# Wavelet Modeling Using Finite Mixtures of Generalized Gaussian Distributions: Application to Texture Discrimination and Retrieval

Mohand Saïd Allili, *Member, IEEE*

***Abstract*—This paper addresses statistical-based texture modeling using wavelets. We propose a new approach to represent the marginal distribution of the wavelet coefficients using finite mixtures of generalized Gaussian (MoGG) distributions. The MoGG captures a wide range of histogram shapes, which provides better description and discrimination of texture than using single probability density functions (pdf's), as proposed by recent state-of-the-art approaches. Moreover, we propose a model similarity measure based on Kullback–Leibler divergence (KLD) approximation using Monte Carlo sampling methods. Through experiments on two popular texture data sets, we show that our approach yields significant performance improvements for texture discrimination and retrieval, as compared with recent methods of statistical-based wavelet modeling.**

***Index Terms*—Image segmentation, Kullback–Leibler divergence (KLD), mixture of generalized Gaussians (MoGG), texture, wavelet decomposition.**

## I. Introduction

TEXTURE analysis and representation play an important role in visual perception. As such, texture information is used in several computer-vision, remote-sensing, medical-imaging, and content-based image retrieval applications. In the last two decades, discrete wavelet transform (DWT) has been proven to be a powerful tool for texture analysis and representation. DWT decomposes an image into independent frequency bands exhibiting details and structures at multiple scales and orientations [37], [50], [52]. It is very efficient to compute, and any signal can be completely reconstituted using its decomposition [8], [21], [50]. Psychological research on human texture perception suggests that two textures are often difficult to discriminate when they produce a similar distribution of responses from a bank of linear filters [31], [40]. For this reason, the past approaches used, for example, different types of wavelet-based signatures, to discriminate between textures. Among the most recent approaches, we find those using statistical distributions modeling of subbands [7], [16], [18], [22], [23].

The main advantage of the statistical modeling of wavelets is that texture discrimination can be cast as a problem of similarity measurement between statistical models, which is relatively more efficient to implement, as compared with methods based, for example, on autoregressive models [1] or Markov random fields [30]. Recently, parametric distributions have been used to model wavelet coefficients distribution for the purpose of texture discrimination and retrieval. The most popular parametric model investigated so far is the generalized Gaussian density (GGD), which has been successfully used for texture classification and retrieval in [22] and [50]. Compared with the Laplacian or Gaussian distributions, the GGD has an additional free parameter that controls the shape of the distribution, which provides it flexibility to fit *platykurtic* and *leptokurtic* shapes of data [2], [12]. Due to this property, the GGD has been also successfully used in image and video coding [15], [33], [34], [48], image denoising [42], image segmentation [3], [4], [9], change detection [10], and texture discrimination and retrieval [16], [18], [22]. Do and Vetterli [22] demonstrated the superiority of using the GGD signature over energy-based methods for texture discrimination and retrieval. Indeed, it can be observed that the frequency subband distribution for a wide range of natural images is symmetrical and sharply peaked around zero (see [18], [22], and [37]). However, one can also find a range of images where this distribution is asymmetrical and/or multimodal, as is the case for images exhibiting approximate periodicities (see [13], [19], and [50]). Cossu *et al.* [19] observed, for instance, that wavelet packets histograms can have Gaussian, leptokurtic, or multimodal forms. One can also observe histograms with heavy-tailed and sharply peaked shapes at the same time (see Fig. 2 in this paper). Using a single GGD or any unimodal distribution to model the wavelet coefficients, in this case, would not perfectly capture the shape of the coefficients distribution, which, in turn, may deteriorate the performance of the wavelet signature for texture discrimination and retrieval.

In this paper, we propose a new statistical framework based on finite mixtures of generalized Gaussian (MoGG) distributions modeling for wavelet coefficients. MoGG modeling, introduced in [3], gathers the advantages of the GGD and the mixture modeling to offer a powerful tool to fit multimodal data histograms with heavy-tailed and sharply peaked modes. Compared with the mixture of Gaussians (MoG) distributions, MoGG achieves high-precision data fitting while using a lesser

Manuscript received June 03, 2011; revised September 11, 2011; accepted September 14, 2011. Date of publication October 06, 2011; date of current version March 21, 2012. This work was supported by the Natural Sciences and Engineering Research Council of Canada. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Charles Creusere.

The author is with the Université du Québec en Outaouais, Département d'Informatique et d'Ingénierie, Gatineau, QC J8X 3X7, Canada (e-mail: mohandsaid.allili@uqo.ca).

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/TIP.2011.2170701

number of components (i.e., model parsimony). These properties make it an ideal tool to represent the different forms of the wavelet subband histograms and provide a precise signature for texture discrimination and retrieval. To estimate the parameters of the MoGG model, we propose a Bayesian method based on optimizing a minimum message length (MML) objective. For the purpose of texture retrieval, we propose a new approach for calculating statistical-based similarity between images based on the Kullback–Leibler divergence (KLD) approximation. Experimental results on the VisTex [41] and Brodatz [14] texture databases have shown significant improvements in texture retrieval accuracy, as compared with recent approaches, while maintaining a comparable level of computational complexity. The application of our approach for texture segmentation will be also shown. A short version of this paper was published in [5]. In the current version, we gather the power of low- and high-pass wavelet subbands for more efficient texture discrimination and retrieval. We also give more details about MoGG model selection and learning and MoGG model comparison. Finally, we present a more thorough experimental validation using two texture data sets and an application for texture segmentation.

This paper is organized as follows. Section II presents related works and our main contributions. Section III presents the MoGG model selection and learning. Section IV presents our proposed MoGG similarity measures using KLD approximations. Section V presents some experimental results for texture retrieval and image segmentation applications using our approach. We end this paper with conclusion and future work perspectives.

## II. Related Works and Contribution

### A. Related Works

Texture analysis and modeling has been a very active research in the last decades; thus, an exhaustive review of all texture analysis methods in the literature is beyond the scope of this paper. Rather, we focus on existing methods for texture modeling with application to image discrimination and retrieval.

Among the most used features for texture analysis and characterization, we find filtering approaches [23], [25], [45] and statistical-based approaches [26], [39], [45], [49]. Randen and Husoy [45] compared several filtering approaches including Laws masks, dyadic Gabor filter banks, quadrature mirror filters, eigenfilters, and optimized finite impulse response filters. They concluded that co-occurrence signatures and Gabor filter banks have the best performance. Grigorescu *et al.* [29] compared Gabor filters with the cell operator, Gabor energy, and features based on complex moments, and arrived at the same conclusion. Liu and Wang [35] used spectral histograms as texture features. They concluded that these features outperform most filtering approaches.

Recently, several works have been carried out on the statistical modeling of wavelet distribution for texture characterization. Since the shape of the wavelet histogram is a critical characteristic for the image content, it is important to have flexible statistical models to capture the various histogram shapes. The authors in [22], [23], and [53] used the GGD to model high-pass wavelet subband histograms, where the parameters of the model are estimated using the maximum-likelihood (ML) method. In their approach, feature extraction and similarity measurement are integrated into a unified framework for texture retrieval. They successfully applied their approach to texture discrimination and retrieval and concluded that the GGD signature outperforms energy-based approaches. Choy *et al.* [16] extended this approach to model classes of texture using the characteristic GGD (CGGD) learnt from a set of GGDs. They concluded that the CGGD signature gives better performance than directly comparing GGD signatures for texture retrieval. The same authors successfully applied the bit-plane probability to model the wavelet coefficients after quantization [17]. They obtained superior performance for texture retrieval, as compared with using a single GGD as a texture signature. Recently, this approach has been improved by using the generalized Gamma distribution $G\Gamma D$ model for wavelet subband characterization [18], [24]. We should note, however, that the aforementioned models approximate only high-pass wavelet coefficients as they exhibit approximate unimodality in nature. The low-pass coefficients are often omitted since they cannot be approximated by these (unimodal) models, although they may carry an important amount of texture information (i.e., image approximation).

### B. Our Contributions

As stated in the introduction, the goal of this paper is to introduce a new method for modeling the distribution of wavelet coefficients, which offers many advantages over state-of-the-art methods. While the main limitation of existing statistical-based wavelet modeling is the inability to cope with different shapes of wavelet distributions, our approach provides an efficient and viable alternative that overcomes this problem. The achieved precision, in turn, drastically improves the performance of texture discrimination and retrieval. Our main contributions are summarized as follows.

First, the proposed model allows the accurate fitting of multimodal, asymmetrical, and heavy-tailed histograms, which is difficult to achieve with any single distribution (e.g., Gaussian, GGD, $G\Gamma D$, etc.). Furthermore, the use of the MoGG distributions with arbitrary location parameters allows the accurate modeling of any histogram with a mode shifted from zero. Second, the new model is able to accurately fit high- and low-pass wavelet coefficients, which is not possible to achieve with any single distribution, as proposed by numerous state-of-the-art methods. This permits leverage of the power of high- and low-pass subband coefficients in the same framework for texture characterization and therefore improves the performance of texture discrimination and retrieval. Third, we design a similarity measure between MoGG models based on the KLD approximation, which permits performance of efficient texture discrimination and searching. Various experiments of texture classification and retrieval, as well as texture image segmentation, demonstrate the efficiency of the proposed method on different databases.

## III. MoGG Distributions

### A. MoGG Model Learning Using MML

The general Gaussian distribution for a univariate random variable $\mathbf{x} \in \mathbb{R}$ is defined in its general form as follows (see [12]):

$$p(x|\mu,\sigma,\beta) = \frac{\beta\sqrt{\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}}}{2\sigma\Gamma(1/\beta)} \exp\left(-A(\beta)\left|\frac{x-\mu}{\sigma}\right|^{\beta}\right) \quad (1)$$

where $A(\beta) = [\Gamma(3/\beta)/\Gamma(1/\beta)]^{\beta/2}$, $\Gamma(\cdot)$ denotes the gamma function, and $\mu$ and $\sigma$ are the distribution mean and standard deviation parameters. Parameter $\beta \geq 1$ controls the kurtosis of the probability density function (pdf) and determines whether the distribution is peaked or flat; the larger the value of $\beta$, the flatter the pdf, and the smaller $\beta$ is, the more peaked the pdf is around its mean. As $\beta \to \infty$, the distribution becomes uniform, whereas when $\beta \to 0$, the distribution becomes a delta function at $\mu$. This parameter gives the pdf flexibility to fit the shape of flat and sharply peaked unimodal histograms [12], [51].

For multimodal data, we propose to use an MoGG distribution, as proposed in [3]. Given an MoGG with $K$ components, the marginal distribution of the random variable $\mathbf{x}$ is given by

$$p(x|\vec{\theta}) = \sum_{i=1}^{K} \pi_i p(x|\mu_i,\sigma_i,\beta_i) \quad (2)$$

where $0 < \pi_i \leq 1$ and $\sum_{i=1}^{K} \pi_i = 1$. Here, $\vec{\theta}$ designates the set of model parameters $\{\pi_i,\mu_i,\sigma_i,\beta_i, i = 1,\ldots,K\}$. These can be estimated using, for example, the ML method [3], [22]. However, the ML estimation is not the best way for model selection since it tends to encourage more complex models (i.e., high number of components) to maximize data fitting [11], [54]. Nonetheless, it can be used to estimate the values of the parameters once the model is selected.

To achieve a fully unsupervised model selection and parameter estimation, we follow a similar approach to [43] and [54] using the MML principle. It is proven that the MML principle provides an optimal way for model selection since it offers a good balance between model complexity and data fitting [3], [54]. Wallace and Freeman [55] advocated the use of MML as an optimal way to encode data since it selects the model with its parameters that gives the shortest overall message length. A more complex model will need a larger first part of the message to encode the model parameters but will give a better fit for data. On the other hand, a simple model will need a shorter first part of the message but may decrease the data fitting accuracy. Therefore, the MML provides a natural tradeoff between model complexity and goodness of fit.

Given a data sample $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$, which represents, in our case, wavelet coefficients, our goal is to estimate the optimal MoGG model, which provides the best tradeoff between data fitting accuracy and model complexity (i.e., the number of components $K$). The message length that encodes the wavelet coefficients in a given subband is given by (see [43] and [54])

$$\text{MessL} \simeq -\log p(\vec{\theta}) + \frac{1}{2}\log\left|I(\vec{\theta})\right| + \frac{c}{2}\left(1 + \log\frac{1}{12}\right) - \log p(\mathcal{X}|\vec{\theta}) \quad (3)$$

where $p(\vec{\theta})$, $I(\vec{\theta})$, and $p(\mathcal{X}|\vec{\theta})$ denote the prior distribution of parameters $\vec{\theta}$, the (expected) Fisher information matrix, and the likelihood of data, respectively. Here, $|I(\vec{\theta})|$ denotes the determinant of matrix $I(\vec{\theta})$. Finally, constant $c = 4K$ in (3) gives the total number of parameters in the mixture model.

To facilitate the calculation of the different parts of function (3), it is common sense to assume the independence of the different groups of parameters in $\vec{\theta}$, which factorizes both the Fisher information $I(\vec{\theta})$ and the prior distribution $p(\vec{\theta})$. Note that parameters $\pi_i$ are defined in simplex $\{\pi_1,\ldots,\pi_K : \sum_{i=1}^{K-1}\pi_i < 1\}$. Therefore, a natural choice as a conjugate prior for these parameters is the Dirichlet distribution with hyperparameters set to 1/2 (i.e., uninformative Jeffrey prior [46]); thus, $p(\pi_1,\ldots,\pi_K) = \prod_{i=1}^{K}\pi_i^{-1/2}$. In the same vein, we choose a prior probability for each of the GGD parameters in the mixture. Following [43], we suppose the mean and the variance of the data are calculated and are equal to $\mu_{\text{pop}}$ and $\sigma^2_{\text{pop}}$, respectively. Then, the priors for parameters $\mu_i$ and $\sigma_i$ can be chosen as $p(\mu_i) = 1/2 \times \sigma_{\text{pop}}$ and $p(\sigma_i) = 1/\sigma_{\text{pop}}$, respectively, where we supposed $0 \leq \sigma_i \leq \sigma_{\text{pop}}$ and $\mu_{\text{pop}} - \sigma_{\text{pop}} \leq \mu_i \leq \mu_{\text{pop}} + \sigma_{\text{pop}}$. Finally, the prior of the shape parameter $\beta_i$ is set to be a uniform distribution, where $p(\beta_i) = 1/\text{Kur}(\mathcal{X})$ in interval $0 \leq \beta_i \leq \text{Kur}(\mathcal{X})$, $\text{Kur}(\mathcal{X})$ being the kurtosis of data $\mathcal{X}$.

The priors selected herein have the following properties over those in [3]. They are scale invariant. That is to say, if we change the units of measurement in a linear fashion, the priors will be invariant. Moreover, the uniform priors of the model mean and variance parameters give a rough idea about their values, knowing mean $\mu_{\text{pop}}$ and variance $\sigma_{\text{pop}}$ of the data. The uniform prior about the shape parameter $\beta_i$ represents our belief of the values it can have in the light of the data. In fact, most of the wavelets coefficient distributions are heavy tailed and/or sharply peaked. Kurtosis $\text{Kur}(\mathcal{X})$ measures the flatness/peakedness of the data, which gives a rough idea about the shape of the distribution.

For the calculation of the Fisher information in (3), we use the complete-data Fisher information matrix $I_c(\vec{\theta})$ [43] for which the labels of the data are known. In this case, the Fisher information will be block diagonal containing $4K$ elements, which is given as follows (see [27]):

$$I_c(\vec{\theta}) = \text{block-diag}\left[\mathcal{M}, n_1\pi_1 I^{(1)}(\vec{\theta}_1),\ldots,n_K\pi_K I^{(1)}(\vec{\theta}_K)\right] \quad (4)$$

where $n_j$ is the number of data classified in the $j$th component of the mixture and $\mathcal{M}$ is the information matrix of the multinomial distribution with parameters $\pi_1,\pi_2,\ldots,\pi_K$. Finally, $I^{(1)}(\vec{\theta}_j)$ is the one-observation Fisher information matrix associated with the internal parameters of the $j$th component of the mixture. The elements of $I_c(\vec{\theta})$ are given in Appendix I.

The cost function (3) is minimized using the expectation–maximization (EM) algorithm, which we have developed in [3]. In this algorithm, the expectation step essentially consists of calculating the posterior probability of each component of the mixture. The maximization step consists of updating the parameters of the mixture in a way that increases the expectation of the complete likelihood of the data. We obtain the following updating equations:

$$\hat{\pi}_k = \frac{1}{n}\sum_{i=1}^{n} p(\theta_k|x_i) \tag{5}$$

$$\hat{\mu}_k = \frac{\sum_{i=1}^{n} p(\theta_k|x_i)|x_i-\mu_k|^{\beta_k-2}x_i}{\sum_{i=1}^{n} p(\theta_k|x_i)|x_i-\mu_k|^{\beta_k-2}} \tag{6}$$

$$\hat{\sigma}_k = \left[\frac{\beta_k A(\beta_k)\sum_{i=1}^{n} p(\theta_k|x_i)|x_i-\mu_k|^{\beta_k}}{\sum_{i=1}^{n} p(\theta_k|x_i)}\right]^{1/\beta_k} \tag{7}$$

where $\theta_k$ denotes the parameters of component $k$ of the mixture and $A(\beta_k)$ is defined in (1). The posterior probability $p(\theta_k|x_i)$ is given by $p(\theta_k|x_i) = \pi_k p(\mathbf{x}_i|\theta_k)/\sum_{j=1}^{K}\pi_j p(\mathbf{x}_i|\theta_j)$. Finally, the shape parameter $\beta_k$ for each component is simultaneously estimated with the other parameters using the Newton–Raphson method, i.e.,

$$\hat{\beta}_k \simeq \beta_k - \left\{\frac{\partial^2 \log\left(p(\mathcal{X}|\vec{\theta})\right)}{\partial\beta_k^2}\right\}^{-1}\frac{\partial \log\left(p(\mathcal{X}|\vec{\theta})\right)}{\partial\beta_k} \tag{8}$$

where $p(\mathcal{X}|\vec{\theta})$ is the likelihood of generating the data by the model. The different steps for model selection and parameter estimation can be summarized in the script of Alg. 1. As suggested in [27], we test several values of $K$, starting from $K_{\max}$ to $K_{\min}$, to find the optimal one. Note that $K_{\min}$ and $K_{\max}$ can be fixed for all the estimations.

**Algorithm 1**: Algorithm for MoGG model selection.

**for** $K = K_{\max} \rightarrow K_{\min}$ **do**

**repeat**

1. Estimate the location parameters $\mu_k$ using (5).
2. Estimate the scale parameters $\sigma_k$ using (6).
3. Estimate the mixture weights $\pi_k$ using (7).
4. Estimate the shape parameters $\beta_k$ using (8).

**until** (Convergence)

Calculate the message length using (3).

**end for**

Return the optimal model $K^*$ with its parameters.

### *B. MoGG Model for Wavelet Representation*

To measure the accuracy of the proposed model for wavelet histogram fitting, we chose two particular texture images (see Fig. 1), where the shape of the detail coefficient histogram of the horizontal subband exhibits a strong heavy-tailed and sharply peaked shape in the first image and an asymmetrical shape in the second image. We compared the fitting accuracy of the proposed model with single GGD and MoG fitting, respectively. To measure the fitting accuracy of each model, we use the goodness-of-fit statistic value $\chi^2$, which is defined as

Fig. 1. Image RT10–37 in [56] and Fabric–14 in [41] used to obtain the results in Fig. 2.

$$\chi^2 = \sum_{x}\frac{\left(\mathbf{H}(x)-\mathbf{p}(x)\right)^2}{\mathbf{p}(x)} \tag{9}$$

where $\mathbf{H}(x)$ and $\mathbf{p}(x)$ represent the empirical and expected frequencies for the wavelet coefficient value $x$, respectively. Fig. 2 shows the different approximations of the detail coefficient histogram of the two images. We can note that using a single GGD does not provide a perfect histogram fitting, whereas the MoGG model can already almost perfectly fit the histogram using a combination of two GGDs in the first example and three GGDs in the second example. To achieve a comparable fitting error, the MoG model requires four and five components for the same examples, respectively. These results demonstrate that the MoGG model is more compact than the MoG model, which is a very desirable property in signal coding for example.

Fig. 3 shows the values of the $\chi^2$ statistic as a function of the number of components used in the MoG and MoGG models, respectively. We note that the MoGG model requires, in general, a lesser number of components to reach a high level of histogram fitting accuracy than using the MoG model. This demonstrates that, to fit the data, the MoGG approach uses less complex models than the MoG approach, and therefore, the wavelets coefficients distribution is better explained by the MoGG model. For applications, such as image coding and compression [15], [33], [34], this aspect is of prominent importance.

## IV. Similarity Measurement Between Two MoGGs

Recently, Do and Vetterli [22] developed a closed-form KLD similarity measurement between univariate centered GGDs, where the authors reported superior performance for texture retrieval than using Euclidian and energy-based similarity measures. However, when dealing with MoGG distributions with arbitrary mean parameters, a closed-form solution for the KLD is intractable. Note that the same problem has been pointed out for similarity measurement between MoGs (see [28] and [32]). To remedy this issue, one can resort to approximating the KLD using techniques such as Monte Carlo sampling methods [32], [47] or bounding approximation of the KLD [28], [32]. Given two MoGG models $f(x) = \sum_{i=1}^{K}\pi_i p(x|\theta_i)$

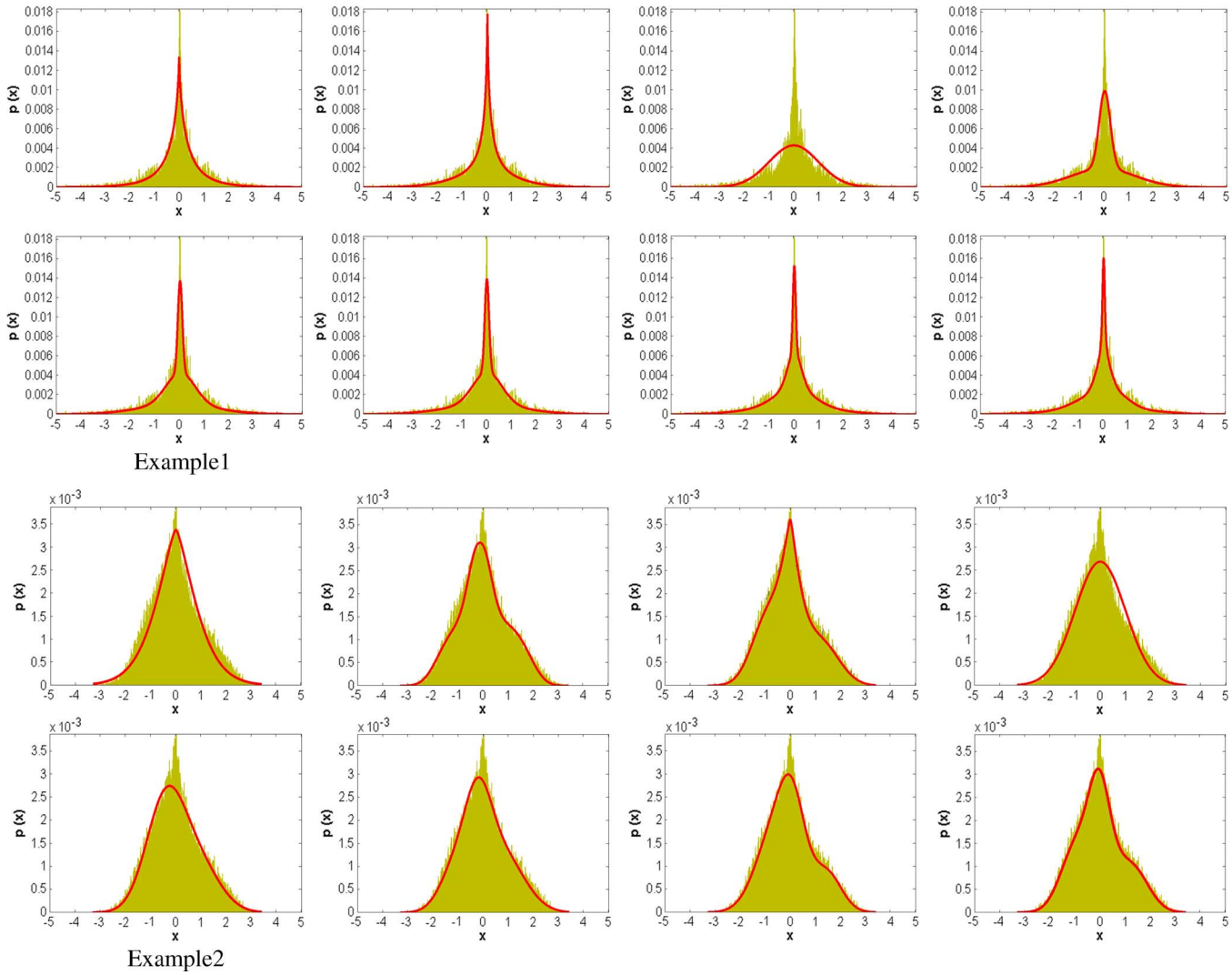


Fig. 2. Approximation of the wavelet histogram using (from left to right and top to bottom) Example 1: MoGG-1, MoGG-2, MoG-1, MoG-2, MoG-3, MoG-4, MoG-5, and MoG-6; Example 2: MoGG-1, MoGG-2, MoGG-3, MoG-1, MoG-2, MoG-3, MoG-4, and MoG-5. The number of components used in each model is added at the end of each model acronym.

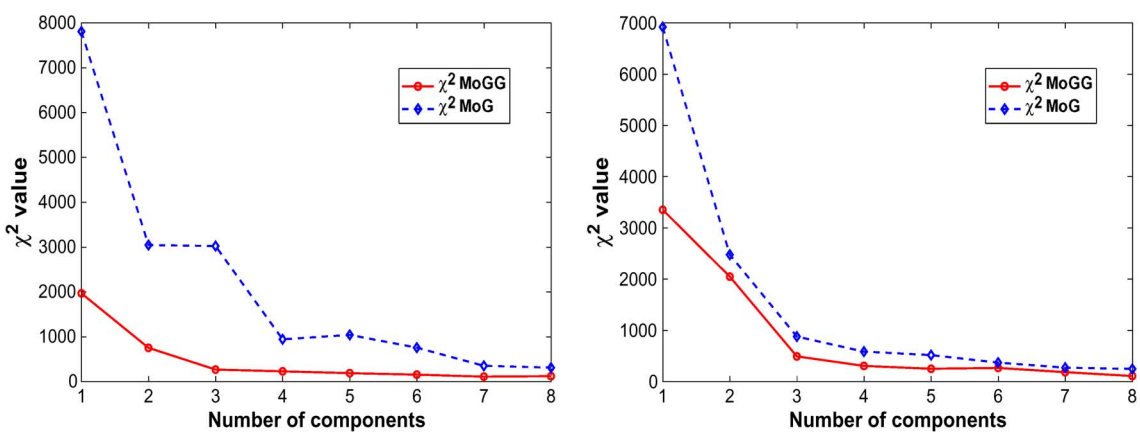


Fig. 3. $\chi^2$ statistic values for the MoGG and MoG approximations of the wavelet coefficient histograms in Fig. 2, as a function of the number of components $K$.

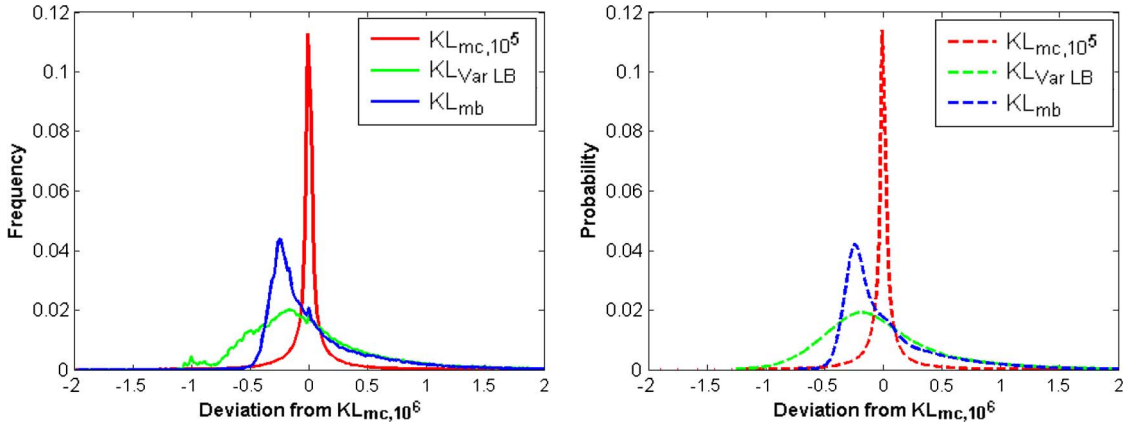


Fig. 4. Distribution of different approximation methods with regard to the RSM. (a) Frequencies of the deviations of the KLD approximations from the RSM. (b) Approximation of the graphs in (a) using three MoGG models. (Red line) Monte Carlo approximation using $10^5$ samples. (Blue line) Variational lower bound and (green line) matching-based approximations.

and $g(x) = \sum_{j=1}^{M} \omega_j p(x|\theta_j)$, the KLD between $f$ and $g$ is defined as [20]

$$\begin{aligned} \mathrm{KL}(f\|g) &= \int f \log\left(\frac{f}{g}\right) dx \\ &= \int \sum_{i=1}^{K} \pi_i f_i \log\left(\frac{\sum_{i=1}^{K} \pi_i f_i}{\sum_{j=1}^{M} \omega_j g_j}\right) dx \end{aligned} \quad (10)$$

where, to alleviate the notations, we used functions $f_i$ and $g_j$ instead of $p(x|\theta_i)$ and $p(x|\theta_j)$, respectively. In [6], we compared the performance of four techniques for approximating the KLD between MoGGs, namely, Metropolis–Hastings sampling, accept–reject sampling, variational lower bound, and matching-based approximations. In what follows, these methods are briefly explained (for more details of how these approximations are derived we refer to [6]).

- *Sampling methods* aim to generate a sufficiently large sample $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ independently drawn from the distribution $f$ in order to approximate the expectation of the KLD in (10) using Monte Carlo integration, i.e.,

$$\mathrm{KL}_{\mathrm{mc}}(f\|g) = \frac{1}{n}\sum_{i=1}^{n} \log\frac{f(x_i)}{g(x_i)} \approx \mathrm{KL}(f\|g) \quad (11)$$

  which converges to $\mathrm{KL}(f\|g)$ when $n \to \infty$. In [6], we tested for both Metropolis–Hastings and accept–reject samplings and conclude that the latter is more accurate in approximating the true KLD.
- *Variational approximation* aims to generate a lower bound for the KLD. We have

$$\mathrm{KL}(f\|g) \triangleq E_f\left[\log(f)\right] - E_f\left[\log(g)\right] \quad (12)$$

  where $E_f$ denotes the expectation with regard to distribution $f$. By introducing a variational parameter $\varphi_{ij}$, where $i \in \{1, \ldots, K\}$ and $j \in \{1, \ldots, M\}$, such that $\varphi_{ij} > 0$ and $\sum_j^M \varphi_{ij} = 1$, and using Jensen's inequality [36], we obtain

$$\begin{aligned} E_f\left[\log(g)\right] &= E_f\left[\log\left(\sum_{j=1}^{M} \varphi_{ij}\frac{\omega_j g_i}{\varphi_{ij}}\right)\right] \\ &\geq E_f\left[\sum_{j=1}^{M} \varphi_{ij}\log\left(\frac{\omega_j g_j}{\varphi_{ij}}\right)\right] \triangleq L_f(g,\varphi). \end{aligned} \quad (13)$$

  After maximizing $L_f(g,\varphi)$ with respect to each parameter $\varphi_{ij}$, we obtain the maximum lower bound for $E_f[\log(g)$. By applying the same reasoning to $E_f[\log(f)]$, we obtain a lower bound for (12), given by [6]

$$\mathrm{KL}_{\mathrm{var}}(f\|g) \triangleq \sum_{i=1}^{K} \pi_i \log\left(\frac{\sum_{l=1}^{K} \pi_l \exp\left(-\mathrm{KL}(f_i\|f_l)\right)}{\sum_{j=1}^{M} \omega_j \exp\left(-\mathrm{KL}(f_i\|g_j)\right)}\right) \quad (14)$$

  where distances $\mathrm{KL}(f_i\|f_l)$ and $\mathrm{KL}(f_i\|g_j)$ can be very accurately approximated using either of the aforementioned sampling algorithms.
- *Matching-based approximation* defines a matching function $m : \{1, \ldots, K\} \to \{1, \ldots, M\}$ that associates for each mixture component in $f$ the most similar component in $g$, with respect to the following formula [28]:

$$m(i) = \arg\min_j \left(\mathrm{KL}(f_i\|g_j) - \log(\omega_j)\right). \quad (15)$$

  The approximation of the KLD is then defined as

$$\mathrm{KL}_{\mathrm{mb}}(f\|g) \triangleq \sum_{i=1}^{K} \pi_i \left(\mathrm{KL}\left(f_i\|g_{m(i)}\right) + \log\left(\pi_i/\omega_{m(i)}\right)\right). \quad (16)$$

Fig. 4 shows two graphs representing the relative frequencies of the signed deviation of each approximation method with respect to a *reference similarity measure* (Monte Carlo KLD approximation with $10^6$ samples). The graphs are generated using 384 mixture models, where the KLD is calculated between each two models. We can note that the variational lower bound approximation behaves similarly to the matching-based one. How-

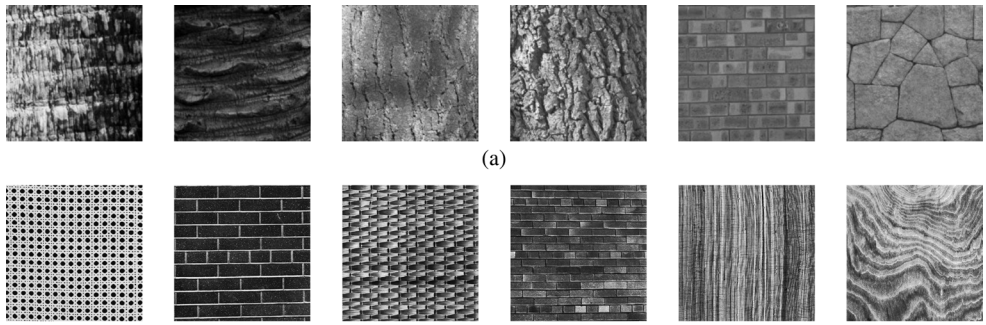

Fig. 5. Twelve images in our test data sets from (a) VisTeX and (b) Brodatz [14].

ever, the former produces less variation than the latter, which makes it more stable. Therefore, we can conclude that variational approximation performs better than matching approximation. On the other hand, we note that sampling-based approximation is better than using the other approximations. Therefore, we can conclude that sampling is better when precision is needed.

## V. Experiments

Texture classification is an important problem in computer vision as it plays a major role in a wide range of applications. To validate our aforementioned contributions, we conducted experiments using two popular databases for two applications, namely, texture image retrieval and mosaic texture segmentation. In what follows, we develop each application separately.

### A. Application to Texture Retrieval

In this application, we use two data sets for texture classification and retrieval, namely, Brodatz [14] and VisTeX data sets. In each data set, we choose 45 images of $512 \times 512$ pixels size and divide each image into sixteen $256 \times 256$ subimages, thus creating a database of $\mathcal{M} = 720$ subimages in each data set. Fig. 5 shows some images of the two data sets. To eliminate the effect of luminance range variation of the subimages and thus reduce bias in the retrieval phase, we normalized the luminance of each subimage as follows. Let $\mu_M$ and $\sigma_M$ be the medians of the means and the standard deviations of the 16 subimages that originate from the same original image. Each subimage $I_i$ of the 16 is normalized by the following formula:

$$I_i = \frac{(I_i - \mu_i)}{\sigma_i} \times \sigma_M + \mu_M \tag{17}$$

where $\mu_i$ and $\sigma_i$ are the mean and the standard deviation of $I_i$. After this step, we estimate for each subimage an MoGG model for each high- and low-pass (approximation) wavelets subband. The MoGG models constitute the signature of the image in the corresponding subbands. To test the retrieval performance, a sample of image queries is randomly selected from each database. A query image is compared with all images in the database, and we retain the $N$ top similar images ($N \ll \mathcal{M}$). To measure the similarity between the query image $I_Q$ and a target image $I_T$, we sum up the KLD approximations between all corresponding subbands in $I_Q$ and $I_T$, as follows:

$$S(I_Q, I_T) = \sum_{j=1}^{J} \sum_{d=1}^{3} \left( \mathrm{KL}\left( f_Q^{(j,d)} \| f_T^{(j,d)} \right) \right) + \mathrm{KL}\left( f_Q^{(a)} \| f_T^{(a)} \right) \tag{18}$$

where $f_Q^{(j,d)}$ and $f_T^{(j,d)}$ represent the MoGG models in the two images $I_Q$ and $I_T$, respectively, for the subband of the $j$th level and the $d$th direction, $d \in \{horizontal(H), vertical(V), diagonal(D)\}$. Similarly, $f_Q^{(a)}$ and $f_T^{(a)}$ designate the MoGG models for the approximations of the same images, respectively. In our experiments, we use $10^5$ samples for the KLD approximation. To quantify the performance of our approach against state-of-the-art methods, which use single distributions, we compared our method (which we denote MoGG + KLD + WA, where symbols $W$ and $A$ are used to denote high- and low-pass coefficients, respectively) with the following works:

1) GGD + KLD: This method, proposed in [22], models each wavelet subband using a single GGD. As stated before, the GGDs are calculated in a closed-form fashion.
2) GΓD + KLD: This method, proposed in [18], uses a single generalized Gamma distribution (GΓD) to model each wavelet subband. The authors of that work also developed a closed-form solution to measure the KLD between two GΓD models.
3) CONT + HMT: Here, we implemented the method developed in [44], where the authors use hidden Markov trees and MoG distributions to model the contourlet transform coefficients.
4) MoGG + KLD + W: In addition to the aforementioned methods, we implemented a version of our method where we discarded the distance between image approximations in the similarity measurement $S$ defined in (18). This enables us to quantify the contribution of the image approximation in texture retrieval.

First, we compared the aforementioned methods using different $J$ levels of the pyramidal decomposition of images. This part of the experiment aims at determining whether the number of decomposition levels influences the discrimination power of texture. To this end, we use three different levels of decomposition for each method and perform 100 retrievals for images randomly selected in each database. Table I shows the average

TABLE I
AVERAGE RETRIEVAL RATE IN THE TOP 16 IMAGES ETRIEVED IN 100 QUERIES IN THE (A) VISTEX AND (B) BRODATZ DATABASES

| Number of decomposition levels | GGD+KLD | GΓD+KLD | CONT+HMT | MoGG+KLD+W | MoGG+KLD+WA |
|---|---|---|---|---|---|
| $J = 1$ | 83.55 % | 86.81% | 88.83% | 90.41% | 95.91% |
| $J = 2$ | 85.24 % | 90.73% | 92.70% | 92.65% | 96.87% |
| $J = 3$ | 86.04 % | 90.03% | 93.15% | 94.15% | 98.10% |

**(A)**

| Number of decomposition levels | GGD+KLD | GΓD+KLD | CONT+HMT | MoGG+KLD+W | MoGG+KLD+WA |
|---|---|---|---|---|---|
| $J = 1$ | 84.11% | 88.90% | 92.10% | 93.40% | 95.20% |
| $J = 2$ | 89.10% | 92.61% | 93.45% | 96.40% | 98.40% |
| $J = 3$ | 89.94% | 93.03% | 94.12% | 97.32% | 99.30% |

**(B)**

retrieval accuracy in the VisTex and Brodatz databases, respectively. We measure the accuracy of each tested method by the percentage of *relevant* images found among the top $N$ retrieved images (100% retrieval precision means that a query retrieved all of the 16 subimages that originated from the same image). We note that this measure can be related as well to *false accept* or *false reject* ratios as they constitute the percentage of *relevant* images not retrieved by the algorithm. The following observations can be made on the obtained results:

1) Increasing the number of decomposition levels always leads to improvement in the retrieval accuracy. However, the first two levels embody most of the discrimination power in the high-pass subbands. Indeed, the third level brings little improvement in the retrieval accuracy (less than 1%) when combined with the first two levels.
2) Our experiments agree with the conclusion in [18], which used the Brodatz data set in their evaluation. Indeed, $\mathrm{G\Gamma D}$ + KLD outperforms GGD + KLD in both VisTex and Brodatz databases by an average of 1.67%. Using contourlets as proposed in [44] outperforms both GGD + KLD and $\mathrm{G\Gamma D}$ + KLD methods by about 3% and 1.39%, respectively. Our method using only high-pass subbands outperforms the previous three methods in all the decomposition levels. The improvement is about 6.32% against GGD + KLD, 4.65% against $\mathrm{G\Gamma D}$ + KLD, and 2.92% against using contourlets.
3) Combining high- and low-pass subbands drastically improves the retrieval accuracy by about 2.80% over using only high-pass subbands. This demonstrates that an important amount of the discrimination power lies in the loss-pass subband. We note that, due to models limitation, this information is ignored by most of past methods using wavelets or contourlets modeling for texture discrimination.

Fig. 6 shows graphs illustrating a comparison of retrieval precision, as a function of the number of top matches considered. Fig. 6(a) and (b) show the average retrieval rate in the two databases of 100 random queries each, using different levels of decomposition $J$. We can clearly note that using the first two levels of the pyramidal decomposition ($J = 2$) considerably improves the retrieval performance over using only the first level ($J = 1$). Combining the image approximation with the high-pass subbands improves the retrieval performance for all the considered numbers of top matches. For instance, the average improvement is 03.5% for the VisTex database, and 01.3% for the Brodatz database, when $N \leq 20$. Using three levels of decomposition ($J = 3$) increases these numbers to 04.6% and 02.3% in the two databases, respectively. Finally, including the approximation subband increases the retrieval precision rate for all the considered numbers of top matches. For instance, for $N \leq 20$, the average improvement is 03.7% for the VisTex database and 02% for the Brodatz database.

Fig. 6(c) and (d) show the retrieval rates for the compared methods, namely, GGD + KLD, $\mathrm{G\Gamma D}$ + KLD, CONT + HMT, and MoGG + KLD + WA, using the VisTex and the Brodatz databases, respectively. Our experimental results agree with [18] in that $\mathrm{G\Gamma D}$ + KLD outperforms GGD + KLD for all the considered numbers of top image matches. On the other hand, the method CONT + HMT outperforms both GGD + KLD and $\mathrm{G\Gamma D}$ + KLD. Finally, our method combining high- and low-pass subbands considerably improves the retrieval performance over the previous methods. For instance, the average improvement when $N \leq 20$ is 13% against GGD + KLD, 06.4% against $\mathrm{G\Gamma D}$ + KLD, and 03.9% against CONT + HMT, using the VisTex database. The equivalent numbers using the Brodatz database are 10.85%, 04.5%, and 02.5%, respectively. From these results, we can assess the contribution of our method. Indeed, using MoGG modeling with only high-pass subbands already improves state-of-the-art methods based on the same subbands. Moreover, combining high- and low-pass subbands, which is a possibility allowed in our paper

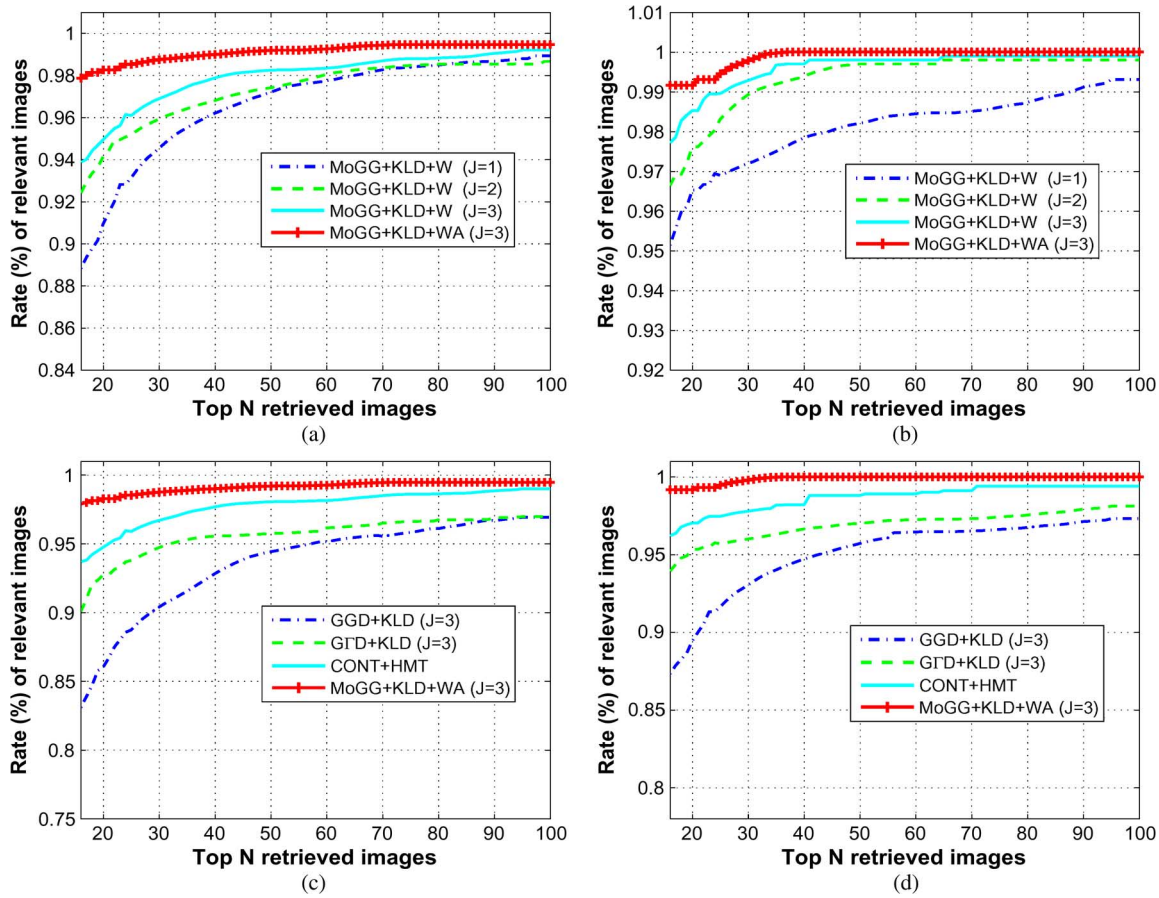


Fig. 6. Average retrieval rate of 100 queries as function of the number of top matches considered for the [(a) and (c)] VisTex and [(b) and (d)] Brodatz databases. For each database, the first row shows retrieval performance of our method using different levels of decomposition $J$ and the image approximation. The second row shows retrieval performances for the methods GGD + KLD, GΓD + KLD, CONT + HMT, and MoGG + KLD + WA.

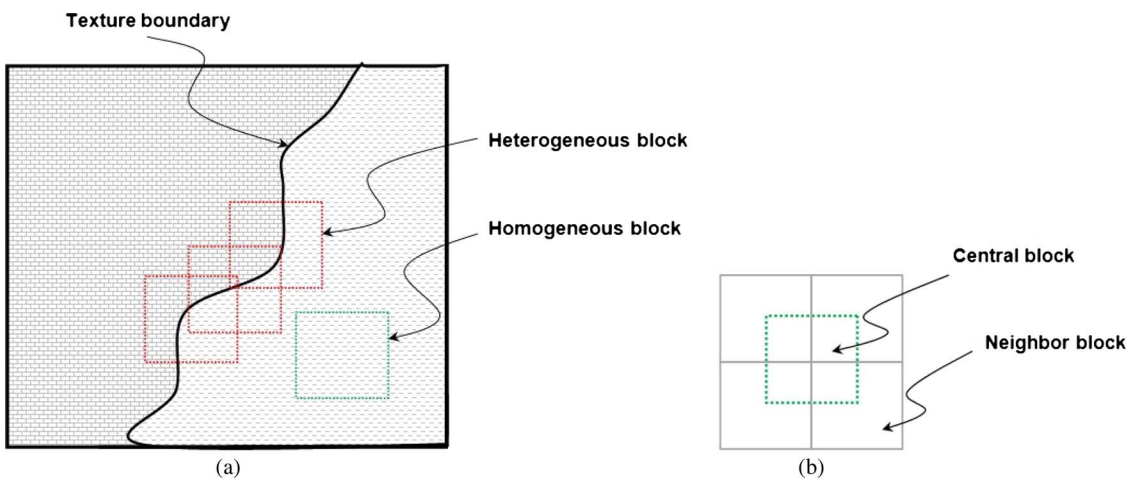


Fig. 7. Illustration of our block-based texture segmentation. (a) Boundary localization using heterogeneous blocks. (b) Neighborhood system used for determining block homogeneity.

and not the others, also brings a significant improvement of the results.

### *B. Application to Texture Segmentation*

Texture segmentation plays an important role in computer vision and pattern recognition. The topic of texture segmentation has been investigated for the last four decades where numerous approaches have been proposed. We emphasize that the goal here is rather to demonstrate the application of the proposed modeling method for texture discrimination. Thus, giving an exhaustive review about texture segmentation is beyond the scope of this paper.

Our procedure for texture segmentation is based on extracting heterogeneous blocks that likely lie on the boundaries

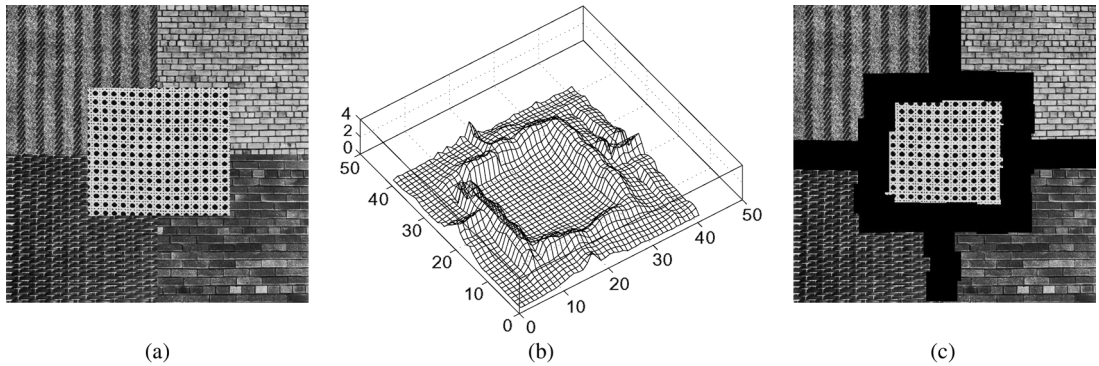


Fig. 8. Illustration of the procedure for finding texture boundary blocks. (a) Original image, (b) KLD distance map for local block neighborhoods, and (c) distance thresholding.

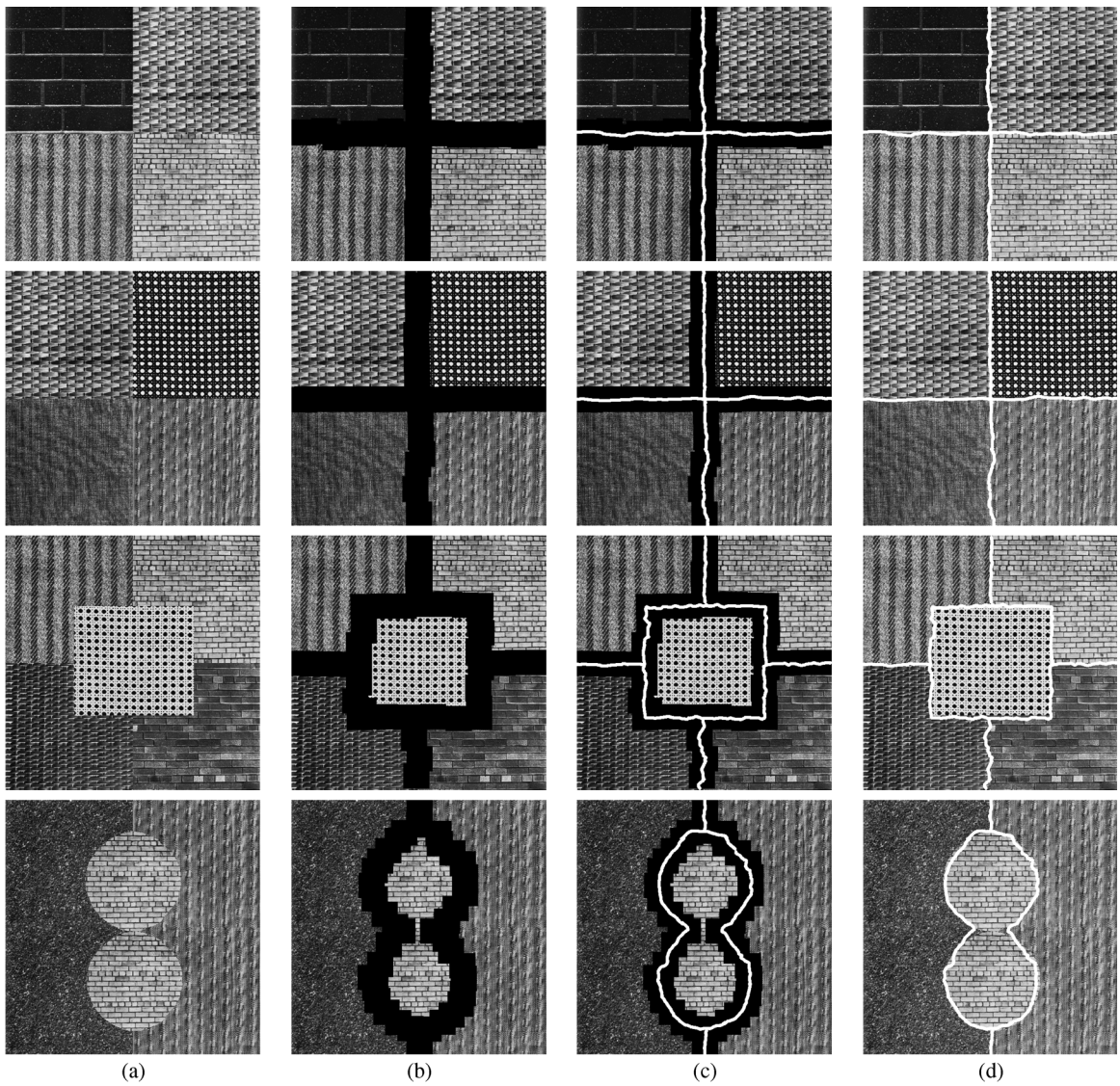


Fig. 9. Examples of texture mosaic segmentation. (a) Original image, (b) heterogeneous block detection, and (c), (d) detected region boundaries.

of a mosaic texture. Given an image composed of different homogenous textures, we aim to detect the strongest discontinuities of texture homogeneity, which likely correspond to region boundaries. Fig. 7 illustrates how the texture boundaries are extracted using our approach. Using a sliding window, our procedure starts by estimating the MoGG parameters for the different subbands for each window location. Among different window sizes, $65 \times 65$ pixels gives the best segmentation quality. Next, we calculate the average distance of each block with its eight neighbor blocks, as illustrated in Fig. 7(b). A

distance map is then calculated for the mosaic image using the similarity measurement defined in (18). Fig. 8(b) shows an example of distance map calculation of the shown example. By this procedure, heterogeneous blocks can be identified by thresholding the distance map image, as shown in Fig. 7(c). Finally, texture boundaries are determined by joining adjacent centers of heterogeneous blocks.

Fig. 9 shows some examples of texture segmentation. The texture mosaics are created using images from the Brodatz database. We show for each image the segmentation result after executing the described procedure. As demonstrated by these examples, the approach is capable of accurately segmenting images with arbitrary numbers of textures.

## VI. Computational Complexity

Considering the MML model selection algorithm illustrated in Alg. 1, each iteration for the estimation of a $K$-component MoGG model using $N$ data points requires the expectation and maximization steps in the EM algorithm [3]. The mixture parameters are initialized using fuzzy c-means, which has complexity $O(KN)$. The computation time of the two EM steps altogether and the calculation of the elements of the information matrix calculated using (19) through (26) are also of complexity $O(KN)$. The computation time of the similarity measurement in (18) is $O(JKn)$, where $J$ is the number of decomposition levels and $n$ is the number of sampled data. It follows that the complexity of performing a query in a database of size $\mathcal{N}$ is $O(JK\mathcal{N}n)$. Finally, for the segmentation application, given a texture image of size $Z = H \times L$ pixels and a block size of $B = W \times W$ pixels, the estimation of the block models is of complexity $O(KBZ)$, and the texture boundary localization is of complexity $O(KZJn)$.

Currently, all our experiments are performed in the MATLAB environment running on a workstation with 2.8-GHz Intel processor. In the texture retrieval application, the average central-processing-unit time for a query is about 15 s, whereas the equivalent amount of time is 1.2 s using the closed-form KLD proposed in [22]. In the segmentation application, the region boundary detection of an image of size 512 × 512 pixels takes about 3 min using the aforementioned computer configuration.

## VII. Conclusion and Discussion

We have introduced a new statistical framework for texture characterization using MoGG modeling. Our approach allows for leveraging the power of low- and high-pass wavelet coefficients and provides a high accurate description for their distribution. The experiments demonstrated the positive impact of this increased accuracy on the texture retrieval performance. Moreover, we proposed an estimation of the MoGG of each subband, which is based on the MML principle, which has a sound theoretical justification. This increasing accuracy certainly comes with a price, which is the computation time that is mainly incurred by the KLD approximation, since a closed form is not possible to derive for the MoGG. Currently, an average query time takes on average ten times the amount it takes for a closed-form solution. However, this is not too much a limitation compared with the increased accuracy. We demonstrated by experiments that our approach significantly improves texture discrimination and retrieval accuracy over recent state-of-the-art methods using closed-form solutions of the KLD for single distributions. Our approach also very favorably compares with methods based on the contourlet transform for texture representation.

Finally, we emphasize the fact that, even if our results are obtained for the DWT, our approach can be readily applied to other similar filtering schemes such as wavelet packets and contourlet transforms. Another promising aspect of improving this research is the extension of our MoGG modeling to take color information into account. For this purpose, using a mixture of multivariate GGD, as proposed in [3], can be a viable solution. Finally, investigating applications for general image classification and retrieval problems will be an interesting research path to pursue.

## Appendix I
## Calculation of the Fisher Information Matrix $I_c(\vec{\theta})$

We have the following derivatives that are required to calculate $I_c(\vec{\theta})$:

$$-\frac{\partial \log\left(p(\mathcal{X}_j|\vec{\theta})\right)}{\partial \mu_j} = \frac{\beta_j}{\sigma_j^{\beta_j}} \sum_{i=l}^{r_j} \text{sign}(x_i-\mu_j)|x_i-\mu_j|^{\beta_j-1} \quad (19)$$

$$-\frac{\partial \log\left(p(\mathcal{X}_j|\vec{\theta})\right)}{\partial \sigma_j} = \sum_{i=l}^{r_j}\left(\frac{1}{\sigma_j} + \frac{\lambda_j A(\beta_j)}{\sigma_j^{\beta_j+1}}|x_i-\mu_j|^{\beta_j}\right) \quad (20)$$

$$\begin{aligned}-\frac{\partial \log\left(p(\mathcal{X}_j|\vec{\theta})\right)}{\partial \beta_j} =& -\frac{n_j}{\beta_j} + \frac{3n_j}{2\beta_j^2}(\Psi(3/\beta_j)-\Psi(1/\beta_j)) \\ & - A(\beta_j)\sum_{i=l}^{r_j}\left[\left|\frac{x_i-\mu_j}{\sigma_j}\right|^{\beta_j}\log\left|\frac{x_i-\mu_j}{\sigma_j}\right|\right] \\ & - B(\beta_j)\sum_{i=l}^{r_j}\left|\frac{x_i-\mu_j}{\sigma_j}\right|^{\beta_j} \end{aligned} \quad (21)$$

where $r_j = l + n_j - 1$ and $\text{sign}(x)$ is equal to 1, if $x \geq 0$, and $-1$ otherwise. $A(\beta_k)$ is defined in (1), and $B(\beta_j)$ is defined as follows:

$$\begin{aligned}B(\beta_j) =& \frac{\partial A(\beta_j)}{\partial \beta_j} \\ =& \frac{1}{2}\left[\log\left(\frac{\Gamma(3/\beta_j)}{\Gamma(1/\beta_j)}\right)\right] - \frac{3\Psi(3/\beta_j)}{2\beta_j} + \frac{3\Psi(1/\beta_j)}{2\beta_j}. \end{aligned} \quad (22)$$

We also have

$$\begin{aligned}&-\frac{\partial^2 \log\left(p(\mathcal{X}_j|\vec{\theta})\right)}{\partial \mu_j \partial \mu_k} \\ &= \begin{cases} \frac{\beta_j(1-\beta_j)}{\sigma_j^{\beta_j}} A(\beta_j)\sum_{i=l}^{r_j}|x_i-\mu_j|^{\beta_j-2}, & \text{if } j=k \\ 0, & \text{if } j \neq k. \end{cases} \end{aligned} \quad (23)$$

$$-\frac{\partial^2 \log\left(p(\mathcal{X}_j|\vec{\theta})\right)}{\partial\beta_j\partial\beta_k} = \begin{cases} \frac{n_j}{\beta_j^2} + \frac{3n_j}{\beta_j^3}\left(\Psi(1/\beta_j) - \Psi(3/\beta_j)\right) - \frac{9n_j}{\beta_j^4}\Psi'(3/\beta_j) - 2B(\beta_j)\sum_{i=l}^{r_j}\left[\left|\frac{x_i-\mu_j}{\sigma_j}\right|^{\beta_j}\log\left|\frac{x_i-\mu_j}{\sigma_j}\right|\right] \\ \quad -C(\beta_j)\sum_{i=l}^{r_j}\left|\frac{x_i-\mu_j}{\sigma_j}\right|^{\beta_j} - A(\beta_j)\sum_{i=l}^{r_j}\left[\left|\frac{x_i-\mu_j}{\sigma_j}\right|^{\beta_j}\left(\log\left|\frac{x_i-\mu_j}{\sigma_j}\right|\right)^2\right] + \frac{3n_j}{2\beta_j^4}\Psi'(1/\beta_j), & \text{if } j=k \\ 0, & \text{if } j\neq k \end{cases} \tag{25}$$

In the same vein, we obtain the following for the scale parameters:

$$-\frac{\partial^2 \log\left(p(\mathcal{X}_j|\vec{\theta})\right)}{\partial\sigma_j\partial\sigma_k} = \begin{cases} -\frac{n_j}{\sigma_j^2} - A(\beta_j)\frac{\beta_j(1+\beta_j)}{\sigma_j^{\beta_j+2}}\sum_{i=l}^{r_j}|x_i-\mu_j|^{\beta_j}, & \text{if } j=k \\ 0, & \text{if } j\neq k \end{cases} \tag{24}$$

and for the shape parameters, see (25), shown at the top of the page.

That is, with

$$\begin{aligned} C(\beta_j) &= \frac{\partial^2 A(\beta_j)}{\partial^2\beta_j} \\ &= \frac{1}{2}A(\beta_j)\left[\frac{1}{2\beta_j^2}\left(\Psi(1/\beta_j) - 3\Psi(3/\beta_j)\right) + \frac{1}{2\beta_j^3}\left(9\Psi'(3/\beta_j) - \Psi'(1/\beta_j)\right)\right] \\ &\quad + \frac{1}{2}B(\beta_j)\left[\log\left(\frac{\Gamma(3/\beta_j)}{\Gamma(1/\beta_j)}\right) - \frac{3\Psi(3/\beta_j)}{2\beta_j} + \frac{\Psi(1/\beta_j)}{2\beta_j}\right] \end{aligned} \tag{26}$$

where we have $\Psi(x) = \partial log(\Gamma(x))/\partial x$ and $\Psi'(x) = \partial^2 log(\Gamma(x))/\partial^2 x$.

## REFERENCES


[1] N. Abbadeni, "Computational perceptual features for texture representation and retrieval," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 236–246, Jan. 2011.

[2] B. Aiazzi, L. Alparone, and S. Baronti, "Estimation based on entropy matching for generalized Gaussian PDF modelling," *IEEE Signal Process. Lett.*, vol. 6, no. 6, pp. 138–140, Jun. 1999.

[3] M. S. Allili, N. Bouguila, and D. Ziou, "Finite general Gaussian mixture modelling and application to image and video foreground segmentation," *J. Electron. Imaging*, vol. 17, no. 1, p. 013 005, Jan.–Mar. 2008.

[4] M. S. Allili, D. Ziou, N. Bouguila, and S. Boutemedjet, "Image and video segmentation by combining unsupervised generalized Gaussian mixture modelling and feature selection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 10, pp. 1373–1377, Oct. 2010.

[5] M. S. Allili, "Wavelet-based texture retrieval using a mixture of generalized Gaussian distributions," in *Proc. IEEE Int. Conf. Pattern Recog.*, 2010, pp. 3143–3146.

[6] M. S. Allili, Similarity measurements between finite mixtures of generalized Gaussian distributions 2011, Tech. Rep..

[7] M. S. Allili and N. Baaziz, "Contourlet-based texture retrieval using a mixture of generalized Gaussian distributions," in *Proc. Comput. Anal. Images Patterns*, 2011, vol. 2, pp. 446–454.

[8] A. N. Akansu and P. R. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*. New York: Academic, 2000.

[9] M. Baccar, L.-A. Gee, and M. A. Abidi, "Reliable location and regression estimates with application to range image segmentation," *J. Math. Imaging Vis.*, vol. 11, no. 3, pp. 195–205, Dec. 1999.

[10] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.

[11] C. Bishops, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.

[12] G.-E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*. New York: Wiley, 1992.

[13] K. Brady, I. H. Jermyn, and J. Zerubia, "Texture analysis: An adaptive probabilistic approach," in *Proc. IEEE Int. Conf. Image Process.*, 2003, pp. 1045–1048.

[14] Texture database [Online]. Available: http://www.ux.uis.no/~tranden/brodatz.html

[15] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.

[16] S.-K. Choy and C.-S. Tong, "Supervised texture classification using characteristic generalized Gaussian density," *J. Math. Imaging Vis.*, vol. 29, no. 1, pp. 35–47, Sep. 2007.

[17] S.-K. Choy and C.-S. Tong, "Statistical properties of bit-plane probability model and its application to supervised texture classification," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1399–1405, Aug. 2008.

[18] S.-K. Choy and C.-S. Tong, "Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 281–289, Feb. 2010.

[19] R. Cossu, I. H. Jermyn, and J. Zerubia, "Texture analysis using probabilistic models of the unimodal and multimodal statistics of adaptive wavelet packet coefficients," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2004, pp. 597–600.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.

[21] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, no. 7, pp. 909–996, Oct. 1988.

[22] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and KL distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[23] 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, Dec. 2002.

[24] E. de Ves, X. Benavent, A. Ruedin, D. Acevedo, and L. Seijas, "Wavelet-based texture retrieval and modelling the magnitudes detail coefficients with generalized gamma distribution," in *Proc. IEEE Int. Conf. Pattern Recog.*, 2010, pp. 221–224.

[25] G. Fahmy, J. A. Black, Jr., and S. Panchanathan, "Texture characterization for joint compression and classification based on human perception in the wavelet domain," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1389–1396, Jun. 2006.

[26] G. Fan and X. G. Xia, "Wavelet-based texture analysis and synthesis using hidden Markov models," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 50, no. 1, pp. 106–120, Jan. 2003.

[27] M. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.

[28] J. Goldberger, S. Gordon, and H. Greenspan, "An efficient image similarity based on approximations of KL-divergence between two Gaussian mixtures," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2003, pp. 487–493.
[29] S. E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on Gabor filters," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1160–1167, Oct. 2002.
[30] M. Haindl and P. Vacha, "Illumination invariant texture retrieval," in *Proc. IEEE Int. Conf. Pattern Recog.*, 2006, pp. 276–279.
[31] D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis," in *Proc. IEEE Int. Conf. Image Process.*, 1995, vol. 3, pp. 648–651.
[32] J. R. Hershey and P. A. Olsen, "Approximating the Kullback–Leibler divergence between Gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2007, pp. 317–320.
[33] R. L. Joshi and T. R. Fischer, "Comparison of generalized Gaussian and Laplacian modelling in DCT image coding," *IEEE Signal Process. Lett.*, vol. 2, no. 5, pp. 81–82, May 1995.
[34] S. Kasaei, M. Deriche, and B. Boashash, "A novel fingerprint image compression technique using wavelets packets and pyramid lattice-vector quantization," *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1365–1378, Dec. 2002.
[35] X. Liu and D. Wang, "Texture classification using spectral histograms," *IEEE Trans. Image Process.*, vol. 12, no. 6, pp. 661–670, Jun. 2003.
[36] D.-J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[37] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[38] M. K. Mandal, T. Aboulnasr, and S. Panchanathan, "Fast wavelet histogram techniques for image indexing," *Comput. Vis. Image Understand.*, vol. 75, no. 1/2, pp. 99–110, Jul./Aug. 1999.
[39] B. S. Manjunathi and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
[40] M. Mirmehdi, X. Xie, and J. Suri, *Handbook of Texture Analysis*. London, U.K.: Imperial College Press, 2008.
[41] MIT Vision and Modelling Group, Vision Texture [Online]. Available: http://vismod.media.mit.edu/
[42] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity schemes," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 909–919, Apr. 1999.
[43] J. J. Oliver, R. A. Baxter, and C. S. Wallace, "Unsupervised learning using MML," in *Proc. Int. Conf. Mach. Learn.*, 1996, pp. 364–372.
[44] D. D.-Y. Po and M. N. Do, "Directional multiscale modelling of images using the contourlet transform," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1610–1620, Jun. 2006.
[45] T. Randen and J.-H. Husoy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 291–310, Apr. 1999.
[46] C. Robert, *The Bayesian Choice*, ser. Springer Texts in Statistics, 2nd ed. New York: Springer-Verlag, 2001.
[47] C. P. Robert and G. Casella, *Monte-Carlo Statistical Methods*. New York: Springer-Verlag, 2004.
[48] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distribution in subband decomposition of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.
[49] M. Tuceryan and A. K. Jain, "Texture analysis," in *The Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds., 2nd ed. , Singapore: World Scientific, 1998, pp. 207–248.
[50] G. Van De Wouwer, P. Scheunders, and D. Van Dyck, "Statistical texture characterization from discrete wavelet representation," *IEEE Trans. Image Process.*, vol. 8, no. 4, pp. 592–598, Apr. 1999.
[51] M. K. Varanasi and B. Aazhang, "Parametric generalized Gaussian density estimation," *J. Acoust. Soc. Amer.*, vol. 86, no. 4, pp. 1404–1415, Oct. 1989.
[52] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[53] G. Verdoolaege and P. Scheunders, "Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination," *Int. J. Comput. Vis.*, vol. 95, no. 3, pp. 265–286, Dec. 2011.
[54] C. S. Wallace, *Statistical and Inductive Inference by Minimum Message Length*, ser. Information Science and Statistics. New York: Springer-Verlag, 2005.
[55] C. S. Wallace and P. R. Freeman, "Estimation and inference by compact coding," *J. R. Stat. Soc. (Ser. B)*, vol. 49, no. 3, pp. 240–265, 1987.
[56] G.-S. Xia, J. Delon, and Y. Gousseau, "Shape-based invariant texture indexing," *Int. J. Comput. Vis.*, vol. 88, no. 3, pp. 382–403, Jul. 2010.

**Mohand Saïd Allili** (M'08) received the M.Sc. and Ph.D. degrees in computer science from the University of Sherbrooke, Sherbrooke, QC, Canada, in 2004 and 2008, respectively.

Since June 2008, he has been an Assistant Professor of computer science with the Department of Computer Science and Engineering, Université du Québec en Outaouais, Canada. His main research interests include computer vision and graphics, image processing, pattern recognition, and machine learning.

Dr. Allili was a recipient of the Best Ph.D. Thesis Award in engineering and natural sciences from the University of Sherbrooke for 2008 and the Best Student Paper and Best Vision Paper awards for two of his papers at the Canadian Conference on Computer and Robot Vision 2007 and 2010, respectively.