, \nu_n)$  position in the discrete reduced frequency domain. The  $\lambda_p$  from (10) and (12) are the elements of the PSD field at these discrete positions and depend on the model,  $\mathcal{M} = k$ , and on the parameters  $\theta_k$ . From this point forward, to explicitly show this dependency, the notation  $\lambda_p^k(\theta_k)$  is used and, more precisely:  $\lambda_p^k(\theta_k) = \lambda^k(\nu_m, \nu_n, \theta_k)$ .

The use of parametric models has the advantage of reducing the number of unknowns. On the other hand, this model defines a highly non-linear dependency of  $\lambda_p^k(\theta_k)$  with respect to  $\theta_k$ , as shown in Table I. This complicated dependency means that there is no conjugate form for this law.

Moreover, the prior information about the texture parameters  $\theta_k = \{\nu_{x_0}, \nu_{y_0}, \sigma_x, \sigma_y, q\}$  is very reduced, thus, uninformative priors will be used. Consequently, a uniform prior is employed:  $\pi(\theta_k | \mathcal{M} = k) = \mathcal{U}_{[\theta_k^m; \theta_k^M]}(\theta_k)$ , where  $\theta_k^m$  and  $\theta_k^M$  are the vectors containing the minima, respectively maxima, allowed values for the texture parameters. These values were chosen to ensure the coherence of the parameter values with their physical interpretation:  $\theta_k^m = \{-0.5, -0.5, 10^{-2}, 10^{-2}, 0.1\}$  and  $\theta_k^M = \{0.5, 0.5, 1, 1, 5\}$ .

This section devoted to the texture models has concluded the problem specification part of this paper. Based on the chosen observation model and the SMGRF image model, Section IV will present considerations regarding the problem difficulty and Section V will give the mathematical developments required to determine the most probable model.

#### IV. INFORMATION QUANTITATIVE ASSESSMENT

The data can be more or less informative and this is directly reflected in the problem difficulty. In order to evaluate this difficulty, this section presents a series of theoretical considerations regarding the available information, quantified by the Fisher information matrix.

Let us now focus on the information regarding any component of  $\Psi$ , denoted by  $\psi$ , by analyzing the diagonal elements of the information matrix, i.e., the expectation of the second derivative of the co-log-likelihood:

$$\mathcal{I}(\psi) = -\mathbb{E}_{\mathbf{y}|\Psi} \left[ \frac{\partial^2}{\partial \psi^2} \log f(\mathbf{y}|\Psi) \right] \quad (13)$$

Due to the separability and Gaussianity of the noise model and the form of (1), the law  $f(\mathbf{y}|\Psi)$  is also separable, Gaussian, zero-mean and of variance:

$$r_p(\Psi) = g_p [\gamma_x s_p \lambda_p(\theta)]^{-1} + \gamma_n^{-1} \quad (14)$$

where  $g_p = |\hat{h}_p|^2$ . Equation (13) contains first and second order derivatives of  $r_p(\Psi)$  with respect to  $\psi$ . When calculating the expectation, knowing that  $\mathbb{E}_{\mathbf{y}|\Psi} [|\hat{y}_p|^2] = r_p(\Psi)$ , the second order derivatives cancel out. Then:

$$\mathcal{I}(\psi) = \sum_{p=1}^P \left[ \frac{1}{r_p(\Psi)} \cdot r'_p(\Psi) \right]^2 \quad (15)$$

where  $r'_p(\Psi)$  is the derivative of  $r_p(\Psi)$  with respect to  $\psi$ .

For the noise parameter,

$$\mathcal{I}(\gamma_n) = \sum_{p=1}^P \left[ \gamma_n^{-1} + \gamma_x^{-1} \cdot g_p [s_p \lambda_p(\theta)]^{-1} \right]^{-2} \quad (16)$$

is a decreasing function, thus, the smaller the  $\gamma_n$ , i.e., the larger the noise level, the easier its estimation.

For a texture coefficient  $\theta$ , element of  $\theta$ ,

$$\mathcal{I}(\theta) = \sum_{p=1}^P \left[ \frac{g_p \cdot \lambda'_p(\theta)}{\lambda_p(\theta) [g_p + \gamma_x / \gamma_n \cdot s_p \lambda_p(\theta)]} \right]^2 \quad (17)$$

depends on the values of the other texture parameters. For instance, the information regarding the central frequency  $\nu_x^0$  is higher if the frequency  $\nu_y^0$  is close to 0 and decreases symmetrically with the absolute value of  $\nu_y^0$  (see Fig. 6(a)).  $\mathcal{I}(\nu^0)$  also increases when the width  $\sigma$  decreases, i.e., the characteristic is more concentrated, as in Fig. 6(b). Moreover, the amount of information available for estimating the texture parameters also varies according to the PSD model. Finally, concerning the signal and noise levels, the lower the SNR, the smaller  $\mathcal{I}(\theta)$  in (17) and, thus, the less information on  $\theta$ .

*Remark:* The higher information around the null frequency is due to the low pass character of the PSF. This becomes obvious in Fig. 6(d) where the amount of information for the high frequencies drops significantly when the filter becomes more selective (the inverse width is higher).

Another interesting case is the noiseless scenario ( $\gamma_n = \infty$ ):

$$\mathcal{I}(\theta) = \sum_{p=1}^P \left[ \frac{1}{\lambda_p(\theta)} \cdot \lambda'_p(\theta) \right]^2 \quad (18)$$

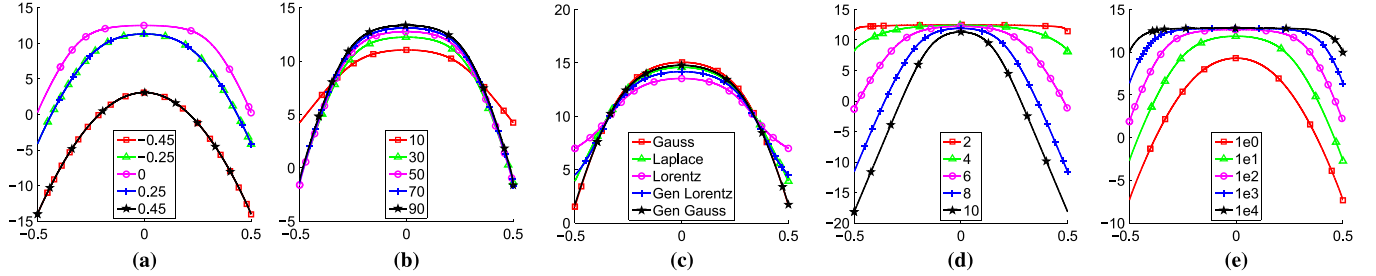


Fig. 6. Variation of the Fisher information for  $\nu_x^0$ , as a function of various factors. Only one parameter is varied at a time in order to quantify its effect on the amount of information for  $\nu_x^0$  (in logarithmic scale). (a)  $\nu_y^0$  sweep; (b)  $\sigma_x^{-1}$  sweep; (c) model sweep; (d) OTF inverse width; (e)  $\gamma_n$  sweep.

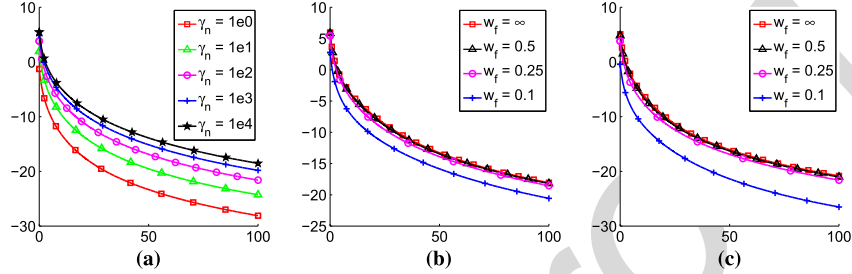


Fig. 7. Fisher information for  $w_x$  for various values of the observation system parameters  $w_f$  and  $\gamma_n$  (in logarithmic scale). (a)  $\gamma_n$  sweep; (b)  $w_f$  sweep-high  $\gamma_n$ ; (c)  $w_f$  sweep-low  $\gamma_n$ .

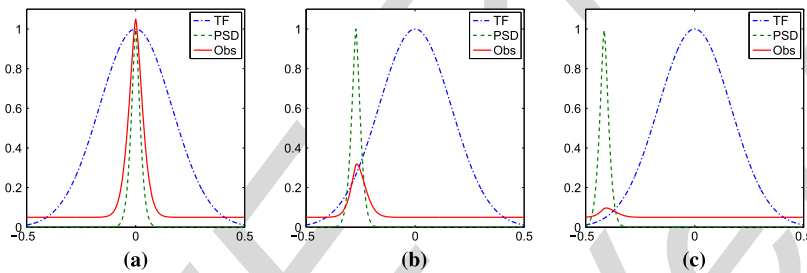


Fig. 8. Different cases emerging from the relative positioning and widths of the TF and PSD. (a) Total overlap; (b) partial overlap; (c) no overlap.

The Fisher information regarding  $\theta$  depends only on the PSD (as in [26]). This has multiple implications on the amount of information for the texture parameters:

- it does not depend on the form of the TF (known TF),
- it is the same, whether the texture is Gaussian or not.

Fig. 7 illustrates the amount of information concerning one width of the PSD for various levels of noise and widths of the PSF. In Fig. 7(a), the variation of the Fisher information with the noise precision shows a significant difference in the amount of information. Situations with SNR < 20 dB are challenging, since the available information is reduced. Figs. 7(b) and 7(c) illustrate the Fisher information variation for a PSD width as a function of the TF width  $w_f$ , the SNR being fixed at 30 dB in Fig. 7(b) and at 10 dB in Fig. 7(c). These plots show that the Fisher information does not strongly depend on the  $w_f$ , especially in the low noise case (high  $\gamma_n$ ) from Fig. 7(b).

Last, but not least, for the auxiliary variables  $s_p$ ,

$$\mathcal{I}(s_p) = \left\{ \frac{g_p}{[s_p(g_p + \gamma_x/\gamma_n \cdot s_p \lambda_p(\theta))]} \right\}^2 \quad (19)$$

the information depends on the corresponding coefficient of the PSD, TF and  $\gamma_n$ . Moreover, for the noiseless case,  $\mathcal{I}(s_p) = s_p^{-2}$ , i.e., the smaller the value, the more the information.

Fig. 8 illustrates various scenarios for a Gaussian TF and a Laplacian texture PSD. Their product plus the noise, of variance  $1/\gamma_n$ , yields the observation's PSD. The extreme cases are either simplifications:

- PSD centered near the null frequency and narrower than the TF, the effect of the convolution being negligible,
- noiseless case ( $\gamma_n \rightarrow \infty$ ),
- asd

or cases with signal alterations so severe that information on the original PSD is no longer present in the data:

- the PSD is positioned in a region with strong TF attenuation, the original image information being lost,
- high noise case ( $\gamma_n \rightarrow 0$ ). This corresponds to a low Signal to Noise Ratio (SNR)  $\gamma_n/\gamma_x$ .

This section offers a prior performance analysis based on the Fisher information matrix. The amount of available information is directly related to the estimation performances, allowing us to predict these performances in various situations.

## V. EVIDENCE CALCULATION

The full description of our Bayesian model choice method relies (i) on the data model and (ii) on the priors for the unknowns.

The hierarchical direct model is shown in Fig. 2, while the texture model specificities are shown in Fig. 3. In the following, the emphasis is on the SMGRF, which encompasses the GRF. By writing (6) as:

$$e_k = \int_{\mathbf{x}, \gamma_x, \gamma_n, \boldsymbol{\theta}_k, \mathbf{s}} f(\mathbf{y}|\boldsymbol{\Psi}, \mathcal{M} = k) \cdot \pi(\boldsymbol{\Psi}|\mathcal{M} = k) d\mathbf{s} d\boldsymbol{\theta}_k d\gamma_x d\gamma_n d\mathbf{x} \quad (20)$$

and plugging in (2) and (8), the intractability of the integral is obvious. For this reason, it must be calculated numerically and the solution chosen here is sampling.

A natural idea is to straightforwardly determine the evidence  $e_k$  from samples of the prior  $\pi(\boldsymbol{\Psi}|\mathcal{M} = k)$ :  $\boldsymbol{\Psi}^{(t)} = (\mathbf{x}^{(t)}, \gamma_n^{(t)}, \gamma_x^{(t)}, \mathbf{s}^{(t)}, \boldsymbol{\theta}_k^{(t)})$ ,  $t = 1 \dots T$  as follows:

$$\bar{e}_k \approx \frac{1}{T} \sum_{t=1}^T f(\mathbf{y}|\boldsymbol{\Psi}^{(t)}, \mathcal{M} = k) \quad (21)$$

It consists in sampling the priors and computing the arithmetic mean of the corresponding likelihood values—Arithmetic Mean Approximation (AMA). Evidence computation based on prior samples can also be done by nested sampling [41].

Nevertheless, when the likelihood is very peaked, as in the current case, most of these samples have weak likelihood, i.e., an insignificant contribution and thus, the algorithm is slow to converge. For this reason, it is more suitable to compute (20) by sampling the posterior  $\pi(\boldsymbol{\Psi}|\mathbf{y}, \mathcal{M} = k)$ .

#### A. Posterior Sampling

The posterior law is proportional to the joint law:

$$\begin{aligned} f(\mathbf{y}, \boldsymbol{\Psi}, \mathcal{M} = k) &= \mathcal{C} \cdot \exp \left[ -\gamma_n \sum_{p=1}^P |y_p - \overset{\circ}{h}_p \overset{\circ}{x}_p|^2 \right] \\ &\cdot \gamma_n^{P+\alpha_n-1} \exp[-\beta_n \gamma_n] \cdot \prod_{p=1}^P [s_p \cdot \lambda_p^k(\boldsymbol{\theta}_k)] \\ &\cdot \gamma_x^{P+\alpha_x-1} \exp[-\beta_x \gamma_x] \exp \left[ -\gamma_x \sum_{p=1}^P |\overset{\circ}{x}_p|^2 s_p \lambda_p^k(\boldsymbol{\theta}_k) \right] \\ &\cdot \mathcal{U}_{[\boldsymbol{\theta}_k^m, \boldsymbol{\theta}_k^M]}(\boldsymbol{\theta}_k) \cdot \prod_{p=1}^P s_p^{\alpha_s-1} \cdot \exp \left[ -\beta_s \sum_{p=1}^P s_p \right] \end{aligned} \quad (22)$$

where the normalization constant is  $\mathcal{C} = K^{-1} \cdot (2\pi)^{-3P} \cdot \beta_n^{\alpha_n} \cdot \Gamma^{-1}(\alpha_n) \cdot \beta_x^{\alpha_x} \cdot \Gamma^{-1}(\alpha_x) \cdot \beta_s^{P\alpha_s} \cdot \Gamma(\alpha_s)^{-P} \cdot (\boldsymbol{\theta}_k^M - \boldsymbol{\theta}_k^m)^{-1}$ .

However, this law cannot be directly sampled, thus MCMC methods will be employed, more precisely, Gibbs sampling. This can be performed via two types of algorithms.

- Across-model approach—joint sampling of the model index and its parameters. The algorithm jumps from one model to another and explores the joint model index plus parameter space, yielding a joint chain of model indexes and parameter values. The most representative algorithm of this type is Reversible Jump MCMC (RJCMC) [42].

- Within-model approach—consists in exhaustively visiting the candidate models and parameter sampling conditionally on the model. It provides  $K$  chains of parameter values, one for each model. For a detailed description see [43], [44] and the more recent survey [45].

Despite the conceptual differences, for a finite candidate models set, the two approaches yield the same result (provided they have reached convergence) but, under two different forms.

The RJCMC algorithm is especially interesting for a very large number of models, when an exhaustive sequential sampling of all the models may be prohibitively expensive. However, in our problem, the number of concurrent models is rather reduced. In this case, the within-model strategy is better, avoiding the non-trivial RJCMC problems concerning the parameter transformation when switching models. Moreover, the within-model approach guarantees that all models have been thoroughly explored and the model selection is not affected by the sampling algorithm. For this reason, our model choice method is based on within-model posterior sampling.

#### B. Evidence Approximations Based on Posterior Samples

Our model selection method is based on evidence approximation from posterior samples and the proposed method will be referred to as the Classifier based on Evidence Approximation from Posterior Samples (CEAPS). This method can be formulated using two different approximations of the evidence, which we present in the following: the Harmonic Mean Approximation (HMA) [44] and the Laplace-Metropolis Approximation (LMA) [16].

1) *Harmonic Mean Approximation*: We consider that  $\boldsymbol{\Psi}^{(t)} = (\mathbf{x}^{(t)}, \gamma_n^{(t)}, \gamma_x^{(t)}, \mathbf{s}^{(t)}, \boldsymbol{\theta}_k^{(t)})$ ,  $t = 1 \dots T$  are samples of the *a posteriori* law. Then, the evidence can be computed as:

$$\tilde{e}_k \approx \left\{ \frac{1}{T} \sum_{t=1}^T \left[ f(\mathbf{y}|\boldsymbol{\Psi}^{(t)}, \mathcal{M} = k) \right]^{-1} \right\}^{-1} \quad (23)$$

i.e., the harmonic mean of the likelihood values for the  $\boldsymbol{\Psi}^{(t)}$ .

Although  $\tilde{e}_k$  converges almost surely to the true value  $e_k$  when  $T \rightarrow \infty$  [46], it does not generally satisfy the central limit theorem [45]. Occasionally, a  $\boldsymbol{\Psi}^{(t)}$  with significant *a priori* probability, but very low likelihood, may occur. Its contribution in the harmonic mean is high and this may trigger infinite variances [44]. Solutions to stabilize this approximation have been provided in [47]. Nevertheless, we have not encountered this difficulty neither in our previous work [28], nor in the current one, where the priors are uniform on a finite interval and the likelihood is very peaked. Hence, the posterior samples are distributed in the regions where the likelihood has significant values. Consequently, the situations where the HMA may diverge or converge too slowly are avoided.

2) *Laplace-Metropolis Approximation*: The evidence (20) can also be expressed as:

$$e_k = \int_{\boldsymbol{\Psi}} \exp \left\{ \underbrace{\log [f(\mathbf{y}|\boldsymbol{\Psi}, \mathcal{M} = k) \cdot \pi(\boldsymbol{\Psi}|\mathcal{M} = k)]}_{F_k(\boldsymbol{\Psi}, \mathbf{y})} \right\} d\boldsymbol{\Psi} \quad (24)$$

with  $F_k(\boldsymbol{\Psi}, \mathbf{y})$  the log-posterior computed for observation  $\mathbf{y}$ .

*Remark:*  $F_k(\Psi, \mathbf{y})$  is called the observed information and indicates the amount of available information, for the given observation  $\mathbf{y}$ .  $F_k(\Psi, \mathbf{y})$  is related to the Fisher information  $\mathcal{I}_k$  introduced in Section IV through (13) as follows:

$$\mathcal{I}_k(\Psi) = -\mathbb{E}_{\mathbf{y}} [F_k''(\Psi, \mathbf{y})] \quad (25)$$

Under the hypothesis that  $F_k(\Psi, \mathbf{y})$  is twice differentiable with a unique maximum in  $\Psi^*$ , the Laplace approximation of a quadratic function (also used for the derivation of BIC) can be applied to evaluate the integral (24):

$$\hat{e}_k \approx \exp [PF_k(\Psi_k^*, \mathbf{y})] \cdot \left(\frac{2\pi}{P}\right)^{\frac{D_k}{2}} \cdot |F_k''(\Psi_k^*, \mathbf{y})|^{-1/2} \quad (26)$$

where  $\Psi_k^*$  represents the MAP value for model  $k$ ,  $F_k''(\Psi_k^*, \mathbf{y})$  is the Hessian of the log-posterior, evaluated at  $\Psi_k^*$  and  $D_k$  is the dimension of model  $k$ . The last factor in (26) is the determinant of the observed information matrix. In fact, computing the evidence in this manner consists in determining the MAP value,  $\Psi^*$ , i.e., the value for which  $F(\Psi, \mathbf{y})$  is maximum, and replacing this value in (26). [48] reviews the Laplace based methods for evidence computation. These approximation methods have relative errors of order  $\mathcal{O}(P^{-1})$ .

This approximation can also be performed based on MCMC samples from the posterior. In this case the method is called Laplace Metropolis Approximation (LMA) [16]. The LMA can be based in the MAP, the Posterior Mean (PM), or the Median *a Posteriori* (MedAP) and the Hessian can be approximated by the covariance matrix of the samples. In our case, this approximation is performed using the PM and the value of the Hessian computed for the PM.

*Remark:* The LMA explicitly penalizes complex models due to the second factor that decreases exponentially with model dimension, while for the HMA the model complexity penalization is implicit.

### C. Gibbs Within-Model Posterior Sampling

The samples from the posterior law are obtained using Gibbs sampling. Among the various strategies, we have chosen to sample  $\gamma_n$ ,  $\gamma_x$ ,  $\mathbf{x}$ ,  $\boldsymbol{\theta}_k$  and  $\mathbf{s}$ . The advantage of this approach is that we obtain rather standard targets and we can perform parallel sampling for  $\mathbf{x}$  and  $\mathbf{s}$ . Then, the *a posteriori* conditional laws for the parameters are the following:

$$\begin{aligned} \mathbf{x} &\sim \prod_p \exp \left[ - \left( \gamma_n |\dot{y}_p - \dot{h}_p \dot{x}_p|^2 + \gamma_x |\dot{x}_p|^2 s_p \lambda_p^k(\boldsymbol{\theta}_k) \right) \right] \\ &= \prod_p \mathcal{N}(m_p, v_p), \text{ with } m_p = \gamma_n \dot{y}_p \dot{h}_p v_p \text{ and } \\ &v_p = (\gamma_n g_p + \gamma_x s_p \lambda_p^k(\boldsymbol{\theta}_k))^{-1} \text{—separable in the Fourier} \\ &\text{domain, i.e., parallel sampling is possible; computation} \\ &\text{cost equivalent to sampling the } a \text{ priori law;} \end{aligned}$$

$$\mathbf{s} \sim \prod_p \left\{ s_p^{\alpha_s} \cdot \exp -s_p \left[ \gamma_x |\dot{x}_p|^2 \lambda_p^k(\boldsymbol{\theta}_k) + \beta_s \right] \right\}$$

$= \prod_p \mathcal{G}(a_s, b_{s_p})$  with  $a_s = \alpha_s + 1$  and  $b_{s_p} = \beta_s + \gamma_x s_p |\dot{x}_p|^2 \lambda_p^k(\boldsymbol{\theta}_k)$ —separable, independent on the observations, allowing for parallel sampling;

$$\begin{aligned} \gamma_n &\sim \gamma_n^{P+\alpha_n-1} \cdot \exp -\gamma_n \left( \beta_n + \sum_p |\dot{y}_p - \dot{h}_p \dot{x}_p|^2 \right) \\ &= \mathcal{G}(a_n, b_n), \text{ with } a_n = \alpha_n + P \text{ and } b_n = \beta_n + \sum_p |\dot{y}_p - \dot{h}_p \dot{x}_p|^2; \end{aligned}$$

$$\begin{aligned} \gamma_x &\sim \gamma_x^{P+\alpha_x-1} \cdot \exp -\gamma_x \left( \beta_x + \sum_p |\dot{x}_p|^2 s_p \lambda_p^k(\boldsymbol{\theta}_k) \right) \\ &= \mathcal{G}(a_x, b_x), \text{ with } a_x = \alpha_x + P \text{ and } b_x = \beta_x + \sum_p |\dot{x}_p|^2 s_p \lambda_p^k(\boldsymbol{\theta}_k); \end{aligned}$$

$$\boldsymbol{\theta}_k \sim \mathcal{U}(\boldsymbol{\theta}_k) \cdot \prod_{p=1}^P \lambda_p^k(\boldsymbol{\theta}_k) \cdot \exp \left[ -\gamma_x s_p |\dot{x}_p|^2 \lambda_p^k(\boldsymbol{\theta}_k) \right]$$

$f(\boldsymbol{\theta}_k | \mathbf{x}, \mathbf{s}, \gamma_x, \mathcal{M} = k)$ —very complicated dependency. Since  $\boldsymbol{\theta}$  has a non-standard law, more complex sampling strategies must be used. We chose to include an MH step [49] in the Gibbs loop. This Metropolis within Gibbs strategy is convergent, as proven in [50], i.e., provides a chain of values  $\Psi^{(t)}$ ,  $t = 1 \dots T$  that has the posterior as limit probability distribution (when the number of iteration tends to infinity).

*Remark:* The alternative to sampling all the unknowns is to integrate a part of them, in order to avoid their sampling steps. This strategy is similar to the collapsed Gibbs sampler method used in [1]. Some of the options are the following:

- $\mathbf{x}$  marginalization and sampling the rest of the unknowns, thus no texture sampling, but even more cumbersome dependency on  $\gamma_x$ ,  $\boldsymbol{\theta}_k$ ,  $\mathbf{s}$  and  $\gamma_n$ . Regarding  $\mathbf{s}$ , it remains separable, thus parallel sampling is feasible, but not of Gamma laws. Furthermore, the posterior for  $\gamma_n$  and  $\gamma_x$  no longer have Gamma forms either:

$$\begin{aligned} f(s_p, \gamma_n, \boldsymbol{\theta}_k | *) &\propto \exp \left[ \frac{1}{r_p} \frac{|\dot{y}_p|^2 g_p}{\gamma_x s_p \lambda_p(\boldsymbol{\theta}_k)} \right] \\ &\cdot \exp - [r_p \gamma_n \gamma_x s_p \lambda_p(\boldsymbol{\theta}_k)] \left[ \dot{x}_p - \frac{1}{r_p} \frac{\dot{y}_p \dot{h}_p}{\gamma_x s_p \lambda_p(\boldsymbol{\theta}_k)} \right]^2 \end{aligned}$$

- $\mathbf{s}$  marginalization, resulting in a complicated law for  $\mathbf{x}$  (loss of the advantage of the SMGRF texture model—the conditional Gaussianity for  $\dot{x}_p$ ). However,  $\dot{x}_p$  remain independent, thus can be sampled in parallel, but by more sophisticated samplers. The new conditional laws are:

$$\begin{aligned} f(\dot{x}_p | *) &\propto \left[ 1 + \frac{\gamma_x \lambda_p^k(\boldsymbol{\theta}_k)}{\beta_s} |\dot{x}_p|^2 \right]^{-\alpha_s-1} \\ &\cdot \exp \left[ -\gamma_n |\dot{y}_p - \dot{h}_p \dot{x}_p|^2 \right] \end{aligned}$$

- integrate both  $\mathbf{x}$  and  $\mathbf{s}$  and sample only  $\gamma_n$ ,  $\gamma_x$  and  $\boldsymbol{\theta}_k$ , which will be distributed under complicated laws.



Moreover, by marginalization, the resulting law is more diluted, thus what is gained by eliminating sampling steps, may be lost in terms of speed of convergence. We have chosen not to integrate any of them since, although this implies more sampling steps, the sampled laws are easier to handle.

#### D. Efficient Metropolis-Hastings for $\theta$

A Random Walk MH (RWMH) step consists in formulating a proposal  $\theta_{\text{pr}}$ , based on a proposition law of the form  $q(\theta_{\text{pr}}|\theta_c) = \theta_c + \phi$ , evaluating the acceptance probability  $\xi(\theta_c, \theta_{\text{pr}})$ , which is a function of the current value and the proposal, and then accepting or rejecting the proposal. If the proposal is accepted,  $\theta^{(t)} = \theta_{\text{pr}}$ , otherwise  $\theta^{(t)} = \theta_c$ . The algorithm is known to be convergent [51]: the samples are asymptotically distributed under the posterior law.

There are numerous options for formulating the proposal and both the convergence speed and mixing properties are directly influenced by the adequacy between the proposal law and the target. Thus, choosing a  $q$  that embeds information about the shape of the target can significantly enhance the algorithm performances. In this context, the RWMH algorithm formulates the proposal based on a stochastic component and, in some cases, a component based on the target. This can contain first or second order derivatives of the target, so as to ensure a convenient exploration of the parameter space.

In [26] we have developed an efficient sampler called the FRWMH. This algorithm is based on the idea of quasi-Newton proposals [23] for a fast exploration of the parameter space and superior mixing properties. In this context, the Hessian used in the quasi-Newton approach is replaced by the Fisher information matrix. The FRWMH algorithm formulates the proposal:

$$\theta_{\text{pr}} = \theta_c + \varepsilon \cdot \mathcal{I}^{-1}(\theta_c) \cdot \nabla_{\theta} \mathcal{L}(\theta_c) + \sqrt{\mathcal{I}^{-1}(\theta_c)} \cdot \mathbf{z}_c \quad (27)$$

where  $\mathcal{L}(\theta) = \log \pi(\theta|\mathbf{y})$  is the log-posterior,  $\mathbf{z}_c \sim \mathcal{N}(0, \mathbf{I})$  is an isotopic displacement and  $\mathcal{I}(\theta)$  is the Fisher information matrix, already defined in Section IV by (13).

In this case, the acceptance probability  $\xi(\theta_c, \theta_{\text{pr}})$  becomes:

$$\xi(\theta_c, \theta_{\text{pr}}) = \min \left\{ 1, \frac{\pi(\theta_{\text{pr}}) \cdot q(\theta_{\text{pr}}, \theta_c)}{\pi(\theta_c) \cdot q(\theta_c, \theta_{\text{pr}})} \right\} \quad (28)$$

with

$$q(\theta_{\text{pr}}, \theta_c) = \mathcal{N}(\theta_{\text{pr}} - \theta_c - \varepsilon \cdot \mathcal{I}^{-1}(\theta_c) \cdot \nabla_{\theta} \mathcal{L}(\theta_c), \mathcal{I}^{-1}(\theta_c))$$

The use of this proposal proved advantageous from multiple points of view. Firstly, this exploits the target curvature similarly to the Newton step from the optimization theory. Secondly, this made way for a series of algorithmic simplifications and performance enhancements:

- for our GRF textures, the second order derivatives vanished under the expectation. Thus, the efficient proposal was formulated only based on first order derivatives,
- the Fisher matrix is positive definite, hence when it is well-conditioned there are no instabilities when taking the inverse, such as those mentioned in [23] for the Hessian,
- also due to the positive definite Fisher matrix, the Newton term always has the direction of gradient ascent, thus the algorithm only makes efficient steps.

The same principle of using the Fisher matrix to formulate an efficient proposal is exploited in the manifold Metropolis adjusted Langevin algorithm (mMALA) [20].

#### E. Implementation Issues

The implementation has raised two numerical problems:

- 1) Since the likelihood has an exponential form for each Fourier coefficient and consists in a product over all the coefficients, this quantity often exceeds Matlab's representation capabilities. To overcome this obstacle, the co-log-likelihood (CLL), i.e., the negative log-likelihood is computed instead of the likelihood.
- 2) The problem is again encountered when computing the evidence from the CLL. The employed solution is to determine the minimum value of each CLL chain, subtract it from all the CLL chains and compute the evidences based on the "offset" values, in a logsumexp manner. The normalization is reversed in the final stage of posterior probability computation.

The resulting CEAPS procedure is given in Algorithm 1, where the "ComputeEvidence" routine can be based on either of the two aforementioned approximations.

These previous sections have presented theoretical aspects regarding the employed method, its mathematical formulation and the corresponding implementation issues. The following section will present a series of tests that evaluate the method performances.

## VI. EXPERIMENTAL RESULTS

This section is devoted to the description and interpretation the performances of our CEAPS model selection method from blurred and noisy textured images. Several experiments are presented.

- A. The first study compares the two MH samplers: the standard RWMH and the FRWMH.
- B. The second test compares the two evidence approximations based on posterior samples, i.e., HMA and LMA.
- C. The evaluation of our CEAPS classifier represents the third test. The classification performances for the CEAPS are first presented for various PSDs and then compared to those of the GMLE classifier.
- D. The classification performances are linked with the Fisher information analysis of Section IV.
- E. Results concerning the deconvolution are given in a visual form, by presenting the original textures, the observations and the deconvolved images.

Algorithm 1 proceeds as follows: a sampler is launched for each of the  $K$  concurrent models. These samplers compute at every iteration the evidence, based on the samples that have been drawn so far. When the difference between two consecutive values of the evidence is smaller than a threshold, the sampling is stopped. The first 10% of the samples are discarded as burn-in and the evidence is recomputed based only on the remaining 90% of the samples. Since the sampling is the most time consuming operation, it determines the duration of the overall algorithm and thus it is important to use efficient samplers.



---

**Algorithm 1:** Classifier based on Evidence Approximation from Posterior Samples (CEAPS) Algorithm
 

---

**input :** Data  $\mathbf{y}$ , models dictionary for  $\mathcal{M} = k, k = 1 \dots K$

**output:** Evidences  $\tilde{e}_k$  + samples for texture  $\mathbf{x}^{(t)}$ , noise parameter  $\gamma_n^{(t)}$ , texture parameters  $\boldsymbol{\theta}_k^{(t)}$  ( $t = 1 \dots T$ )

*% prior model probabilities:*  
 $p_k = 1/K$ ;

*% generate samples of  $\pi(\boldsymbol{\theta}_k, \mathbf{x}, \mathbf{s}, \gamma_n | \mathbf{y}, \mathcal{M} = k)$*

**for**  $k = 1$  **to**  $K$  **do**

*% Gibbs sampler for  $(\gamma_n, \mathbf{s}, \boldsymbol{\theta}_k, \mathbf{x})$ ,  $\mathcal{M} = k$  fixed:*

$t = 1$ ;

initialization  $\boldsymbol{\theta}_k^{(t)}, \mathbf{x}^{(t)} = \mathbf{y}, \mathbf{s}^{(t)}$ ;

$\mathbf{CLL}_k(t) = \text{CoLogL}(\mathbf{x}^{(t)}, \mathbf{s}^{(t)}, \boldsymbol{\theta}_k^{(t)})$ ;

$m(k) = \min(\mathbf{CLL}_k)$ ;

$\mathbf{nCLL}_k = \mathbf{CLL}_k - m(k)$ ;

*% compute the evidence using HMA or LMA*

$\tilde{e}_k = \text{ComputeEvidence}(\mathbf{nCLL}_k)$ ;

$e_{rec}(t) = 0$ ;

**while**  $|e_{rec}(t) - \tilde{e}_k| > \epsilon$  **do**

$t = t + 1$ ;

$\gamma_n^{(t)} \sim f(\gamma_n | \mathbf{y}, \mathbf{x}^{(t-1)}, \alpha_n, \beta_n)$

$\gamma_x^{(t)} \sim f(\gamma_x | \mathbf{x}^{(t-1)}, \boldsymbol{\theta}_k^{(t-1)}, \mathbf{s}^{(t-1)}, \alpha_x, \beta_x)$

$\mathbf{s}^{(t)} \sim f(\mathbf{s} | \mathbf{x}^{(t-1)}, \boldsymbol{\theta}_k^{(t-1)}, \alpha_s, \beta_s)$

$\boldsymbol{\theta}_k^{(t)} = \text{FRWMH}$  with target  $f(\boldsymbol{\theta}_k | \mathbf{x}^{(t-1)}, \mathbf{s}^{(t)})$

$\mathbf{x}^{(t)} \sim f(\mathbf{x} | \mathbf{y}, \gamma_x^{(t)}, \gamma_n^{(t)}, \mathbf{s}^{(t)}, \boldsymbol{\theta}_k^{(t)})$

$\mathbf{CLL}_k(t) = \text{CoLogL}(\mathbf{x}^{(t)}, \mathbf{s}^{(t)}, \boldsymbol{\theta}_k^{(t)})$ ;

$m(k) = \min(\mathbf{CLL}_k)$ ;

$\mathbf{nCLL}_k = \mathbf{CLL}_k - m(k)$ ;

$\tilde{e}_k = \text{ComputeEvidence}(\mathbf{nCLL}_k)$ ;

$e_{rec}(t) = \text{RecursiveEvidence}(e_{rec}(t-1), \tilde{e}_k)$ ;

**end**

**end**

*% determine the posterior model probabilities*

**for**  $k \leftarrow 1$  **to**  $K$  **do**

$\Pr(\mathcal{M} = k | \mathbf{y}) = p_k \cdot \tilde{e}_k / \sum_{l=1}^K p_l \cdot \tilde{e}_l \cdot \exp[m(k) - m(l)]$

*% Compute the parameter estimates by PM*

$\hat{\boldsymbol{\theta}}_k = \text{PM}(\boldsymbol{\theta}_k)$ ;

**end**

---

### A. RWMH vs FRWMH

The first tests investigate the speed performances of two sampling algorithms, the isotropic RWMH and our efficient FRWMH, in the context of the complicated laws for  $\boldsymbol{\theta}$ .

Our tests indicate that the use of the FRWMH yields an algorithmic speed increase by a factor of at least 10 as compared to the RWMH. This is due to the FRWMH directional form of the proposal, which permits the algorithm to attain the high probability region in a very small number of iterations. Once in this region, the directional component exhibits negligible values and the algorithm explores this high probability area of the parameter space due to the stochastic component of its proposal.

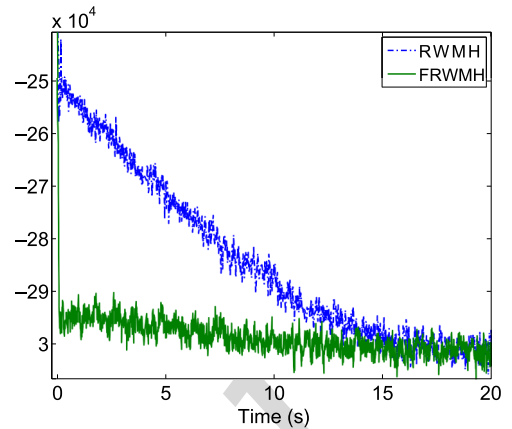


Fig. 9. CLL evolution—posterior sampling starting from the same initialization, using RWMH and FRWMH.

This translates into a very short burn-in period, as opposed to the isotropic RWMH, which has a significantly longer burn-in period, depending on the initialization. This efficiency is illustrated in Fig. 9 where the CLL chains for the two samplers are represented.

Using the efficient FRWMH sampler for the  $\boldsymbol{\theta}$  parameters, the overall sampling process takes roughly<sup>2</sup> 1 minute/model, thus the full run takes 8 minutes (the number of candidate models:  $K = 4$  shapes of  $\boldsymbol{\lambda}$  for the GRF type and  $K = 4$  shapes of  $\boldsymbol{\lambda}$  for the SMGRF).

### B. LMA and HMA

The posterior samples drawn by FRWMH within Gibbs sampling are used to compute the evidences. This is achieved via two approximations: the PM-based LMA and the HMA.

These approximations are computed using the same set of posterior samples in order to evaluate their accuracy in the same conditions. The numerical results show that the difference between the two evidence approximations is less than 0.1%, thus confirming that both approximations are viable for the problem in question.

Moreover, since the sampling is the most costly part of the evidence computation, the choice of approximation does not affect the overall speed performance. Consequently, the two evidence approximations imply a similar computational load.

In the tests presented in the following section, the CEAPS is based on the HMA.

### C. CEAPS Performances

Let us now present the performances of the selection method itself. The experimental setup consists in testing our method on synthetic textures, using 20 sets of parameter values for the PSD. Each set was used for each PSD model to generate both GRF and SMGRF texture realizations. The observations are obtained in a scenario with Gaussian blur of standard deviation  $w = 0.3$  and SNR = 20 dB. This corresponds to a partial overlap configuration, such as the one depicted in Fig. 8(b).

<sup>2</sup>Algorithms have been implemented using the computing environment *Matlab* on a Personal Computer, with a Intel Xeon 2 GHz and 1 GB of RAM.

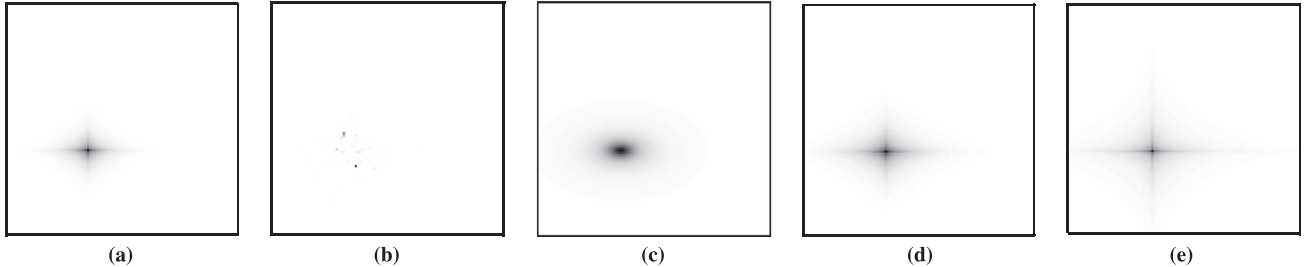


Fig. 10. Selection problem when the true model is not among the candidates. (a) True  $\lambda$ ; (b) Observed PSD; (c)  $\lambda$  candidate 1; (d)  $\lambda$  candidate 2; (e)  $\lambda$  candidate 3.

TABLE II  
CEAPS MODEL SELECTION PERFORMANCE FOR GRF TEXTURES (CORRECT CLASSIFICATIONS RATE IN %) FOR A PARTIAL OVERLAP CASE

True model	Estimated model			
	Lo	GL	Exp	GG
Lo	85	10	1	4
GL	21	69	3	7
La	2	4	87	7
GG	4	8	16	72

TABLE III  
CEAPS MODEL SELECTION PERFORMANCE FOR SMGRF TEXTURES (CORRECT CLASSIFICATIONS RATE IN %) FOR A PARTIAL OVERLAP CASE

True model	Estimated model			
	Lo	GL	Exp	GG
Lo	87	9	1	3
GL	19	73	2	6
La	1	7	83	9
GG	3	6	20	71

1) *CEAPS*: The algorithm was run on each texture realization and Tables II and III summarize the classification results for GRF and SMGRF textures, respectively. We observe on the main diagonal of both tables the percentages of correct classifications. As expected, the CEAPS chooses the correct PSD model in most cases. There are, however, situations where the method chooses another model.

As anticipated by the Fisher information analysis, in the majority of cases, there is enough available information on the central frequencies to ensure the samples are distributed under the correct law and have a finite variance. Nevertheless, the information concerning the widths is more sensitive to the noise level and the PSD model and thus more prone to errors. These errors are important for the method behavior, since they trigger misclassifications. The majority of missclassification cases are due to high noise levels and consist in mistakenly considering a PSD with thicker tails as the most adequate model. In this situation, the thicker tails account for the noise and the noise level is underestimated.

Nevertheless, in the context of our model choice problem, where the nested models help testing the method's ability to penalize model complexity, choosing another model can be regarded as not necessarily a failure. In this setting, the underlined percentages from Tables II and III represent the “good” miss-classifications, for instance, a Generalized Gaussian with  $q = 1$  that is classified as an exponential. This illustrates the

method's capacity to penalize model dimension, i.e., eliminate the parameters that do not significantly increase the model fit.

The method is not only able to distinguish between the different PSD forms, but also between the laws for the Fourier coefficients. More specifically, in 82% of the cases the algorithm correctly determined if the texture was from the GRF or the SMGRF class. This means that, on the one hand, the method has the ability to discriminate among a GRF and a SMGRF having the same form for the parametric part of the PSD. On the other hand, the PSD models themselves are structured enough to allow the algorithm to simultaneously identify the PSD model and whether all the PSD coefficients are identically scaled or not.

In the case where the real model that generated the data is not among the candidates, for each candidate model  $k = 1 \dots K$ , the chain of samples for  $\theta_k$  converges to the values that make model  $k$  best resemble the true model. Then, the evidences are computed based on these estimates and the model with the MAP probability is selected.

For instance, if the true model was an SMGRF with a certain shape for  $\lambda$  and the candidate models are only GRFs with various shapes for  $\lambda$ , an important remark is that the estimated widths for  $\lambda$  will be overestimated (i.e.,  $\sigma_x$  and  $\sigma_y$  will be underestimated), since the new  $\lambda$  will have to fit a wider range of frequencies (due to the presence of the auxiliary variables). Then, the method will choose the shape of  $\lambda$  that best fits the frequency components of the test image. Such a case is depicted in Fig. 10, where we show the true  $\lambda$ , the PSD of the observed texture (product among the true  $\lambda$  and auxiliary variables) and the PSDs of some of the candidate models (PSD is equivalent to  $\lambda$  in this case since the candidate models are GRFs).

2) *CEAPS vs GMLE*: A crucial point is that the GMLE cannot solve the problem of interest. This is due to the presence of indirect data introduced by the:

- non-Gaussian texture model,
- blurred and noisy observations.

Although the comparison cannot be performed on our problem, it is done on a simplified version of the problem dealing with direct observations (no noise, no convolution) of GRF textures.

Table IV lists the average classification success rate for the CEAPS and the GMLE, every method being tested on 20 texture realizations, with various parameter values for each type of PSD shape. More specifically, this represents an averaging over the PSD models and PSD parameters. The lower classification performance of the GMLE is due to the fact that it does not have any mechanism of model complexity penalization, thus it

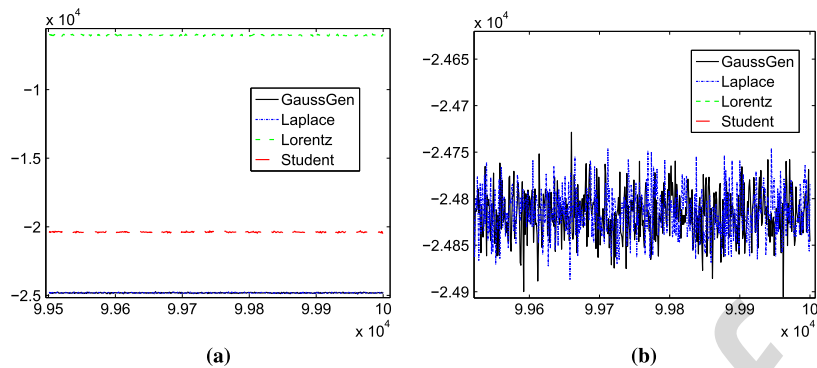


Fig. 11. CLL chains to illustrate a typical situation where GMLE fails to select the good model. (a) CLL chains for candidate models; (b) zoom on the CLL chains for the most likely models.

TABLE IV  
AVERAGE MODEL SELECTION PERFORMANCE (CLASSIFICATION SUCCESS RATE IN %) COMPARISON BETWEEN CEAPS AND GMLE. THE AVERAGING IS DONE OVER PSD MODEL AND PSD PARAMETERS

Algorithm	Classification accuracy (%)
CEAPS	89
GMLE	86

chooses in 100% of the cases the most complex among the embedded models. On the contrary, as previously explained, the CEAPS penalizes model dimension and selects the less complex model that fits the data.

Table IV shows that the GMLE has a lower success rate and the reason for this is illustrated in Fig. 11. In this figure, we plotted a case where GMLE selects the Generalized Gaussian model, since its minimum CLL is the global CLL minimum among all models. However, the minimum co-log PM is that of the Laplacian model, which is indeed the true model. This is a typical failure situation for GMLE.

As already stated in Section II, the evidence based classifier is optimal from the risk point of view, property that can be seen in this table through the CEAPS performances.

#### D. Results and Fisher Information

The experiments show that high noise scenarios,  $\text{SNR} < 20$  dB, are challenging since the samples for the widths have a too strong variance (this variance is high when the Fisher information is low). Furthermore, these samples, used to compute the evidences, have a direct impact on the model selection process.

In practice, above a certain noise level, the method tends to favor the thicker tailed PSDs, by considering that these tails account for the noise. We have seen in the informational analysis, summarized in Fig. 7, that for high noise levels there is a smaller amount of information, thus more uncertainty in the estimation, which eventually triggers estimation errors for the PSD widths and even miss-classifications. More specifically, the noise level is undervaluated and either the PSD widths are overevaluated, or a model with thicker tails is selected.

#### E. Image Reconstructions

Using the samples employed to compute the evidences, we can also compute PM estimates for the texture parameters, the noise precision and the unknown image. Consequently,

as an additional result, our algorithm provides a PM estimate of the original image, conditionally on the selected model. Fig. 12 shows examples of the reconstruction. We can observe situations (Figs. 12(c), 12(f) and 12(i)) where CEAPS successfully restores the texture even if the observations are severely degraded. This illustrates the method's high capacity to handle the blur and the noise. This is due to the strength of the information given by the structure of the PSD and to the method's optimality from the classification and estimation risk point of view. Nevertheless, there are also situations, such as Fig. 12(l), where the image is degraded to an extent that impairs a reconstruction, in most cases, this being due to a low information scenario.

## VII. CONCLUSION AND PERSPECTIVES

This paper presents a method for texture model choice from blurred and noisy observations. The textured images are modeled by Scale Mixture of Gaussian Fields with parametric power spectral density and parametric probability density for the scale variables. In a Bayesian framework, we are able to determine the posterior model probabilities based on the evidences, this approach being optimal from the classification risk point of view.

The employed within-model simulation technique consists in a sweep of all possible models and the computation of the evidence for each model. This quantity can be determined only by numerical methods, since the required integral is intractable. We have compared two methods for numerically computing the evidence based on samples from the *a posteriori* law, the Laplace Metropolis Approximation and the Harmonic Mean Approximation, which yield the same results in the context of our problem. We have presented the performances of our Classifier based on Evidence Approximation from Posterior Samples, which is the optimal model choice strategy from the mean classification error point of view. This is reflected in the classification results that show the method's ability to select the true model.

As a secondary result, this approach provides chains of samples for the parameters, conditionally on each model  $\mathcal{M} = k$ . These samples can be used to obtain estimates that are optimal from the mean square error point of view, by using the Posterior Mean estimator. Moreover, using these estimates and the selected model, the original image can be reconstructed.



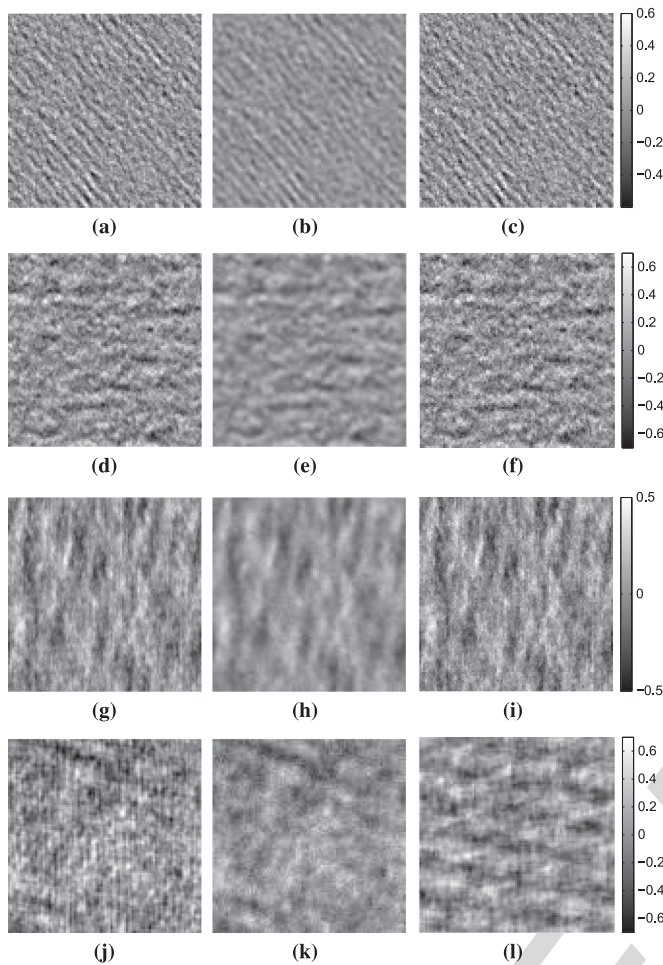


Fig. 12. Reconstruction results—4 cases to be read from left to right, on the first column the original, unobserved texture,  $\mathbf{x}|\theta^*$ ,  $\mathcal{M} = k^*$ , in the center the distorted observations,  $\mathbf{y}$ , and on the right column the results of the deconvolution, for the selected model,  $\hat{\mathbf{x}}|\hat{\theta}$ ,  $\mathcal{M} = \hat{k}$ .

Further developments include, but are not limited to, extending the SMGRF texture model to dependent Fourier coefficients, which would make it more versatile. Another idea is the use of multi-modal power spectral densities, in order to obtain more structured or quasi-periodic textures. From another standpoint, the method can be coupled with a beforehand learning step to enrich the set of possible texture model. From a different perspective, the method can be adapted to deal with unknown point spread functions (blind or semi-blind approach) and the estimation of their parameters [52]–[54].

#### REFERENCES

- [1] G. Kail, J.-Y. Tourneret, F. Hlawatsch, and N. Dobigeon, "Blind deconvolution of sparse pulse sequences under a minimum distance constraint: A partially collapsed Gibbs sampler method," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2727–2743, Jun. 2012.
- [2] L.-Y. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," in *Proc. 27th Conf. Comput. Graph. Interact. Tech.*, 2000, pp. 479–488.
- [3] F. Zhou, J.-F. Feng, and Q.-Y. Shi, "Texture feature based on local Fourier transform," in *Proc. Int. Conf. Image Process.*, 2001, vol. 2, pp. 610–613.
- [4] G.-S. Xia, S. Ferradans, G. Peyré, and J.-F. Aujol, "Synthesizing and mixing stationary Gaussian texture models," *SIAM J. Imag. Sci.*, vol. 8, no. 1, pp. 476–508, 2014.

- [5] M. N. Do and M. Vetterli, "Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models," *IEEE Trans. Multimedia*, vol. 4, no. 4, pp. 517–527, 2002.
- [6] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 49–71, 2000.
- [7] R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-33, pp. 959–963, Aug. 1985.
- [8] R. Chellappa and R. Kashyap, "Texture synthesis using 2-d noncausal autoregressive models," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 1, pp. 194–203, 1985.
- [9] S. C. Zhu, Y. Wu, and D. Mumford, "Filters, random fields and maximum entropy (frame)—Towards a unified theory for texture modeling," *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 1–20, 1998.
- [10] G. L. Gimelfarb, *Image Textures and Gibbs Random Fields*. Boston, MA, USA: Kluwer Academic, 1999, vol. 16.
- [11] E. P. Simoncelli, "Statistical modeling of photographic images," in *Handbook of Video and Image Processing*. New York, NY, USA: Academic, 2005.
- [12] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [13] S. D. Spiegelhalter, N. G. Best, B. P. Carlin, and A. V. D. Linde, "Bayesian measures of model complexity and fit," *J. Royal Statist. Soc.: Ser. B (Statist. Methodol.)*, vol. 64, no. 4, pp. 583–639, 2002.
- [14] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.
- [15] M. J. Beal, "Variational algorithms for approximate bayesian inference," Ph.D. dissertation, Univ. of Cambridge, Cambridge, U.K., 2003.
- [16] A. E. Raftery, "Hypothesis testing and model selection via posterior simulation," in *Practical Markov Chain Monte Carlo*, W. R. Gilks, D. J. Spiegelhalter, and S. Richardson, Eds. London, U.K.: Chapman & Hall, 1995.
- [17] J. Rosenthal, "Optimal proposal distributions and adaptive MCMC," in *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, Eds. Boca Raton, FL, USA: CRC, 2011, pp. 93–112.
- [18] B. Cai, R. Meyer, and F. Perron, "Metropolis-Hastings algorithms with adaptive proposals," *Statist. Comput.*, vol. 18, no. 4, pp. 421–433, 2008.
- [19] G. Roberts and O. Stramer, "Langevin diffusions and Metropolis-Hastings algorithms," *Methodol. Comput. Appl. Probabil.*, vol. 4, pp. 337–358, 2003.
- [20] M. Girolami and B. Calderhead, "Riemannian manifold Hamiltonian Monte Carlo (with discussion)," *J. Roy. Statist. Soc. B (Statist. Methodol.)*, vol. 73, pp. 123–214, 2011.
- [21] R. M. Neal, "MCMC using Hamiltonian dynamics," in *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, Eds. Boca Raton, FL, USA: CRC, 2011, pp. 93–112.
- [22] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, no. 2, pp. 251–276, 1998.
- [23] Y. Qi and T. P. Minka, "Hessian-based Markov Chain Monte-Carlo algorithms," presented at the 1st Cape Cod Workshop Monte Carlo Methods, Cape Cod, MA, USA, Sep. 2002.
- [24] **[AU: Provide more information on source/where to find--ed.]** T. Bui-Thanh and O. Ghattas, "Scaled stochastic Newton algorithm for Markov chain Monte Carlo simulations," 2012.
- [25] J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas, "A stochastic newton MCMC method for large-scale statistical inverse problems with application to seismic inversion," *SIAM J. Scientif. Comput.*, vol. 34, no. 3, pp. A1460–A1487, 2012.
- [26] C. Vacar, J.-F. Giovannelli, and Y. Berthoumiou, "Langevin and Hessian with Fisher approximation stochastic sampling for parameter estimation of structured covariance," in *Proc. IEEE ICASSP*, Prague, 2011, pp. 3964–3967.
- [27] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, ser. Springer Texts in Statist. New York, NY, USA: Springer-Verlag, 2007.
- [28] C. Vacar, J.-F. Giovannelli, and A.-M. Roman, "Bayesian texture model selection using harmonic mean," presented at the IEEE ICIP, Orlando, FL, USA, 2012.
- [29] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, 1995.

- [30] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, 1992.
- [31] J. F. Giovannelli, "Unsupervised Bayesian convex deconvolution based on a field with an explicit partition function," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 16–26, 2008.
- [32] J. M. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: A gem algorithm exploiting a class of heavy-tailed priors," *IEEE Trans. Image Process.*, vol. 15, no. 4, pp. 937–951, 2006.
- [33] Y. Zhang and N. Kingsbury, "Image deconvolution using a Gaussian scale mixtures model to approximate the wavelet sparseness constraint," in *Proc. IEEE ICASSP*, 2009, pp. 681–684.
- [34] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Appl. Computat. Harmon. Anal.*, vol. 11, pp. 89–123, 2001.
- [35] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, 2003.
- [36] D. K. Hammond and E. P. Simoncelli, "Image modeling and denoising with orientation-adapted Gaussian scale mixtures," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2089–2101, 2008.
- [37] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. London, U.K.: Chapman & Hall, 2005, vol. 104, Monographs on Statist. Appl. Probabil..
- [38] X. Guyon, *Random Fields on a Network: Modelling, Statistics, and Applications, Probability and Its Applications*. New-York, USA: Springer-Verlag, 1995.
- [39] A. Oppenheim and A. S. Willsky, *Signals and Systems*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983.
- [40] B. Picinbono, *Principles of Signals and Systems: Deterministic Signals*. Norwood, MA, USA: Artech House, 1988.
- [41] J. Skilling, "Nested sampling for general Bayesian computation," *Bayesian Anal.*, vol. 1, no. 4, 2006.
- [42] P. J. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.
- [43] A. E. Gelfand and D. K. Dey, "Bayesian model choice: Asymptotics and exact calculations," *J. Roy. Statist. Soc. B*, vol. 56, no. 3, pp. 501–514, 1994.
- [44] M. A. Newton and A. E. Raftery, "Approximate Bayesian inference with the weighted likelihood bootstrap," *J. Roy. Stat. Soc. B*, vol. 56, no. 1, pp. 3–48, 1994.
- [45] **[AU: More info on source and where to find--ed.]** C. P. Robert and D. Wraith, *Computat. Methods for Bayesian Model Choice 2009*.
- [46] A. G. Pakes, "On the convergence of moments of geometric and harmonic means," *Statistica Neerlandica*, vol. 53, no. 1, pp. 96–110, 1999.
- [47] A. E. Raftery, M. A. Newton, J. M. Satagopan, and P. N. Krivitsky, "Estimating the integrated likelihood via posterior simulation using the harmonic mean identity," *Bayesian Statist.*, pp. 1–45, 2007.
- [48] R. E. Kass and A. E. Raftery, "Bayes factors," *J. Amer. Statist. Assoc.*, vol. 90, pp. 773–795, 1995.
- [49] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, NY, USA: Springer-Verlag, 2004.
- [50] L. Tierney, "Markov chain for exploring posterior distribution," *Ann. Statist.*, vol. 22, no. 4, pp. 1701–1762, Dec. 1994.
- [51] C. P. Robert and G. Casella, *Monte-Carlo Statistical Methods*, ser. Springer Texts in Statistics. New York, NY, USA: Springer-Verlag, 2004.
- [52] F. Orieux, J. F. Giovannelli, and T. Rodet, "Bayesian estimation of regularization and point spread function parameters for Wiener-Hunt deconvolution," *J. Opt. Soc. Amer. A*, vol. 27, no. 7, pp. 1593–1607, 2010.
- [53] S.-U. Park, N. Dobigeon, and A. O. Hero, "Semi-blind sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3838–3849, 2012.
- [54] C. Vacar, J.-F. Giovannelli, and Y. Berthoumieu, "Bayesian texture and instrument parameter estimation from blurred and noisy images using MCMC," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 707–711, 2014.



**Cornelia Vacar** received the B.Sc. degree in telecommunications from the Technical University of Cluj-Napoca in 2009. She received the M.Sc. and Ph.D. degrees in image processing from the University of Bordeaux, France, in 2010 and 2014, respectively.

Her research interests focus on Bayesian methods for image deconvolution, texture classification, and image segmentation. She is currently activating within FEI Visualization Sciences Group.



**Jean-François Giovannelli** was born in Béziers, France, in 1966. He graduated from the École Nationale Supérieure de l'Électronique et de ses Applications, Cergy, France, in 1990. He received the Ph.D. and H.D.R. degree in signal–image processing at the Université Paris-Sud, Orsay, France, in 1995 and 2005, respectively.

From 1997 to 2008, he was an Assistant Professor with the Université Paris-Sud and a Researcher with the Laboratoire des Signaux et Systèmes, Groupe Problèmes Inverses. He is currently a Professor with the Université de Bordeaux, France, and a Researcher with the Laboratoire de l'Intégration du Matériau au Système, Groupe Signal-Image. His research focuses on inverse problems in signal and image processing, mainly unsupervised and myopic problems. From a methodological stand point the developed regularization methods are both deterministic (penalty, constraints, etc.) and Bayesian. Regarding numerical algorithms, the work relies on optimization and stochastic sampling. His application fields essentially concern astronomical, medical, proteomics, and geophysical imaging.



**Yannick Berthoumieu** received the Ph.D. degree in January 1996 in signal processing from the University of Bordeaux I, France.

In 1998, he joined the Ecole Nationale Supérieure d'Électronique, Informatique et Radiocommunications de Bordeaux, as an Associate Professor. He is currently a Full Professor with the Department of Telecommunications, Bordeaux Institute of Technology. His major research interests are concerned with multidimensional signal analysis, stochastic modeling, texture analysis, and video processing.

Since 2003, he has been with the Signal and Image Processing Group, IMS Lab, University of Bordeaux.

# Bayesian Texture Classification From Indirect Observations Using Fast Sampling

Cornelia Vacar, Jean-François Giovannelli, and Yannick Berthoumieu

**Abstract**—A Bayesian method for texture model choice from blurred and noisy (i.e., indirect) observations is presented. The textures are modeled by stationary Random Fields, with various distribution laws, either Gaussian or Scale Mixtures of Gaussians. The power spectral densities of the fields are modeled by parametric functions and the aim is to select the most appropriate model among a set of candidates. This is achieved by computing the *a posteriori* model probabilities through parameter marginalization. The marginalization is done by sampling and harmonic mean approach, considering separately each model, in a within-model sampling strategy. The highly nonlinear dependency with respect to the parameters imposes the use of the Metropolis-Hastings sampler. Moreover, to achieve efficient sampling, the paper proposes a new fast algorithm based on the Fisher information matrix, the Fisher Metropolis-Hastings.

**Index Terms**—Bayes, harmonic mean, Metropolis-Hastings, model choice, texture modeling and analysis.

## I. TEXTURE AND MODEL CHOICE

THE central problem in this work is selecting a texture model from a blurred and noisy image. It is solved in a Bayesian framework, where the posterior probability for each model is determined from the model evidences. The classification method is optimal from the risk point of view, however, these evidences are intractable and thus are numerically computed by Monte Carlo Markov Chain (MCMC) methods.

Texture represents a central aspect in image processing. A desirable feature of a texture model is to provide a large capability of description, while being relatively easy to handle and numerically efficient. The literature devoted to texture modeling is fairly rich, covering various classes of models that can be analyzed from different points of view. The representation domain can be spatial [1], [2], Fourier [3], [4], wavelet [5], [6] and the model can be deterministic (pseudo-periodic, based on structural elements), or probabilistic (stochastic-like), as for instance the random field-based model [4], [7]–[11].

In this work, the textured images are modeled by Random Fields (RF), with different parametric forms for the pixel inter-

action and probability distributions. This type of model is very tractable, equally adapted to synthesis and analysis tasks and known to exhibit good performances for stochastic-like texture processing. Nevertheless, it can also be used in the case of more deterministic or pseudo-periodic textures, through specific pixel interactions. From the texture modeling point of view, the contribution of this paper is to introduce a new non-Gaussian model, fully adapted to analysis and synthesis tasks. The particularities of this model are the following:

- Independent Fourier coefficients for the textured image. These coefficients are marginally non-Gaussian, but conditionally Gaussian, to ensure model tractability. This is achieved using Scale Mixtures of GRFs (SMGRF).
- Formulation in the Fourier domain in order to exploit the multiplicative form of the blur in the frequency domain.

Model choice has applications in a wide range of fields, for instance microbiology, proteomics, genomics, economics, statistics, signal and image analysis. The Generalized Maximum Likelihood Estimator (GMLE) method for model selection consists in determining the model with the highest likelihood for the MLEs of the parameters. However, this estimator is not applicable for indirect observations or the non-Gaussian texture model. Among the most frequently used model selection methods are the well-known Akaike Information Criterion (AIC) [12], Deviance Information Criterion (DIC) [13] and the Bayesian Information Criterion (BIC) [14]. Nevertheless, all these criteria are based on two levels of approximations. First, a second order Taylor approximation of the likelihood around its maximum is done in order to derive the form of the criteria. On a second level, except for very simple cases, the maximum likelihood value cannot be computed analytically and, thus, numerical methods (such as optimization) are employed for its computation. In this context, the evidence based methods represent an alternative Bayesian approach [15] that does not rely on an approximation at the first level. Nevertheless, since the evidences are intractable, numerical methods are employed for their computation, this corresponding to an approximation on the aforementioned second level, the numerical one. The evidence-based classification can take the form of Bayes factors or the posterior model probabilities computation method presented in this work.

The latter relies on the formulation of an optimal decision function from the risk viewpoint. Practically, it consists in computing the *a posteriori* model probabilities, based on (i) the joint law for the model, the data, the model parameters and on (ii) computing the evidence for each model. This computation is done by marginalizing the parameters from the joint law. Since

Manuscript received April 30, 2014; revised August 27, 2014, June 20, 2015, and August 26, 2015; accepted August 26, 2015. Date of publication September 17, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Animashree Anandkumar.

The authors are with IMS (Univ. Bordeaux, CNRS, B-INP), UMR 5218, F-33400 Talence, France.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2015.2480040



this law is complicated, the integral is intractable and it is approximated either using MCMC sampling, or the Laplace approximation [16].

In the context of evidence computation using MCMC, some of the parameters have very complicated laws and their sampling can be costly, thus requiring the use of efficient solutions. The recent literature regarding efficient sampling is mainly focused on two types of algorithms: optimal tuning of the standard Metropolis Hastings (MH) samplers [17], or adapted proposition laws [18]. The adapted proposals exploit the information encoded in the target law in order to achieve an efficient exploration. They can be based on first order derivatives of the target, this being the case of the Metropolis Adjusted Langevin Algorithm (MALA) [19], [20] and of the Hamiltonian methods [21]. Another class of algorithms takes advantage of the target curvature through Hessian-based (Newton-like) proposals. Such terms have first been employed in optimization algorithms [22], but they have recently been successfully adapted to sampling [23]–[25]. From this stand point, our contribution is algorithmic and consists in integrating an improved Newton-like proposal [26] using the Fisher matrix instead of the inverse Hessian, i.e., the Fisher Random Walk MH (FRWMH), presented in [26]. In this manner, an important performance increase is achieved, due to the specific nature of our problem: there is no need for computing second order derivatives and thus a Newton type proposal is built using only first order derivatives.

To summarize, the main contributions are: (i) addressing the problem of texture model choice from indirect observations, (ii) tackling this problem in an optimal manner, (iii) introducing a new non-Gaussian model for the texture, tractable and efficient and (iv) enhancing the speed performances of the MCMC sampling phase, by using the FRWMH algorithm.

The paper is structured as follows: Section II presents the problem and the evidence-based model choice, Section III details the texture models and Section IV gives a Fisher information analysis of the problem. Section V is devoted to evidence computation and sampling, while Section VI presents the results. Section VII concludes the paper with comments on the method and perspectives for future work.

## II. PROBLEM STATEMENT AND PRELIMINARIES

Let us now present the method to choose the texture model, say  $\mathcal{M} = k$ , among  $K$  possible models. This is carried out starting from indirect observations, i.e., blurred and noisy data. The observation model is:

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

where  $\mathbf{y}$ ,  $\mathbf{x}$  and  $\mathbf{n} \in \mathbb{C}$  are the lexicographically ordered observations, unobserved textured image and noise, respectively. The images are of size  $N \times N$  and  $P = N^2$  denotes the number of pixels. The point spread function (PSF) is represented by the convolution matrix  $H$ , of size  $P \times P$ . The Fourier transform of the PSF is referred to as the Transfer Function (TF) and its coefficients are denoted  $\hat{h}_p$ .

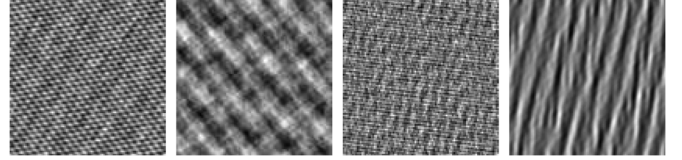


Fig. 1. Texture realizations of the employed image model.

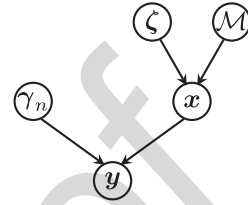


Fig. 2. Observation model graph and variable dependency.

### A. Distributions

The noise  $\mathbf{n}$  is zero-mean stationary Gaussian, with covariance matrix  $R_n(\boldsymbol{\gamma})$ , driven by the unknown parameter set  $\boldsymbol{\gamma}$ . From this model, the conditional law for  $\mathbf{y}$  writes:

$$f(\mathbf{y}|\mathbf{x}, \boldsymbol{\gamma}) \propto \det [R_n(\boldsymbol{\gamma})]^{-1} \exp \left[ -\|\mathbf{y} - H\mathbf{x}\|_{R_n(\boldsymbol{\gamma})}^2 \right] \quad (2)$$

The method is equally adapted to any type of noise correlation, however, our numerical study focuses on the white noise case, i.e.,  $R_n(\boldsymbol{\gamma}) = \gamma_n^{-1} \mathbf{I}$ , with  $\gamma_n$  the precision (inverse variance).

The texture models are parametric and driven by the parameter set  $\boldsymbol{\zeta}$ . Thus, for each model,  $k$ , we have the corresponding law  $f(\mathbf{x}|\boldsymbol{\zeta}_k, \mathcal{M} = k)$  and its associated parameter set  $\boldsymbol{\zeta}_k$ . The exact form of these laws and the structure of the parameter sets will be specified in Section III. Typical realizations of such textures are shown in Fig. 1.

In this context, our model selection method relies on a Bayesian hierarchical framework and Fig. 2 illustrates the variable interdependency. We are considering a case where the prior information about the parameters is very reduced, consequently, uninformative priors [27] are used. For  $\gamma_n$ , the conjugate form with respect to the likelihood (2) is a Gamma law:

$$\pi(\gamma_n) = \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \gamma_n^{\alpha_n - 1} \exp[-\beta_n \gamma_n] = \mathcal{G}(\alpha_n, \beta_n) \quad (3)$$

This law becomes an uninformative Jeffreys prior in the limit case ( $\alpha_n \rightarrow 0, \beta_n \rightarrow 0$ ).

The *a priori* distribution for the model is fully described by the  $p_k = \Pr(\mathcal{M} = k)$  probabilities. Our numerical study relies on an uninformative prior, i.e., equiprobable models:  $p_k = 1/K, k = 1 \dots K$ .

The priors for the texture parameters  $\boldsymbol{\zeta}_k$  will be defined in accordance with the employed texture models, in Section III.

### B. Decision Function

A decision function, denoted  $\Delta$ , associates a model  $k$  to the data  $\mathbf{y}$ , i.e.,  $\Delta(\mathbf{y}) = k$ . Moreover, a cost function quantifies the decision error. Let this function be  $C(k, k^*)$ , where  $k$  is the chosen model and  $k^*$  represents the true model. This cost has

the properties  $C(k, k^*) \geq 0$  and  $C(k^*, k^*) < C(k, k^*)$ , when  $k \neq k^*$ . Based on the cost, the risk is defined as:

$$\rho(\Delta) = \mathbb{E}_{\mathcal{M}, \mathbf{y}} [C(\Delta(\mathbf{y}), k)]$$

and the optimal decision function is  $\Delta_c = \arg \min_{\Delta} \rho(\Delta)$ . The optimality is achieved because the cost is averaged over all possible data, models and parameter values. Moreover, for a binary cost function  $C(k, k^*) = 1 - \delta(k, k^*)$ , the method implicitly selects the model with the Maximum *a Posteriori* (MAP) probability:

$$\hat{\mathcal{M}} = \arg \max_k \Pr(\mathcal{M} = k | \mathbf{y}) \quad (4)$$

Using Bayes' rule, the posterior model probabilities are:

$$\Pr(\mathcal{M} = k | \mathbf{y}) = \frac{f(\mathbf{y} | \mathcal{M} = k) \cdot p_k}{f(\mathbf{y})} \quad (5)$$

and require the two following quantities.

- 1) The probability distribution of the data,  $f(\mathbf{y})$ . Fortunately, it does not depend on the model, thus can be calculated by normalization.
- 2) The **evidence**/marginal likelihood,  $e_k = f(\mathbf{y} | \mathcal{M} = k)$ , obtained from the joint law of data and unknowns, given the model, by marginalizing the unobserved texture, the noise and the texture parameters:

$$e_k = \int_{\Psi} f(\mathbf{y}, \Psi | \mathcal{M} = k) d\Psi \quad (6)$$

where  $\Psi = \{\gamma_n, \mathbf{x}, \zeta_k\}$  gathers all the unknowns.

This law is written using the conditioning rule and the hierarchy shown in Fig. 2:

$$f(\mathbf{y}, \Psi | \mathcal{M} = k) = f(\mathbf{y} | \mathbf{x}, \gamma_n) \cdot f(\mathbf{x} | \zeta_k, \mathcal{M} = k) \cdot \pi(\zeta_k | \mathcal{M} = k) \cdot \pi(\gamma_n) \quad (7)$$

It now becomes clear that (6) is intractable due to the numerous layers of non-linearity and must be computed numerically. In this work, this is done by an MCMC algorithm.

A similar problem has been addressed in our previous work [28], in the context of direct observations,  $\mathbf{y} = \mathbf{x}$ , and for Gaussian textures. In that case, (6) reduced to a single integral with respect to the texture parameters.

### III. TEXTURE MODELING

From a theoretical standpoint, the proposed method can handle any texture model driven by the parameter set  $\zeta$ . In this context, the Gaussian RFs (GRF) model is simple and tractable. Its extension to a scale mixture of Gaussians (SMG) expands the representation capabilities without particularly affecting the algorithm efficiency. This model is based on a set of auxiliary variables  $\mathbf{s}$  such that, conditionally on these variables, the field  $\mathbf{x} | \mathbf{s}$  is Gaussian, but marginally  $\mathbf{x}$  is no longer Gaussian.

This idea, originating from statistics, has been exploited in problems of deconvolution and denoising [29]–[33] to formulate heavier tailed regularization terms and for modeling natural images [34]–[36]. The SMGs have been used so far to model the sparse character of the differences between neighboring pixels

(in the spatial domain) and of the wavelet coefficients (in the wavelet domain).

This is where our work is different. Here, we take into account that the blur introduced by the observation model is easily written as a multiplication in the Fourier domain. Moreover, the Fourier domain is well adapted for representing textural characteristics. Consequently, our SMG model is defined in the Fourier domain.

Hence the originality since, although the SMG model has been previously employed in image analysis, to the best of our knowledge, it has not been used so far to model the Fourier coefficients of a textured image.

For the sake of computational efficiency, we have made the assumption that the textures are zero-mean and stationary, thus the covariance matrix has a Toeplitz-block-Toeplitz structure. For general models,<sup>1</sup> the exact likelihood cannot be computed due to the huge dimension of the covariance to be inverted. But, by Whittle's approximation, this matrix has a Circulant-block-Circulant form, therefore is diagonalizable by discrete Fourier transform. The reader is invited to refer to [37], [38] for approximation properties and asymptotic behavior. Consequently, the Fourier coefficients, denoted by  $\hat{x}_p$ ,  $p = 1 \dots P$ , are decorrelated, conditionally on the power spectral density (PSD). The conditional law of the image is separable in the Fourier domain and writes:

$$f(\mathbf{x} | \zeta) \propto \prod_{p=1}^P \zeta_p \cdot \exp \left[ - \sum_{p=1}^P \zeta_p |\hat{x}_p|^2 \right] \quad (8)$$

Each Fourier coefficient follows a zero-mean Gaussian distribution, of variance given by the corresponding PSD element:

$$\hat{x}_p | \zeta_p \sim \mathcal{N}(0, \zeta_p^{-1}) \quad (9)$$

Due to the fact that the law for  $\mathbf{x} | \zeta$  is Gaussian, the Fourier coefficients are independent, conditionally on  $\zeta$ .

The model complexity and thus its representation capabilities can be enhanced, while keeping the limitation to independent Fourier coefficients, by changing i) the law of the Fourier coefficients and ii) the form of the PSD.

#### A. Fourier Coefficient Distributions

This work focuses on two models for the Fourier coefficient law, between which only the form of  $\zeta_p$  changes:

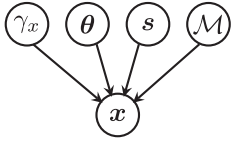
- Gaussian texture (GRF) -  $\zeta_p = \gamma_x \lambda_p(\boldsymbol{\theta})$
- non-Gaussian texture (SMGRF) -  $\zeta_p = \gamma_x s_p \lambda_p(\boldsymbol{\theta})$

where  $\lambda_p(\boldsymbol{\theta})$  are the shape elements of the parametric PSD, driven by the parameter set  $\boldsymbol{\theta}$ . We will denote by  $\boldsymbol{\lambda}$  the parametric component of the PSD, i.e., the collection of all  $\lambda_p(\boldsymbol{\theta})$ .  $\gamma_x$  is a global scale parameter for the PSD and  $s_p$  is a local scale parameter.

The scale parameter pdf determines whether the textures follow a GRF or a SMGRF model and these parameters allow us to switch between the two laws, without changing the parametric part of the PSD,  $\boldsymbol{\lambda}$ .

In this context, Fig. 3 completes the problem description from Fig. 2 by showing the variable hierarchy for the two types of texture models.

<sup>1</sup>This is not the case if the field is defined through its precision matrix (inverse covariance), e.g. for Markov field.



- Non-Gaussian texture:

$$s = \{s_p\}_{p=1\dots P}$$

$$s_p \sim \mathcal{G}(\alpha_s, \beta_s)$$

- Gaussian texture:

$$s_p = 1, p = 1\dots P$$

Fig. 3. SMGRF texture model.

1) *Gaussian Model*: In the GRF case, the scales are identical:  $s_p = s = 1$ , for  $p = 1 \dots P$ , thus (8) and (9) become:

$$f(\mathbf{x}|\gamma_x, \boldsymbol{\theta}, s) \propto \gamma_x^P \left[ \prod_{p=1}^P \lambda_p(\boldsymbol{\theta}) \right] \cdot \exp \left[ - \sum_{p=1}^P \gamma_x \lambda_p(\boldsymbol{\theta}) |\hat{x}_p|^2 \right] \hat{x}_p | \gamma_x, \boldsymbol{\theta} \sim \mathcal{N} \left( 0, [\gamma_x \lambda_p(\boldsymbol{\theta})]^{-1} \right) \quad (10)$$

This is the model used in our previous papers [26], [28] for fast parameter sampling and pixel interaction model choice.

2) *Non-Gaussian Model*: In this case, a prior is assigned to  $s_p$  auxiliary variables and they are included in the estimation framework. Theoretically, any prior with positive support can be used. Some immediate examples are Markov or Wishart. Nevertheless, for numerical efficiency, it is essential for this prior to allow for parallel posterior processing (i.e., the  $s_p$  should be separable *a posteriori*) and to have some conjugacy property with respect to (8). For this reason, the paper is limited to *a priori* independent auxiliary variables, following a Gamma distribution  $\mathcal{G}(\alpha_s, \beta_s)$ :

$$f(s_p | \alpha_s, \beta_s) \propto s_p^{\alpha_s - 1} \exp[-\beta_s s_p] \quad (11)$$

Due to this separability, although  $f(\mathbf{x}|\gamma_x, \boldsymbol{\theta})$  (the marginal law with respect to  $\mathbf{s}$ ) is non-Gaussian, it remains separable. Here it has a Student's t form, hence (8) becomes:

$$f(\mathbf{x}|\gamma_x, \boldsymbol{\theta}, \alpha_s, \beta_s) \propto \prod_{p=1}^P \left[ 1 + \frac{\gamma_x \lambda_p(\boldsymbol{\theta})}{\beta_s} |\hat{x}_p|^2 \right]^{-\alpha_s - 1} \quad (12)$$

3) *Variance*: To individually evaluate the impact of each feature ( $\boldsymbol{\lambda}$ , scale parameter, Fourier coefficient pdf), we chose to keep the same marginal variance of the Fourier coefficients for the two models. In this manner, it is possible to compare textures with the same  $\boldsymbol{\lambda}$ , but different pdfs, or reversely, with the same pdf, but different forms for  $\boldsymbol{\lambda}$ .

The variances of the Fourier coefficients, conditionally on  $\boldsymbol{\theta}$  and  $\gamma_x$ , are given by:

$$\text{var}_{SMGRF}[\hat{x}_p | \gamma_x, \boldsymbol{\theta}] = \frac{\beta_s}{2\alpha_s - 1} \cdot \frac{1}{\gamma_x \lambda_p(\boldsymbol{\theta})}$$

$$\text{var}_{GRF}[\hat{x}_p | \gamma_x, \boldsymbol{\theta}] = \frac{1}{\gamma_x \lambda_p(\boldsymbol{\theta})}$$

Thus, the equality of these variances imposes a constraint on the parameters of the prior for  $s_p$ :  $\beta_s = 2\alpha_s - 1$ , with  $\alpha_s > 0.5$ . Then, the Fourier coefficients for the two types of textures have the same  $\boldsymbol{\lambda}$ , but different pdfs.

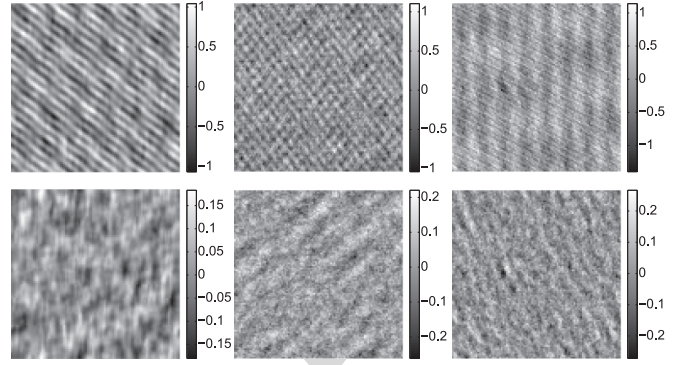


Fig. 4. Examples of texture pairs with the same PSD, but different law for the Fourier coefficients. On the first row, the SMGRF textures and on the second row the corresponding GRF textures.



Fig. 5. Examples of shapes for  $\boldsymbol{\lambda}$  that can be used with our GRF and SMGRF texture models. The representation is done in the Fourier domain on the reduced frequency domain  $[-0.5, 0.5] \times [-0.5, 0.5]$ .

Due to these common aspects, both models yield stochastic textures, however, the SMGRF is able to generate more complex patterns. Fig. 4 shows pairs of textures with the same  $\boldsymbol{\lambda}$ , but different pdfs. The  $\boldsymbol{\lambda}$  that were used to generate these textures have parametric forms, such as those depicted in Fig. 5, and will be analytically described in the next section (see Table I). The spectral content that is modeled using these forms consists in highlighting a cluster of spatially connected frequency components. For this reason, we can notice in Fig. 4 the strongly stochastic appearance of the GRF textures, with a dominant type of spectral components. The SMGRF textures present a spectrum (corresponding to the same parametric form of  $\boldsymbol{\lambda}$ ) with a more diverse range of frequencies and thus can represent more complex textures than the GRF. This is due to the presence of the auxiliary variables which increase the spectral complexity by increasing the amplitude of certain frequencies in the  $\boldsymbol{\lambda}$ . These supplementary components in addition to the grouped together high amplitude frequency components of the  $\boldsymbol{\lambda}$  will represent the new, richer PSD.

*Remark*: The SMGRF textures could also be obtained via the GRF model, however, for a non-parametric  $\boldsymbol{\lambda}$ . The  $\boldsymbol{\lambda}_{GRF} = \mathbf{s} \cdot \boldsymbol{\lambda}$ , which does not have a form that can be easily represented using a small number of parameters (is not compressible).

## B. Power Spectral Density Models

The form of the PSD actually encodes an important part of the textural content information. It must be stressed that there is virtually no constraint on the form of the PSD: it can be constant, corresponding to a white noise, or contain a single frequency. It can have a pulse-like form, this resulting in deterministic-like textures, or a more blobby shape, with one or several components, which would correspond to stochastic-like textures.

TABLE I  
EXAMPLES OF PSD MODELS USED IN THE SEQUEL OF THE PAPER

$\mathcal{M}$	Model $k$	Expression of $\lambda^k(\nu_x, \nu_y, \theta_k)$	$\theta_k$
1	Lorentz	$\left[1 + \frac{(\nu_x - \nu_{x_0})^2}{\sigma_x^2} + \frac{(\nu_y - \nu_{y_0})^2}{\sigma_y^2}\right]$	$\nu_{x_0}, \nu_{y_0}, \sigma_x, \sigma_y$
2	Generalized Lorentz	$\left[1 + \frac{(\nu_x - \nu_{x_0})^2}{\sigma_x^2} + \frac{(\nu_y - \nu_{y_0})^2}{\sigma_y^2}\right]^q$	$\nu_{x_0}, \nu_{y_0}, \sigma_x, \sigma_y, q$
3	Exponential	$\exp \frac{1}{2} \left[ \frac{ \nu_x - \nu_{x_0} }{\sigma_x} + \frac{ \nu_y - \nu_{y_0} }{\sigma_y} \right]$	$\nu_{x_0}, \nu_{y_0}, \sigma_x, \sigma_y$
4	Generalized Gauss	$\exp \frac{1}{2} \left[ \frac{ \nu_x - \nu_{x_0} ^q}{\sigma_x^q} + \frac{ \nu_y - \nu_{y_0} ^q}{\sigma_y^q} \right]$	$\nu_{x_0}, \nu_{y_0}, \sigma_x, \sigma_y, q$

From an algorithmic point of view, it is advantageous to use shapes encoded through a small number of parameters. For this reason, this paper focuses on parametric, unimodal functions, with a relatively reduced number of parameters: Lorentz ( $\mathcal{M} = 1$ ), Generalized Lorentz ( $\mathcal{M} = 2$ ), Exponential ( $\mathcal{M} = 3$ ) and Generalized Gaussian ( $\mathcal{M} = 4$ ). Each model  $k$  is driven by the corresponding  $\theta_k$  and Table I explicitly shows the  $\lambda^k(\theta_k)$ .

The parameters of these functions are the central frequencies  $\nu_{x_0}, \nu_{y_0}$ , the widths  $\sigma_x, \sigma_y$  and the positive shape parameter  $q$ , specific to models  $\mathcal{M} = 2$  and  $\mathcal{M} = 4$ .

We have chosen to include embedded models, for instance  $\mathcal{M} = 1$  and  $\mathcal{M} = 3$  are nested in  $\mathcal{M} = 2$  and  $\mathcal{M} = 4$ , respectively, for  $q = 1$ . This enables an analysis of the method capacity to penalize model complexity when the extra parameters do not trigger a significant model fit increase.

The textured images are spatially discrete, thus the PSD is defined on the reduced frequency domain [39], [40], i.e., the variables  $(\nu_x, \nu_y) \in [-0.5, 0.5]^2$ . Furthermore, let us consider that the Fourier coefficient  $p$  has the  $(\nu_m, \nu_n)$  position in the discrete reduced frequency domain. The  $\lambda_p$  from (10) and (12) are the elements of the PSD field at these discrete positions and depend on the model,  $\mathcal{M} = k$ , and on the parameters  $\theta_k$ . From this point forward, to explicitly show this dependency, the notation  $\lambda_p^k(\theta_k)$  is used and, more precisely:  $\lambda_p^k(\theta_k) = \lambda^k(\nu_m, \nu_n, \theta_k)$ .

The use of parametric models has the advantage of reducing the number of unknowns. On the other hand, this model defines a highly non-linear dependency of  $\lambda_p^k(\theta_k)$  with respect to  $\theta_k$ , as shown in Table I. This complicated dependency means that there is no conjugate form for this law.

Moreover, the prior information about the texture parameters  $\theta_k = \{\nu_{x_0}, \nu_{y_0}, \sigma_x, \sigma_y, q\}$  is very reduced, thus, uninformative priors will be used. Consequently, a uniform prior is employed:  $\pi(\theta_k | \mathcal{M} = k) = \mathcal{U}_{[\theta_k^m; \theta_k^M]}(\theta_k)$ , where  $\theta_k^m$  and  $\theta_k^M$  are the vectors containing the minima, respectively maxima, allowed values for the texture parameters. These values were chosen to ensure the coherence of the parameter values with their physical interpretation:  $\theta_k^m = \{-0.5, -0.5, 10^{-2}, 10^{-2}, 0.1\}$  and  $\theta_k^M = \{0.5, 0.5, 1, 1, 5\}$ .

This section devoted to the texture models has concluded the problem specification part of this paper. Based on the chosen observation model and the SMGRF image model, Section IV will present considerations regarding the problem difficulty and Section V will give the mathematical developments required to determine the most probable model.

#### IV. INFORMATION QUANTITATIVE ASSESSMENT

The data can be more or less informative and this is directly reflected in the problem difficulty. In order to evaluate this difficulty, this section presents a series of theoretical considerations regarding the available information, quantified by the Fisher information matrix.

Let us now focus on the information regarding any component of  $\Psi$ , denoted by  $\psi$ , by analyzing the diagonal elements of the information matrix, i.e., the expectation of the second derivative of the co-log-likelihood:

$$\mathcal{I}(\psi) = -\mathbb{E}_{\mathbf{y}|\Psi} \left[ \frac{\partial^2}{\partial \psi^2} \log f(\mathbf{y}|\Psi) \right] \quad (13)$$

Due to the separability and Gaussianity of the noise model and the form of (1), the law  $f(\mathbf{y}|\Psi)$  is also separable, Gaussian, zero-mean and of variance:

$$r_p(\Psi) = g_p [\gamma_x s_p \lambda_p(\theta)]^{-1} + \gamma_n^{-1} \quad (14)$$

where  $g_p = |\hat{h}_p|^2$ . Equation (13) contains first and second order derivatives of  $r_p(\Psi)$  with respect to  $\psi$ . When calculating the expectation, knowing that  $\mathbb{E}_{\mathbf{y}|\Psi} [|\hat{y}_p|^2] = r_p(\Psi)$ , the second order derivatives cancel out. Then:

$$\mathcal{I}(\psi) = \sum_{p=1}^P \left[ \frac{1}{r_p(\Psi)} \cdot r'_p(\Psi) \right]^2 \quad (15)$$

where  $r'_p(\Psi)$  is the derivative of  $r_p(\Psi)$  with respect to  $\psi$ .

For the noise parameter,

$$\mathcal{I}(\gamma_n) = \sum_{p=1}^P \left[ \gamma_n^{-1} + \gamma_x^{-1} \cdot g_p [s_p \lambda_p(\theta)]^{-1} \right]^{-2} \quad (16)$$

is a decreasing function, thus, the smaller the  $\gamma_n$ , i.e., the larger the noise level, the easier its estimation.

For a texture coefficient  $\theta$ , element of  $\theta$ ,

$$\mathcal{I}(\theta) = \sum_{p=1}^P \left[ \frac{g_p \cdot \lambda'_p(\theta)}{\lambda_p(\theta) [g_p + \gamma_x / \gamma_n \cdot s_p \lambda_p(\theta)]} \right]^2 \quad (17)$$

depends on the values of the other texture parameters. For instance, the information regarding the central frequency  $\nu_x^0$  is higher if the frequency  $\nu_y^0$  is close to 0 and decreases symmetrically with the absolute value of  $\nu_y^0$  (see Fig. 6(a)).  $\mathcal{I}(\nu^0)$  also increases when the width  $\sigma$  decreases, i.e., the characteristic is more concentrated, as in Fig. 6(b). Moreover, the amount of information available for estimating the texture parameters also varies according to the PSD model. Finally, concerning the signal and noise levels, the lower the SNR, the smaller  $\mathcal{I}(\theta)$  in (17) and, thus, the less information on  $\theta$ .

*Remark:* The higher information around the null frequency is due to the low pass character of the PSF. This becomes obvious in Fig. 6(d) where the amount of information for the high frequencies drops significantly when the filter becomes more selective (the inverse width is higher).

Another interesting case is the noiseless scenario ( $\gamma_n = \infty$ ):

$$\mathcal{I}(\theta) = \sum_{p=1}^P \left[ \frac{1}{\lambda_p(\theta)} \cdot \lambda'_p(\theta) \right]^2 \quad (18)$$



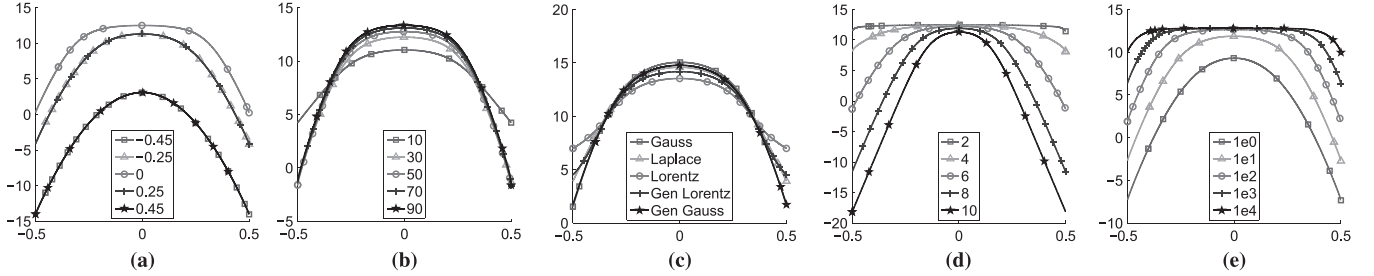


Fig. 6. Variation of the Fisher information for  $\nu_x^0$ , as a function of various factors. Only one parameter is varied at a time in order to quantify its effect on the amount of information for  $\nu_x^0$  (in logarithmic scale). (a)  $\nu_y^0$  sweep; (b)  $\sigma_x^{-1}$  sweep; (c) model sweep; (d) OTF inverse width; (e)  $\gamma_n$  sweep.

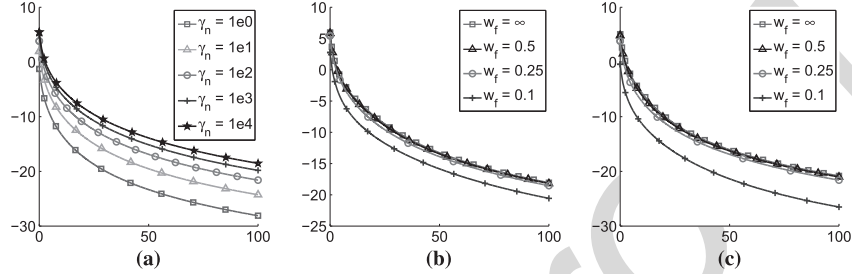


Fig. 7. Fisher information for  $w_x$  for various values of the observation system parameters  $w_f$  and  $\gamma_n$  (in logarithmic scale). (a)  $\gamma_n$  sweep; (b)  $w_f$  sweep-high  $\gamma_n$ ; (c)  $w_f$  sweep-low  $\gamma_n$ .

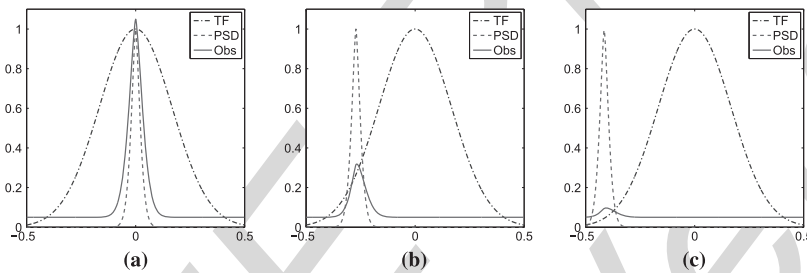


Fig. 8. Different cases emerging from the relative positioning and widths of the TF and PSD. (a) Total overlap; (b) partial overlap; (c) no overlap.

The Fisher information regarding  $\theta$  depends only on the PSD (as in [26]). This has multiple implications on the amount of information for the texture parameters:

- it does not depend on the form of the TF (known TF),
- it is the same, whether the texture is Gaussian or not.

Fig. 7 illustrates the amount of information concerning one width of the PSD for various levels of noise and widths of the PSF. In Fig. 7(a), the variation of the Fisher information with the noise precision shows a significant difference in the amount of information. Situations with SNR < 20 dB are challenging, since the available information is reduced. Figs. 7(b) and 7(c) illustrate the Fisher information variation for a PSD width as a function of the TF width  $w_f$ , the SNR being fixed at 30 dB in Fig. 7(b) and at 10 dB in Fig. 7(c). These plots show that the Fisher information does not strongly depend on the  $w_f$ , especially in the low noise case (high  $\gamma_n$ ) from Fig. 7(b).

Last, but not least, for the auxiliary variables  $s_p$ ,

$$\mathcal{I}(s_p) = \left\{ \frac{g_p}{[s_p(g_p + \gamma_x/\gamma_n \cdot s_p \lambda_p(\theta))]} \right\}^2 \quad (19)$$

the information depends on the corresponding coefficient of the PSD, TF and  $\gamma_n$ . Moreover, for the noiseless case,  $\mathcal{I}(s_p) = s_p^{-2}$ , i.e., the smaller the value, the more the information.

Fig. 8 illustrates various scenarios for a Gaussian TF and a Laplacian texture PSD. Their product plus the noise, of variance  $1/\gamma_n$ , yields the observation's PSD. The extreme cases are either simplifications:

- PSD centered near the null frequency and narrower than the TF, the effect of the convolution being negligible,
- noiseless case ( $\gamma_n \rightarrow \infty$ ),
- asd

or cases with signal alterations so severe that information on the original PSD is no longer present in the data:

- the PSD is positioned in a region with strong TF attenuation, the original image information being lost,
- high noise case ( $\gamma_n \rightarrow 0$ ). This corresponds to a low Signal to Noise Ratio (SNR)  $\gamma_n/\gamma_x$ .

This section offers a prior performance analysis based on the Fisher information matrix. The amount of available information is directly related to the estimation performances, allowing us to predict these performances in various situations.

## V. EVIDENCE CALCULATION

The full description of our Bayesian model choice method relies (i) on the data model and (ii) on the priors for the unknowns.

The hierarchical direct model is shown in Fig. 2, while the texture model specificities are shown in Fig. 3. In the following, the emphasis is on the SMGRF, which encompasses the GRF. By writing (6) as:

$$e_k = \int_{\mathbf{x}, \gamma_x, \gamma_n, \boldsymbol{\theta}_k, \mathbf{s}} f(\mathbf{y}|\boldsymbol{\Psi}, \mathcal{M} = k) \cdot \pi(\boldsymbol{\Psi}|\mathcal{M} = k) d\mathbf{s} d\boldsymbol{\theta}_k d\gamma_x d\gamma_n d\mathbf{x} \quad (20)$$

and plugging in (2) and (8), the intractability of the integral is obvious. For this reason, it must be calculated numerically and the solution chosen here is sampling.

A natural idea is to straightforwardly determine the evidence  $e_k$  from samples of the prior  $\pi(\boldsymbol{\Psi}|\mathcal{M} = k)$ :  $\boldsymbol{\Psi}^{(t)} = (\mathbf{x}^{(t)}, \gamma_n^{(t)}, \gamma_x^{(t)}, \mathbf{s}^{(t)}, \boldsymbol{\theta}_k^{(t)})$ ,  $t = 1 \dots T$  as follows:

$$\bar{e}_k \approx \frac{1}{T} \sum_{t=1}^T f(\mathbf{y}|\boldsymbol{\Psi}^{(t)}, \mathcal{M} = k) \quad (21)$$

It consists in sampling the priors and computing the arithmetic mean of the corresponding likelihood values—Arithmetic Mean Approximation (AMA). Evidence computation based on prior samples can also be done by nested sampling [41].

Nevertheless, when the likelihood is very peaked, as in the current case, most of these samples have weak likelihood, i.e., an insignificant contribution and thus, the algorithm is slow to converge. For this reason, it is more suitable to compute (20) by sampling the posterior  $\pi(\boldsymbol{\Psi}|\mathbf{y}, \mathcal{M} = k)$ .

#### A. Posterior Sampling

The posterior law is proportional to the joint law:

$$\begin{aligned} f(\mathbf{y}, \boldsymbol{\Psi}, \mathcal{M} = k) &= \mathcal{C} \cdot \exp \left[ -\gamma_n \sum_{p=1}^P |y_p - \overset{\circ}{h}_p \overset{\circ}{x}_p|^2 \right] \\ &\cdot \gamma_n^{P+\alpha_n-1} \exp[-\beta_n \gamma_n] \cdot \prod_{p=1}^P [s_p \cdot \lambda_p^k(\boldsymbol{\theta}_k)] \\ &\cdot \gamma_x^{P+\alpha_x-1} \exp[-\beta_x \gamma_x] \exp \left[ -\gamma_x \sum_{p=1}^P |x_p|^2 s_p \lambda_p^k(\boldsymbol{\theta}_k) \right] \\ &\cdot \mathcal{U}[\boldsymbol{\theta}_k^m, \boldsymbol{\theta}_k^M](\boldsymbol{\theta}_k) \cdot \prod_{p=1}^P s_p^{\alpha_s-1} \cdot \exp \left[ -\beta_s \sum_{p=1}^P s_p \right] \end{aligned} \quad (22)$$

where the normalization constant is  $\mathcal{C} = K^{-1} \cdot (2\pi)^{-3P} \cdot \beta_n^{\alpha_n} \cdot \Gamma^{-1}(\alpha_n) \cdot \beta_x^{\alpha_x} \cdot \Gamma^{-1}(\alpha_x) \cdot \beta_s^{P\alpha_s} \cdot \Gamma(\alpha_s)^{-P} \cdot (\boldsymbol{\theta}_k^M - \boldsymbol{\theta}_k^m)^{-1}$ .

However, this law cannot be directly sampled, thus MCMC methods will be employed, more precisely, Gibbs sampling. This can be performed via two types of algorithms.

- Across-model approach—joint sampling of the model index and its parameters. The algorithm jumps from one model to another and explores the joint model index plus parameter space, yielding a joint chain of model indexes and parameter values. The most representative algorithm of this type is Reversible Jump MCMC (RJCMC) [42].

- Within-model approach—consists in exhaustively visiting the candidate models and parameter sampling conditionally on the model. It provides  $K$  chains of parameter values, one for each model. For a detailed description see [43], [44] and the more recent survey [45].

Despite the conceptual differences, for a finite candidate models set, the two approaches yield the same result (provided they have reached convergence) but, under two different forms.

The RJMCMC algorithm is especially interesting for a very large number of models, when an exhaustive sequential sampling of all the models may be prohibitively expensive. However, in our problem, the number of concurrent models is rather reduced. In this case, the within-model strategy is better, avoiding the non-trivial RJMCMC problems concerning the parameter transformation when switching models. Moreover, the within-model approach guarantees that all models have been thoroughly explored and the model selection is not affected by the sampling algorithm. For this reason, our model choice method is based on within-model posterior sampling.

#### B. Evidence Approximations Based on Posterior Samples

Our model selection method is based on evidence approximation from posterior samples and the proposed method will be referred to as the Classifier based on Evidence Approximation from Posterior Samples (CEAPS). This method can be formulated using two different approximations of the evidence, which we present in the following: the Harmonic Mean Approximation (HMA) [44] and the Laplace-Metropolis Approximation (LMA) [16].

1) *Harmonic Mean Approximation*: We consider that  $\boldsymbol{\Psi}^{(t)} = (\mathbf{x}^{(t)}, \gamma_n^{(t)}, \gamma_x^{(t)}, \mathbf{s}^{(t)}, \boldsymbol{\theta}_k^{(t)})$ ,  $t = 1 \dots T$  are samples of the *a posteriori* law. Then, the evidence can be computed as:

$$\tilde{e}_k \approx \left\{ \frac{1}{T} \sum_{t=1}^T \left[ f(\mathbf{y}|\boldsymbol{\Psi}^{(t)}, \mathcal{M} = k) \right]^{-1} \right\}^{-1} \quad (23)$$

i.e., the harmonic mean of the likelihood values for the  $\boldsymbol{\Psi}^{(t)}$ .

Although  $\tilde{e}_k$  converges almost surely to the true value  $e_k$  when  $T \rightarrow \infty$  [46], it does not generally satisfy the central limit theorem [45]. Occasionally, a  $\boldsymbol{\Psi}^{(t)}$  with significant *a priori* probability, but very low likelihood, may occur. Its contribution in the harmonic mean is high and this may trigger infinite variances [44]. Solutions to stabilize this approximation have been provided in [47]. Nevertheless, we have not encountered this difficulty neither in our previous work [28], nor in the current one, where the priors are uniform on a finite interval and the likelihood is very peaked. Hence, the posterior samples are distributed in the regions where the likelihood has significant values. Consequently, the situations where the HMA may diverge or converge too slowly are avoided.

2) *Laplace-Metropolis Approximation*: The evidence (20) can also be expressed as:

$$e_k = \int_{\boldsymbol{\Psi}} \exp \left\{ \underbrace{\log [f(\mathbf{y}|\boldsymbol{\Psi}, \mathcal{M} = k) \cdot \pi(\boldsymbol{\Psi}|\mathcal{M} = k)]}_{F_k(\boldsymbol{\Psi}, \mathbf{y})} \right\} d\boldsymbol{\Psi} \quad (24)$$

with  $F_k(\boldsymbol{\Psi}, \mathbf{y})$  the log-posterior computed for observation  $\mathbf{y}$ .



*Remark:*  $F_k(\Psi, \mathbf{y})$  is called the observed information and indicates the amount of available information, for the given observation  $\mathbf{y}$ .  $F_k(\Psi, \mathbf{y})$  is related to the Fisher information  $\mathcal{I}_k$  introduced in Section IV through (13) as follows:

$$\mathcal{I}_k(\Psi) = -\mathbb{E}_{\mathbf{y}} [F_k''(\Psi, \mathbf{y})] \quad (25)$$

Under the hypothesis that  $F_k(\Psi, \mathbf{y})$  is twice differentiable with a unique maximum in  $\Psi^*$ , the Laplace approximation of a quadratic function (also used for the derivation of BIC) can be applied to evaluate the integral (24):

$$\hat{e}_k \approx \exp [PF_k(\Psi_k^*, \mathbf{y})] \cdot \left(\frac{2\pi}{P}\right)^{\frac{D_k}{2}} \cdot |F_k''(\Psi_k^*, \mathbf{y})|^{-1/2} \quad (26)$$

where  $\Psi_k^*$  represents the MAP value for model  $k$ ,  $F_k''(\Psi_k^*, \mathbf{y})$  is the Hessian of the log-posterior, evaluated at  $\Psi_k^*$  and  $D_k$  is the dimension of model  $k$ . The last factor in (26) is the determinant of the observed information matrix. In fact, computing the evidence in this manner consists in determining the MAP value,  $\Psi^*$ , i.e., the value for which  $F(\Psi, \mathbf{y})$  is maximum, and replacing this value in (26). [48] reviews the Laplace based methods for evidence computation. These approximation methods have relative errors of order  $\mathcal{O}(P^{-1})$ .

This approximation can also be performed based on MCMC samples from the posterior. In this case the method is called Laplace Metropolis Approximation (LMA) [16]. The LMA can be based in the MAP, the Posterior Mean (PM), or the Median *a Posteriori* (MedAP) and the Hessian can be approximated by the covariance matrix of the samples. In our case, this approximation is performed using the PM and the value of the Hessian computed for the PM.

*Remark:* The LMA explicitly penalizes complex models due to the second factor that decreases exponentially with model dimension, while for the HMA the model complexity penalization is implicit.

### C. Gibbs Within-Model Posterior Sampling

The samples from the posterior law are obtained using Gibbs sampling. Among the various strategies, we have chosen to sample  $\gamma_n$ ,  $\gamma_x$ ,  $\mathbf{x}$ ,  $\theta_k$  and  $\mathbf{s}$ . The advantage of this approach is that we obtain rather standard targets and we can perform parallel sampling for  $\mathbf{x}$  and  $\mathbf{s}$ . Then, the *a posteriori* conditional laws for the parameters are the following:

$$\begin{aligned} \mathbf{x} &\sim \prod_p \exp \left[ - \left( \gamma_n |\hat{y}_p - \hat{h}_p \hat{x}_p|^2 + \gamma_x |\hat{x}_p|^2 s_p \lambda_p^k(\theta_k) \right) \right] \\ &= \prod_p \mathcal{N}(m_p, v_p), \text{ with } m_p = \gamma_n \hat{y}_p \hat{h}_p v_p \text{ and } \\ &v_p = (\gamma_n g_p + \gamma_x s_p \lambda_p^k(\theta_k))^{-1} \text{—separable in the Fourier} \\ &\text{domain, i.e., parallel sampling is possible; computation} \\ &\text{cost equivalent to sampling the } a \text{ priori law;} \end{aligned}$$

$$\mathbf{s} \sim \prod_p \left\{ s_p^{\alpha_s} \cdot \exp -s_p \left[ \gamma_x |\hat{x}_p|^2 \lambda_p^k(\theta_k) + \beta_s \right] \right\}$$

$= \prod_p \mathcal{G}(a_s, b_{s_p})$  with  $a_s = \alpha_s + 1$  and  $b_{s_p} = \beta_s + \gamma_x s_p |\hat{x}_p|^2 \lambda_p^k(\theta_k)$ —separable, independent on the observations, allowing for parallel sampling;

$$\begin{aligned} \gamma_n &\sim \gamma_n^{P+\alpha_n-1} \cdot \exp -\gamma_n \left( \beta_n + \sum_p |\hat{y}_p - \hat{h}_p \hat{x}_p|^2 \right) \\ &= \mathcal{G}(a_n, b_n), \text{ with } a_n = \alpha_n + P \text{ and } b_n = \beta_n + \sum_p |\hat{y}_p - \hat{h}_p \hat{x}_p|^2; \end{aligned}$$

$$\begin{aligned} \gamma_x &\sim \gamma_x^{P+\alpha_x-1} \cdot \exp -\gamma_x \left( \beta_x + \sum_p |\hat{x}_p|^2 s_p \lambda_p^k(\theta_k) \right) \\ &= \mathcal{G}(a_x, b_x), \text{ with } a_x = \alpha_x + P \text{ and } b_x = \beta_x + \sum_p |\hat{x}_p|^2 s_p \lambda_p^k(\theta_k); \end{aligned}$$

$$\theta_k \sim \mathcal{U}(\theta_k) \cdot \prod_{p=1}^P \lambda_p^k(\theta_k) \cdot \exp \left[ -\gamma_x s_p |\hat{x}_p|^2 \lambda_p^k(\theta_k) \right]$$

$f(\theta_k | \mathbf{x}, \mathbf{s}, \gamma_x, \mathcal{M} = k)$ —very complicated dependency. Since  $\theta$  has a non-standard law, more complex sampling strategies must be used. We chose to include an MH step [49] in the Gibbs loop. This Metropolis within Gibbs strategy is convergent, as proven in [50], i.e., provides a chain of values  $\Psi^{(t)}$ ,  $t = 1 \dots T$  that has the posterior as limit probability distribution (when the number of iteration tends to infinity).

*Remark:* The alternative to sampling all the unknowns is to integrate a part of them, in order to avoid their sampling steps. This strategy is similar to the collapsed Gibbs sampler method used in [1]. Some of the options are the following:

- $\mathbf{x}$  marginalization and sampling the rest of the unknowns, thus no texture sampling, but even more cumbersome dependency on  $\gamma_x$ ,  $\theta_k$ ,  $\mathbf{s}$  and  $\gamma_n$ . Regarding  $\mathbf{s}$ , it remains separable, thus parallel sampling is feasible, but not of Gamma laws. Furthermore, the posterior for  $\gamma_n$  and  $\gamma_x$  no longer have Gamma forms either:

$$\begin{aligned} f(s_p, \gamma_n, \theta_k | *) &\propto \exp \left[ \frac{1}{r_p} \frac{|\hat{y}_p|^2 g_p}{\gamma_x s_p \lambda_p(\theta_k)} \right] \\ &\cdot \exp - \left[ r_p \gamma_n \gamma_x s_p \lambda_p(\theta_k) \right] \left[ \hat{x}_p - \frac{1}{r_p} \frac{\hat{y}_p \hat{h}_p}{\gamma_x s_p \lambda_p(\theta_k)} \right]^2 \end{aligned}$$

- $\mathbf{s}$  marginalization, resulting in a complicated law for  $\mathbf{x}$  (loss of the advantage of the SMGRF texture model—the conditional Gaussianity for  $\hat{x}_p$ ). However,  $\hat{x}_p$  remain independent, thus can be sampled in parallel, but by more sophisticated samplers. The new conditional laws are:

$$\begin{aligned} f(\hat{x}_p | *) &\propto \left[ 1 + \frac{\gamma_x \lambda_p^k(\theta_k)}{\beta_s} |\hat{x}_p|^2 \right]^{-\alpha_s-1} \\ &\cdot \exp \left[ -\gamma_n |\hat{y}_p - \hat{h}_p \hat{x}_p|^2 \right] \end{aligned}$$

- integrate both  $\mathbf{x}$  and  $\mathbf{s}$  and sample only  $\gamma_n$ ,  $\gamma_x$  and  $\theta_k$ , which will be distributed under complicated laws.

Moreover, by marginalization, the resulting law is more diluted, thus what is gained by eliminating sampling steps, may be lost in terms of speed of convergence. We have chosen not to integrate any of them since, although this implies more sampling steps, the sampled laws are easier to handle.

#### D. Efficient Metropolis-Hastings for $\theta$

A Random Walk MH (RWMH) step consists in formulating a proposal  $\theta_{\text{pr}}$ , based on a proposition law of the form  $q(\theta_{\text{pr}}|\theta_c) = \theta_c + \phi$ , evaluating the acceptance probability  $\xi(\theta_c, \theta_{\text{pr}})$ , which is a function of the current value and the proposal, and then accepting or rejecting the proposal. If the proposal is accepted,  $\theta^{(t)} = \theta_{\text{pr}}$ , otherwise  $\theta^{(t)} = \theta_c$ . The algorithm is known to be convergent [51]: the samples are asymptotically distributed under the posterior law.

There are numerous options for formulating the proposal and both the convergence speed and mixing properties are directly influenced by the adequacy between the proposal law and the target. Thus, choosing a  $q$  that embeds information about the shape of the target can significantly enhance the algorithm performances. In this context, the RWMH algorithm formulates the proposal based on a stochastic component and, in some cases, a component based on the target. This can contain first or second order derivatives of the target, so as to ensure a convenient exploration of the parameter space.

In [26] we have developed an efficient sampler called the FRWMH. This algorithm is based on the idea of quasi-Newton proposals [23] for a fast exploration of the parameter space and superior mixing properties. In this context, the Hessian used in the quasi-Newton approach is replaced by the Fisher information matrix. The FRWMH algorithm formulates the proposal:

$$\theta_{\text{pr}} = \theta_c + \varepsilon \cdot \mathcal{I}^{-1}(\theta_c) \cdot \nabla_{\theta} \mathcal{L}(\theta_c) + \sqrt{\mathcal{I}^{-1}(\theta_c)} \cdot \mathbf{z}_c \quad (27)$$

where  $\mathcal{L}(\theta) = \log \pi(\theta|\mathbf{y})$  is the log-posterior,  $\mathbf{z}_c \sim \mathcal{N}(0, \mathbf{I})$  is an isotopic displacement and  $\mathcal{I}(\theta)$  is the Fisher information matrix, already defined in Section IV by (13).

In this case, the acceptance probability  $\xi(\theta_c, \theta_{\text{pr}})$  becomes:

$$\xi(\theta_c, \theta_{\text{pr}}) = \min \left\{ 1, \frac{\pi(\theta_{\text{pr}}) \cdot q(\theta_{\text{pr}}, \theta_c)}{\pi(\theta_c) \cdot q(\theta_c, \theta_{\text{pr}})} \right\} \quad (28)$$

with

$$q(\theta_{\text{pr}}, \theta_c) = \mathcal{N}(\theta_{\text{pr}} - \theta_c - \varepsilon \cdot \mathcal{I}^{-1}(\theta_c) \cdot \nabla_{\theta} \mathcal{L}(\theta_c), \mathcal{I}^{-1}(\theta_c))$$

The use of this proposal proved advantageous from multiple points of view. Firstly, this exploits the target curvature similarly to the Newton step from the optimization theory. Secondly, this made way for a series of algorithmic simplifications and performance enhancements:

- for our GRF textures, the second order derivatives vanished under the expectation. Thus, the efficient proposal was formulated only based on first order derivatives,
- the Fisher matrix is positive definite, hence when it is well-conditioned there are no instabilities when taking the inverse, such as those mentioned in [23] for the Hessian,
- also due to the positive definite Fisher matrix, the Newton term always has the direction of gradient ascent, thus the algorithm only makes efficient steps.

The same principle of using the Fisher matrix to formulate an efficient proposal is exploited in the manifold Metropolis adjusted Langevin algorithm (mMALA) [20].

#### E. Implementation Issues

The implementation has raised two numerical problems:

- 1) Since the likelihood has an exponential form for each Fourier coefficient and consists in a product over all the coefficients, this quantity often exceeds Matlab's representation capabilities. To overcome this obstacle, the co-log-likelihood (CLL), i.e., the negative log-likelihood is computed instead of the likelihood.
- 2) The problem is again encountered when computing the evidence from the CLL. The employed solution is to determine the minimum value of each CLL chain, subtract it from all the CLL chains and compute the evidences based on the "offset" values, in a logsumexp manner. The normalization is reversed in the final stage of posterior probability computation.

The resulting CEAPS procedure is given in Algorithm 1, where the "ComputeEvidence" routine can be based on either of the two aforementioned approximations.

These previous sections have presented theoretical aspects regarding the employed method, its mathematical formulation and the corresponding implementation issues. The following section will present a series of tests that evaluate the method performances.

## VI. EXPERIMENTAL RESULTS

This section is devoted to the description and interpretation the performances of our CEAPS model selection method from blurred and noisy textured images. Several experiments are presented.

- A. The first study compares the two MH samplers: the standard RWMH and the FRWMH.
- B. The second test compares the two evidence approximations based on posterior samples, i.e., HMA and LMA.
- C. The evaluation of our CEAPS classifier represents the third test. The classification performances for the CEAPS are first presented for various PSDs and then compared to those of the GMLE classifier.
- D. The classification performances are linked with the Fisher information analysis of Section IV.
- E. Results concerning the deconvolution are given in a visual form, by presenting the original textures, the observations and the deconvolved images.

Algorithm 1 proceeds as follows: a sampler is launched for each of the  $K$  concurrent models. These samplers compute at every iteration the evidence, based on the samples that have been drawn so far. When the difference between two consecutive values of the evidence is smaller than a threshold, the sampling is stopped. The first 10% of the samples are discarded as burn-in and the evidence is recomputed based only on the remaining 90% of the samples. Since the sampling is the most time consuming operation, it determines the duration of the overall algorithm and thus it is important to use efficient samplers.

---

**Algorithm 1:** Classifier based on Evidence Approximation from Posterior Samples (CEAPS) Algorithm
 

---

**input :** Data  $\mathbf{y}$ , models dictionary for  $\mathcal{M} = k, k = 1 \dots K$

**output:** Evidences  $\tilde{e}_k$  + samples for texture  $\mathbf{x}^{(t)}$ , noise parameter  $\gamma_n^{(t)}$ , texture parameters  $\boldsymbol{\theta}_k^{(t)}$  ( $t = 1..T$ )

*% prior model probabilities:*  
 $p_k = 1/K$ ;

*% generate samples of  $\pi(\boldsymbol{\theta}_k, \mathbf{x}, \mathbf{s}, \gamma_n | \mathbf{y}, \mathcal{M} = k)$*

**for**  $k = 1$  **to**  $K$  **do**

*% Gibbs sampler for  $(\gamma_n, \mathbf{s}, \boldsymbol{\theta}_k, \mathbf{x})$ ,  $\mathcal{M} = k$  fixed:*  
 $t = 1$ ;  
 initialization  $\boldsymbol{\theta}_k^{(t)}, \mathbf{x}^{(t)} = \mathbf{y}, \mathbf{s}^{(t)}$ ;  
 $\mathbf{C}\mathbf{L}\mathbf{L}_k(t) = \text{CoLogL}(\mathbf{x}^{(t)}, \mathbf{s}^{(t)}, \boldsymbol{\theta}_k^{(t)})$ ;  
 $m(k) = \min(\mathbf{C}\mathbf{L}\mathbf{L}_k)$ ;  
 $\mathbf{n}\mathbf{C}\mathbf{L}\mathbf{L}_k = \mathbf{C}\mathbf{L}\mathbf{L}_k - m(k)$ ;  
*% compute the evidence using HMA or LMA*  
 $\tilde{e}_k = \text{ComputeEvidence}(\mathbf{n}\mathbf{C}\mathbf{L}\mathbf{L}_k)$ ;  
 $e_{rec}(t) = 0$ ;

**while**  $|e_{rec}(t) - \tilde{e}_k| > \epsilon$  **do**  
 $t = t + 1$ ;  
 $\gamma_n^{(t)} \sim f(\gamma_n | \mathbf{y}, \mathbf{x}^{(t-1)}, \alpha_n, \beta_n)$   
 $\gamma_x^{(t)} \sim f(\gamma_x | \mathbf{x}^{(t-1)}, \boldsymbol{\theta}_k^{(t-1)}, \mathbf{s}^{(t-1)}, \alpha_x, \beta_x)$   
 $\mathbf{s}^{(t)} \sim f(\mathbf{s} | \mathbf{x}^{(t-1)}, \boldsymbol{\theta}_k^{(t-1)}, \alpha_s, \beta_s)$   
 $\boldsymbol{\theta}_k^{(t)} = \text{FRWMH with target } f(\boldsymbol{\theta}_k | \mathbf{x}^{(t-1)}, \mathbf{s}^{(t)})$   
 $\mathbf{x}^{(t)} \sim f(\mathbf{x} | \mathbf{y}, \gamma_x^{(t)}, \gamma_n^{(t)}, \mathbf{s}^{(t)}, \boldsymbol{\theta}_k^{(t)})$   
 $\mathbf{C}\mathbf{L}\mathbf{L}_k(t) = \text{CoLogL}(\mathbf{x}^{(t)}, \mathbf{s}^{(t)}, \boldsymbol{\theta}_k^{(t)})$ ;  
 $m(k) = \min(\mathbf{C}\mathbf{L}\mathbf{L}_k)$ ;  
 $\mathbf{n}\mathbf{C}\mathbf{L}\mathbf{L}_k = \mathbf{C}\mathbf{L}\mathbf{L}_k - m(k)$ ;  
 $\tilde{e}_k = \text{ComputeEvidence}(\mathbf{n}\mathbf{C}\mathbf{L}\mathbf{L}_k)$ ;  
 $e_{rec}(t) = \text{RecursiveEvidence}(e_{rec}(t-1), \tilde{e}_k)$ ;

**end**

**end**

*% determine the posterior model probabilities*

**for**  $k \leftarrow 1$  **to**  $K$  **do**

$$\Pr(\mathcal{M} = k | \mathbf{y}) = p_k \cdot \tilde{e}_k / \sum_{l=1}^K p_l \cdot \tilde{e}_l \cdot \exp[m(k) - m(l)]$$

*% Compute the parameter estimates by PM*  
 $\hat{\boldsymbol{\theta}}_k = \text{PM}(\boldsymbol{\theta}_k)$ ;

**end**

---

### A. RWMH vs FRWMH

The first tests investigate the speed performances of two sampling algorithms, the isotropic RWMH and our efficient FRWMH, in the context of the complicated laws for  $\boldsymbol{\theta}$ .

Our tests indicate that the use of the FRWMH yields an algorithmic speed increase by a factor of at least 10 as compared to the RWMH. This is due to the FRWMH directional form of the proposal, which permits the algorithm to attain the high probability region in a very small number of iterations. Once in this region, the directional component exhibits negligible values and the algorithm explores this high probability area of the parameter space due to the stochastic component of its proposal.

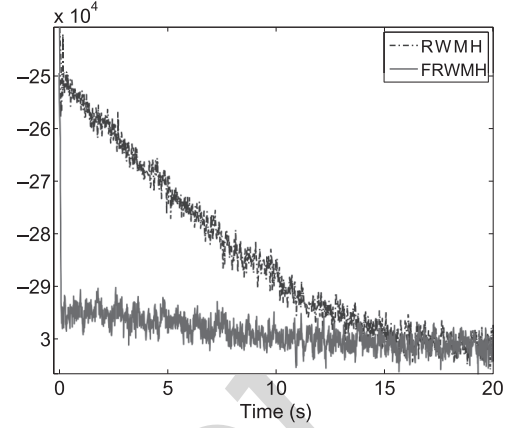


Fig. 9. CLL evolution—posterior sampling starting from the same initialization, using RWMH and FRWMH.

This translates into a very short burn-in period, as opposed to the isotropic RWMH, which has a significantly longer burn-in period, depending on the initialization. This efficiency is illustrated in Fig. 9 where the CLL chains for the two samplers are represented.

Using the efficient FRWMH sampler for the  $\boldsymbol{\theta}$  parameters, the overall sampling process takes roughly<sup>2</sup> 1 minute/model, thus the full run takes 8 minutes (the number of candidate models:  $K = 4$  shapes of  $\boldsymbol{\lambda}$  for the GRF type and  $K = 4$  shapes of  $\boldsymbol{\lambda}$  for the SMGRF).

### B. LMA and HMA

The posterior samples drawn by FRWMH within Gibbs sampling are used to compute the evidences. This is achieved via two approximations: the PM-based LMA and the HMA.

These approximations are computed using the same set of posterior samples in order to evaluate their accuracy in the same conditions. The numerical results show that the difference between the two evidence approximations is less than 0.1%, thus confirming that both approximations are viable for the problem in question.

Moreover, since the sampling is the most costly part of the evidence computation, the choice of approximation does not affect the overall speed performance. Consequently, the two evidence approximations imply a similar computational load.

In the tests presented in the following section, the CEAPS is based on the HMA.

### C. CEAPS Performances

Let us now present the performances of the selection method itself. The experimental setup consists in testing our method on synthetic textures, using 20 sets of parameter values for the PSD. Each set was used for each PSD model to generate both GRF and SMGRF texture realizations. The observations are obtained in a scenario with Gaussian blur of standard deviation  $w = 0.3$  and SNR = 20 dB. This corresponds to a partial overlap configuration, such as the one depicted in Fig. 8(b).

<sup>2</sup>Algorithms have been implemented using the computing environment *Matlab* on a Personal Computer, with a Intel Xeon 2 GHz and 1 GB of RAM.

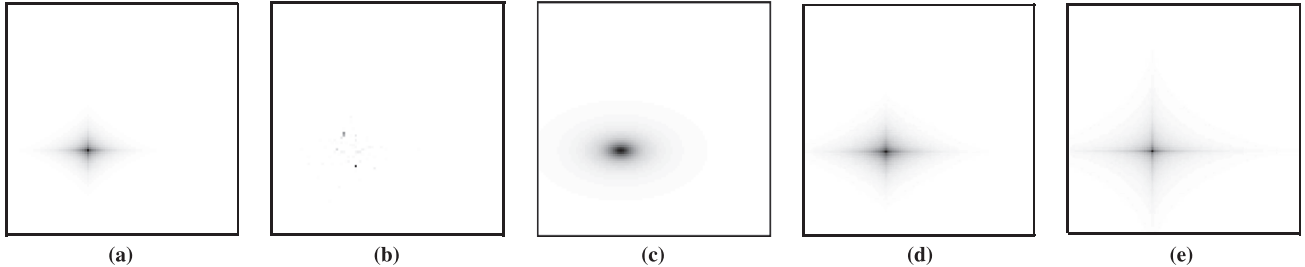


Fig. 10. Selection problem when the true model is not among the candidates. (a) True  $\lambda$ ; (b) Observed PSD; (c)  $\lambda$  candidate 1; (d)  $\lambda$  candidate 2; (e)  $\lambda$  candidate 3.

TABLE II  
CEAPS MODEL SELECTION PERFORMANCE FOR GRF TEXTURES (CORRECT CLASSIFICATIONS RATE IN %) FOR A PARTIAL OVERLAP CASE

True model	Estimated model			
	Lo	GL	Exp	GG
Lo	85	10	1	4
GL	<u>21</u>	69	3	7
La	2	4	87	7
GG	4	8	<u>16</u>	72

TABLE III  
CEAPS MODEL SELECTION PERFORMANCE FOR SMGRF TEXTURES (CORRECT CLASSIFICATIONS RATE IN %) FOR A PARTIAL OVERLAP CASE

True model	Estimated model			
	Lo	GL	Exp	GG
Lo	87	9	1	3
GL	<u>19</u>	73	2	6
La	1	7	83	9
GG	3	6	<u>20</u>	71

1) *CEAPS*: The algorithm was run on each texture realization and Tables II and III summarize the classification results for GRF and SMGRF textures, respectively. We observe on the main diagonal of both tables the percentages of correct classifications. As expected, the CEAPS chooses the correct PSD model in most cases. There are, however, situations where the method chooses another model.

As anticipated by the Fisher information analysis, in the majority of cases, there is enough available information on the central frequencies to ensure the samples are distributed under the correct law and have a finite variance. Nevertheless, the information concerning the widths is more sensitive to the noise level and the PSD model and thus more prone to errors. These errors are important for the method behavior, since they trigger misclassifications. The majority of missclassification cases are due to high noise levels and consist in mistakenly considering a PSD with thicker tails as the most adequate model. In this situation, the thicker tails account for the noise and the noise level is underestimated.

Nevertheless, in the context of our model choice problem, where the nested models help testing the method's ability to penalize model complexity, choosing another model can be regarded as not necessarily a failure. In this setting, the underlined percentages from Tables II and III represent the “good” miss-classifications, for instance, a Generalized Gaussian with  $q = 1$  that is classified as an exponential. This illustrates the

method's capacity to penalize model dimension, i.e., eliminate the parameters that do not significantly increase the model fit.

The method is not only able to distinguish between the different PSD forms, but also between the laws for the Fourier coefficients. More specifically, in 82% of the cases the algorithm correctly determined if the texture was from the GRF or the SMGRF class. This means that, on the one hand, the method has the ability to discriminate among a GRF and a SMGRF having the same form for the parametric part of the PSD. On the other hand, the PSD models themselves are structured enough to allow the algorithm to simultaneously identify the PSD model and whether all the PSD coefficients are identically scaled or not.

In the case where the real model that generated the data is not among the candidates, for each candidate model  $k = 1 \dots K$ , the chain of samples for  $\theta_k$  converges to the values that make model  $k$  best resemble the true model. Then, the evidences are computed based on these estimates and the model with the MAP probability is selected.

For instance, if the true model was an SMGRF with a certain shape for  $\lambda$  and the candidate models are only GRFs with various shapes for  $\lambda$ , an important remark is that the estimated widths for  $\lambda$  will be overestimated (i.e.,  $\sigma_x$  and  $\sigma_y$  will be underestimated), since the new  $\lambda$  will have to fit a wider range of frequencies (due to the presence of the auxiliary variables). Then, the method will choose the shape of  $\lambda$  that best fits the frequency components of the test image. Such a case is depicted in Fig. 10, where we show the true  $\lambda$ , the PSD of the observed texture (product among the true  $\lambda$  and auxiliary variables) and the PSDs of some of the candidate models (PSD is equivalent to  $\lambda$  in this case since the candidate models are GRFs).

2) *CEAPS vs GMLE*: A crucial point is that the GMLE cannot solve the problem of interest. This is due to the presence of indirect data introduced by the:

- non-Gaussian texture model,
- blurred and noisy observations.

Although the comparison cannot be performed on our problem, it is done on a simplified version of the problem dealing with direct observations (no noise, no convolution) of GRF textures.

Table IV lists the average classification success rate for the CEAPS and the GMLE, every method being tested on 20 texture realizations, with various parameter values for each type of PSD shape. More specifically, this represents an averaging over the PSD models and PSD parameters. The lower classification performance of the GMLE is due to the fact that it does not have any mechanism of model complexity penalization, thus it

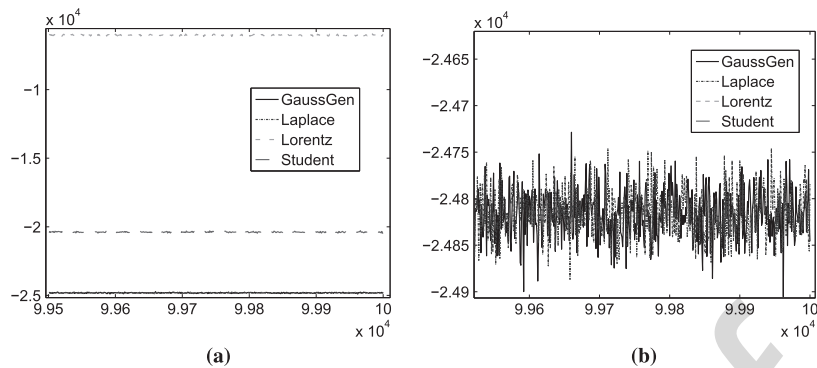


Fig. 11. CLL chains to illustrate a typical situation where GMLE fails to select the good model. (a) CLL chains for candidate models; (b) zoom on the CLL chains for the most likely models.

TABLE IV

AVERAGE MODEL SELECTION PERFORMANCE (CLASSIFICATION SUCCESS RATE IN %) COMPARISON BETWEEN CEAPS AND GMLE. THE AVERAGING IS DONE OVER PSD MODEL AND PSD PARAMETERS

Algorithm	Classification accuracy (%)
CEAPS	89
GMLE	86

chooses in 100% of the cases the most complex among the embedded models. On the contrary, as previously explained, the CEAPS penalizes model dimension and selects the less complex model that fits the data.

Table IV shows that the GMLE has a lower success rate and the reason for this is illustrated in Fig. 11. In this figure, we plotted a case where GMLE selects the Generalized Gaussian model, since its minimum CLL is the global CLL minimum among all models. However, the minimum co-log PM is that of the Laplacian model, which is indeed the true model. This is a typical failure situation for GMLE.

As already stated in Section II, the evidence based classifier is optimal from the risk point of view, property that can be seen in this table through the CEAPS performances.

#### D. Results and Fisher Information

The experiments show that high noise scenarios,  $\text{SNR} < 20$  dB, are challenging since the samples for the widths have a too strong variance (this variance is high when the Fisher information is low). Furthermore, these samples, used to compute the evidences, have a direct impact on the model selection process.

In practice, above a certain noise level, the method tends to favor the thicker tailed PSDs, by considering that these tails account for the noise. We have seen in the informational analysis, summarized in Fig. 7, that for high noise levels there is a smaller amount of information, thus more uncertainty in the estimation, which eventually triggers estimation errors for the PSD widths and even miss-classifications. More specifically, the noise level is undervaluated and either the PSD widths are overevaluated, or a model with thicker tails is selected.

#### E. Image Reconstructions

Using the samples employed to compute the evidences, we can also compute PM estimates for the texture parameters, the noise precision and the unknown image. Consequently,

as an additional result, our algorithm provides a PM estimate of the original image, conditionally on the selected model. Fig. 12 shows examples of the reconstruction. We can observe situations (Figs. 12(c), 12(f) and 12(i)) where CEAPS successfully restores the texture even if the observations are severely degraded. This illustrates the method's high capacity to handle the blur and the noise. This is due to the strength of the information given by the structure of the PSD and to the method's optimality from the classification and estimation risk point of view. Nevertheless, there are also situations, such as Fig. 12(l), where the image is degraded to an extent that impairs a reconstruction, in most cases, this being due to a low information scenario.

## VII. CONCLUSION AND PERSPECTIVES

This paper presents a method for texture model choice from blurred and noisy observations. The textured images are modeled by Scale Mixture of Gaussian Fields with parametric power spectral density and parametric probability density for the scale variables. In a Bayesian framework, we are able to determine the posterior model probabilities based on the evidences, this approach being optimal from the classification risk point of view.

The employed within-model simulation technique consists in a sweep of all possible models and the computation of the evidence for each model. This quantity can be determined only by numerical methods, since the required integral is intractable. We have compared two methods for numerically computing the evidence based on samples from the *a posteriori* law, the Laplace Metropolis Approximation and the Harmonic Mean Approximation, which yield the same results in the context of our problem. We have presented the performances of our Classifier based on Evidence Approximation from Posterior Samples, which is the optimal model choice strategy from the mean classification error point of view. This is reflected in the classification results that show the method's ability to select the true model.

As a secondary result, this approach provides chains of samples for the parameters, conditionally on each model  $\mathcal{M} = k$ . These samples can be used to obtain estimates that are optimal from the mean square error point of view, by using the Posterior Mean estimator. Moreover, using these estimates and the selected model, the original image can be reconstructed.



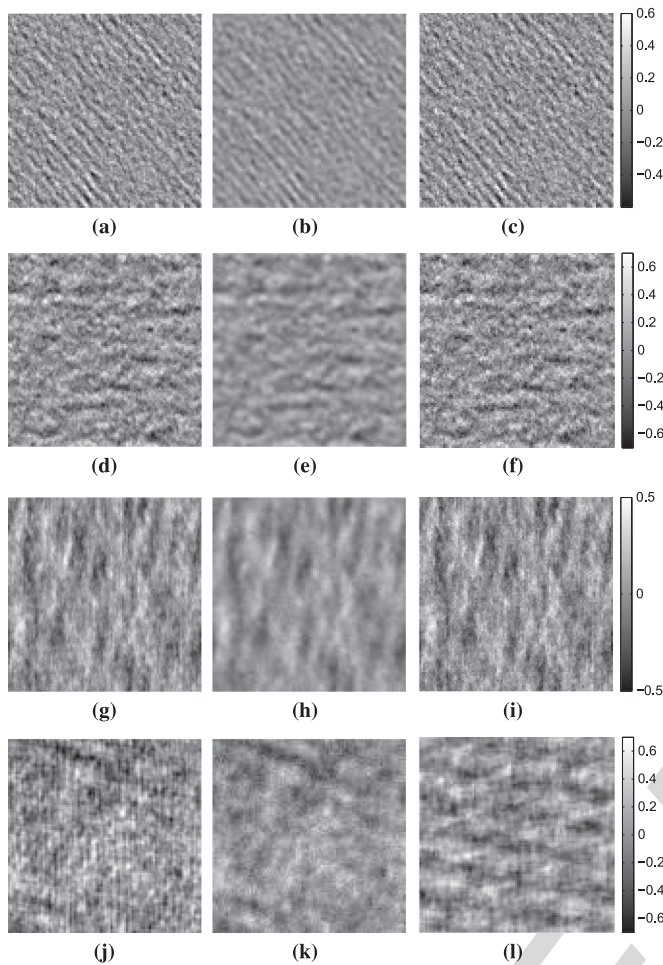


Fig. 12. Reconstruction results—4 cases to be read from left to right, on the first column the original, unobserved texture,  $\mathbf{x}|\theta^*$ ,  $\mathcal{M} = k^*$ , in the center the distorted observations,  $\mathbf{y}$ , and on the right column the results of the deconvolution, for the selected model,  $\hat{\mathbf{x}}|\hat{\theta}$ ,  $\mathcal{M} = \hat{k}$ .

Further developments include, but are not limited to, extending the SMGRF texture model to dependent Fourier coefficients, which would make it more versatile. Another idea is the use of multi-modal power spectral densities, in order to obtain more structured or quasi-periodic textures. From another standpoint, the method can be coupled with a beforehand learning step to enrich the set of possible texture model. From a different perspective, the method can be adapted to deal with unknown point spread functions (blind or semi-blind approach) and the estimation of their parameters [52]–[54].

#### REFERENCES

- [1] G. Kail, J.-Y. Tourneret, F. Hlawatsch, and N. Dobigeon, “Blind deconvolution of sparse pulse sequences under a minimum distance constraint: A partially collapsed Gibbs sampler method,” *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2727–2743, Jun. 2012.
- [2] L.-Y. Wei and M. Levoy, “Fast texture synthesis using tree-structured vector quantization,” in *Proc. 27th Conf. Comput. Graph. Interact. Tech.*, 2000, pp. 479–488.
- [3] F. Zhou, J.-F. Feng, and Q.-Y. Shi, “Texture feature based on local Fourier transform,” in *Proc. Int. Conf. Image Process.*, 2001, vol. 2, pp. 610–613.
- [4] G.-S. Xia, S. Ferradans, G. Peyré, and J.-F. Aujol, “Synthesizing and mixing stationary Gaussian texture models,” *SIAM J. Imag. Sci.*, vol. 8, no. 1, pp. 476–508, 2014.
- [5] M. N. Do and M. Vetterli, “Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models,” *IEEE Trans. Multimedia*, vol. 4, no. 4, pp. 517–527, 2002.
- [6] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 49–71, 2000.
- [7] R. Chellappa and S. Chatterjee, “Classification of textures using Gaussian Markov random fields,” *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-33, pp. 959–963, Aug. 1985.
- [8] R. Chellappa and R. Kashyap, “Texture synthesis using 2-d noncausal autoregressive models,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 1, pp. 194–203, 1985.
- [9] S. C. Zhu, Y. Wu, and D. Mumford, “Filters, random fields and maximum entropy (frame)—Towards a unified theory for texture modeling,” *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 1–20, 1998.
- [10] G. L. Gimelfarb, *Image Textures and Gibbs Random Fields*. Boston, MA, USA: Kluwer Academic, 1999, vol. 16.
- [11] E. P. Simoncelli, “Statistical modeling of photographic images,” in *Handbook of Video and Image Processing*. New York, NY, USA: Academic, 2005.
- [12] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [13] S. D. Spiegelhalter, N. G. Best, B. P. Carlin, and A. V. D. Linde, “Bayesian measures of model complexity and fit,” *J. Royal Statist. Soc.: Ser. B (Statist. Methodol.)*, vol. 64, no. 4, pp. 583–639, 2002.
- [14] G. Schwarz, “Estimating the dimension of a model,” *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.
- [15] M. J. Beal, “Variational algorithms for approximate bayesian inference,” Ph.D. dissertation, Univ. of Cambridge, Cambridge, U.K., 2003.
- [16] A. E. Raftery, “Hypothesis testing and model selection via posterior simulation,” in *Practical Markov Chain Monte Carlo*, W. R. Gilks, D. J. Spiegelhalter, and S. Richardson, Eds. London, U.K.: Chapman & Hall, 1995.
- [17] J. Rosenthal, “Optimal proposal distributions and adaptive MCMC,” in *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, Eds. Boca Raton, FL, USA: CRC, 2011, pp. 93–112.
- [18] B. Cai, R. Meyer, and F. Perron, “Metropolis-Hastings algorithms with adaptive proposals,” *Statist. Comput.*, vol. 18, no. 4, pp. 421–433, 2008.
- [19] G. Roberts and O. Stramer, “Langevin diffusions and Metropolis-Hastings algorithms,” *Methodol. Comput. Appl. Probabil.*, vol. 4, pp. 337–358, 2003.
- [20] M. Girolami and B. Calderhead, “Riemannian manifold Hamiltonian Monte Carlo (with discussion),” *J. Roy. Statist. Soc. B (Statist. Methodol.)*, vol. 73, pp. 123–214, 2011.
- [21] R. M. Neal, “MCMC using Hamiltonian dynamics,” in *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, Eds. Boca Raton, FL, USA: CRC, 2011, pp. 93–112.
- [22] S.-I. Amari, “Natural gradient works efficiently in learning,” *Neural Comput.*, vol. 10, no. 2, pp. 251–276, 1998.
- [23] Y. Qi and T. P. Minka, “Hessian-based Markov Chain Monte-Carlo algorithms,” presented at the 1st Cape Cod Workshop Monte Carlo Methods, Cape Cod, MA, USA, Sep. 2002.
- [24] **[AU: Provide more information on source/where to find--ed.]** T. Bui-Thanh and O. Ghattas, “Scaled stochastic Newton algorithm for Markov chain Monte Carlo simulations,” 2012.
- [25] J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas, “A stochastic newton MCMC method for large-scale statistical inverse problems with application to seismic inversion,” *SIAM J. Scientif. Comput.*, vol. 34, no. 3, pp. A1460–A1487, 2012.
- [26] C. Vacar, J.-F. Giovannelli, and Y. Berthoumiou, “Langevin and Hessian with Fisher approximation stochastic sampling for parameter estimation of structured covariance,” in *Proc. IEEE ICASSP*, Prague, 2011, pp. 3964–3967.
- [27] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, ser. Springer Texts in Statist. New York, NY, USA: Springer-Verlag, 2007.
- [28] C. Vacar, J.-F. Giovannelli, and A.-M. Roman, “Bayesian texture model selection using harmonic mean,” presented at the IEEE ICIP, Orlando, FL, USA, 2012.
- [29] D. Geman and C. Yang, “Nonlinear image recovery with half-quadratic regularization,” *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, 1995.



- [30] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, 1992.
- [31] J. F. Giovannelli, "Unsupervised Bayesian convex deconvolution based on a field with an explicit partition function," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 16–26, 2008.
- [32] J. M. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: A gem algorithm exploiting a class of heavy-tailed priors," *IEEE Trans. Image Process.*, vol. 15, no. 4, pp. 937–951, 2006.
- [33] Y. Zhang and N. Kingsbury, "Image deconvolution using a Gaussian scale mixtures model to approximate the wavelet sparseness constraint," in *Proc. IEEE ICASSP*, 2009, pp. 681–684.
- [34] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Appl. Computat. Harmon. Anal.*, vol. 11, pp. 89–123, 2001.
- [35] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, 2003.
- [36] D. K. Hammond and E. P. Simoncelli, "Image modeling and denoising with orientation-adapted Gaussian scale mixtures," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2089–2101, 2008.
- [37] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. London, U.K.: Chapman & Hall, 2005, vol. 104, Monographs on Statist. Appl. Probabil..
- [38] X. Guyon, *Random Fields on a Network: Modelling, Statistics, and Applications, Probability and Its Applications*. New-York, USA: Springer-Verlag, 1995.
- [39] A. Oppenheim and A. S. Willsky, *Signals and Systems*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983.
- [40] B. Picinbono, *Principles of Signals and Systems: Deterministic Signals*. Norwood, MA, USA: Artech House, 1988.
- [41] J. Skilling, "Nested sampling for general Bayesian computation," *Bayesian Anal.*, vol. 1, no. 4, 2006.
- [42] P. J. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.
- [43] A. E. Gelfand and D. K. Dey, "Bayesian model choice: Asymptotics and exact calculations," *J. Roy. Statist. Soc. B*, vol. 56, no. 3, pp. 501–514, 1994.
- [44] M. A. Newton and A. E. Raftery, "Approximate Bayesian inference with the weighted likelihood bootstrap," *J. Roy. Stat. Soc. B*, vol. 56, no. 1, pp. 3–48, 1994.
- [45] **[AU: More info on source and where to find--ed.]** C. P. Robert and D. Wraith, *Computat. Methods for Bayesian Model Choice 2009*.
- [46] A. G. Pakes, "On the convergence of moments of geometric and harmonic means," *Statistica Neerlandica*, vol. 53, no. 1, pp. 96–110, 1999.
- [47] A. E. Raftery, M. A. Newton, J. M. Satagopan, and P. N. Krivitsky, "Estimating the integrated likelihood via posterior simulation using the harmonic mean identity," *Bayesian Statist.*, pp. 1–45, 2007.
- [48] R. E. Kass and A. E. Raftery, "Bayes factors," *J. Amer. Statist. Assoc.*, vol. 90, pp. 773–795, 1995.
- [49] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, NY, USA: Springer-Verlag, 2004.
- [50] L. Tierney, "Markov chain for exploring posterior distribution," *Ann. Statist.*, vol. 22, no. 4, pp. 1701–1762, Dec. 1994.
- [51] C. P. Robert and G. Casella, *Monte-Carlo Statistical Methods*, ser. Springer Texts in Statistics. New York, NY, USA: Springer-Verlag, 2004.
- [52] F. Orieux, J. F. Giovannelli, and T. Rodet, "Bayesian estimation of regularization and point spread function parameters for Wiener-Hunt deconvolution," *J. Opt. Soc. Amer. A*, vol. 27, no. 7, pp. 1593–1607, 2010.
- [53] S.-U. Park, N. Dobigeon, and A. O. Hero, "Semi-blind sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3838–3849, 2012.
- [54] C. Vacar, J.-F. Giovannelli, and Y. Berthoumiu, "Bayesian texture and instrument parameter estimation from blurred and noisy images using MCMC," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 707–711, 2014.



**Cornelia Vacar** received the B.Sc. degree in telecommunications from the Technical University of Cluj-Napoca in 2009. She received the M.Sc. and Ph.D. degrees in image processing from the University of Bordeaux, France, in 2010 and 2014, respectively.

Her research interests focus on Bayesian methods for image deconvolution, texture classification, and image segmentation. She is currently activating within FEI Visualization Sciences Group.



**Jean-François Giovannelli** was born in Béziers, France, in 1966. He graduated from the École Nationale Supérieure de l'Électronique et de ses Applications, Cergy, France, in 1990. He received the Ph.D. and H.D.R. degree in signal–image processing at the Université Paris-Sud, Orsay, France, in 1995 and 2005, respectively.

From 1997 to 2008, he was an Assistant Professor with the Université Paris-Sud and a Researcher with the Laboratoire des Signaux et Systèmes, Groupe Problèmes Inverses. He is currently a Professor with the Université de Bordeaux, France, and a Researcher with the Laboratoire de l'Intégration du Matériau au Système, Groupe Signal-Image. His research focuses on inverse problems in signal and image processing, mainly unsupervised and myopic problems. From a methodological stand point the developed regularization methods are both deterministic (penalty, constraints, etc.) and Bayesian. Regarding numerical algorithms, the work relies on optimization and stochastic sampling. His application fields essentially concern astronomical, medical, proteomics, and geophysical imaging.



**Yannick Berthoumiu** received the Ph.D. degree in January 1996 in signal processing from the University of Bordeaux I, France.

In 1998, he joined the Ecole Nationale Supérieure d'Électronique, Informatique et Radiocommunications de Bordeaux, as an Associate Professor. He is currently a Full Professor with the Department of Telecommunications, Bordeaux Institute of Technology. His major research interests are concerned with multidimensional signal analysis, stochastic modeling, texture analysis, and video processing.

Since 2003, he has been with the Signal and Image Processing Group, IMS Lab, University of Bordeaux.