IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 1, 2019, accepted August 17, 2019, date of publication September 4, 2019, date of current version September 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939466

A Finite Mixture of Weibull-Based Statistical Model for Texture Retrieval in the Complex Wavelet Domain

HASSAN RAMI[1], AHMED DRISSI EL MALIANI[1], AND MOHAMMED EL HASSOUNI[2], (Member, IEEE)
[1]LRIT URAC 29, Faculty of Sciences, Mohammed V University, Rabat 10106, Morocco
[2]DESTEC, FLSHR, University of Mohammed V, Rabat 1014, Morocco

Corresponding author: Mohammed El Hassouni (mohamed.elhassouni@um5.ac.ma)

ABSTRACT This paper presents a new statistical model for texture retrieval in the complex wavelet domain. For this purpose, a finite mixture of Weibull distributions (MoWbl) is proposed to characterize the statistical distribution of magnitudes of complex wavelet coefficients. Despite the ability of the mixture model on capturing a wide range of distribution shapes, choosing an appropriate number of mixture components is a challenging task. To this end, we adopt an unsupervised learning of the model parameters based on the Figueiredo-Jain algorithm and maximum-likelihood estimates. As found in all retrieval statistical-based frameworks, the presence of a similarity measure is trivial. Generally, the failure of a retrieval mixture based system is closely related to the choice of the similarity measure that relies mainly on approximations of some divergences and distances. To overcome this limitation, we propose a canonical form of Weibull distribution which allows us to develop an analytic expression of Cauchy-Schwarz divergence (CSD) for MoWbl distributions. Experiments, conducted on three popular datasets, show that the proposed model yields better performance in terms of goodness-of-fit, retrieval, and execution time compared to some related statistical models for texture retrieval.

INDEX TERMS Texture retrieval, statistical analysis, finite mixture of weibulls, Cauchy-Schwarz divergence.

I. INTRODUCTION

Besides visual content in images such as color and shape, texture content is a very important feature for content-based image retrieval systems. Various texture analysis techniques [1]–[4] consider texture as the outcome of a deterministic dynamical system subject to random noise. The effectiveness of wavelet-based signatures has been recognized for texture discrimination since wavelet analysis is supported by the visual cortex studies [5]. Randen and Husoy [6] have compared various filtering approaches and have concluded that statistical features yield better performance for texture description than energy based methods. Indeed, since a texture can be considered as a realization of a random process, statistical-based models are more suitable to discriminate between textures. Specifically, parametric distributions have been successfully used in wavelet domain to describe the shape of normalized histograms. For instance, histograms of wavelet coefficients provided by the orthogonal wavelet transform (OWT) are distributed according to a generalized Gaussian distribution (GGD) [7], [8]. Histograms of wavelet coefficients are mostly non-Gaussian and have leptokurtic forms [9]. Moreover, some histograms can be asymmetric and may have a multimodal form such as the case for wavelet packets [10]. In this situations, a finite mixture model (FMM) [11] is suitable for a more accurate histogram fitting and a more precise texture image description. Allili proposed recently a finite mixture of generalized Gaussian (MoGG) distributions for texture modeling [12] where they modeled both wavelet detail and approximation subbands with a mixture rather than a single generalized Gaussian distribution (GGD). They showed that various histograms raise multimodal forms which are hard to fit using a single GGD. Later on, they have modeled the histograms of the Contourlet Transform (CT) in [13] to overcome some limitations of the

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

DWT using the same MoGG model. Li *et al.* [14] proposed a generalized Gamma mixture model (GΓMM) to analyze high-resolution synthetic aperture radar (SAR) images. They fit histograms of SAR images using an unsupervised learning algorithm to estimate the parameters of the mixture model. Recently, a Wishart mixture model (WMM) has been proposed for change detection in PolSAR images [15]. The polarimetric information in PolSAR images is represented by a complex vector and each data point in the image is represented by a covariance matrix. Hence, the Wishart mixture model is a good choice to model these covariance matrices. Biomedical imaging also benefits from the mixture distributions such as Rayleigh mixture model (RMM) proposed in [16] to characterize plaque in intravascular ultrasound images.

In a recent work [17], we proposed to use Cauchy-Schwarz divergence (CSD) [18] for MoGG models to obtain a closed-form expression of the similarity measure. This choice has been inspired by the work of Nielsen [19] where several closed-form expressions of CSD have been computed for numerous distributions that belong to the exponential family such as Laplacian, Rayleigh and Bernoulli. Although GGD is not part of the exponential family, we fixed the shape parameter to obtain a closed-form expression of the CSD. Indeed, we stressed out how a CSD is such a promising alternative for KLD since its mathematical expression allows a closed-form expression in case of MoGG with a fixed shape parameter. Starting from these interesting properties of the CSD, we aim to take advantage of the finite mixture modeling without being confronted to the approximation of the divergence between two finite mixture distributions. In order to achieve this, we choose to model histograms of magnitudes of complex wavelet coefficients. Complex wavelet transforms have gained attention in recent years because they provide interesting properties that are not present in real valued coefficients such as the shift invariance of the magnitude and a large orientation selectivity [20]–[23]. As any information retrieval problem, finding the appropriate statistical distribution that describes well the data is subject to many research studies. Complex coefficients have the advantage of providing two sources of information, magnitudes and phases. Commonly in the literature, magnitudes have been characterized using different positive distributions such as Weibull [24], Gamma distribution [25] and Generalized Gamma distribution [23]. Since magnitudes are positive values, it is obvious that all of this positive distributions are more suitable than a GGD to take advantage of the exponential family distributions and model the marginal behavior of complex wavelet coefficients histograms as well. Additionally, Weibull distribution has been successfully used to model magnitudes of complex wavelet coefficients in some previous texture retrieval models [22], [24], and [26]–[28]. Angles of complex coefficients are generally exploited using circular distributions such as Wrapped Cauchy and von Mises to characterize the relative phase [29], [30]. From all of these reasons stems our motivation to model magnitudes of complex wavelet coefficients using a finite mixture of Weibull (MoWbl) distributions to retrieve textured images. The proposed MoWbl model characterizes the normalized histograms of the magnitudes of complex wavelet coefficients issued from the dual-tree complex wavelet transform (DTCWT [31]) and the magnitudes computed with horizontal and vertical detail sub-bands of a discrete wavelet transform (DWT [7]). Based on the unsupervised algorithm proposed by Figueiredo and Jain [32], we develop equations for the minimum message length (MML) based estimators for a finite mixture of Weibull distribution to obtain the optimal parameter set including the number of components. Since the Weibull distribution can be part of the exponential family, we propose herein a canonical parametrization for the Weibull distribution. This technique has also been used in [15] for Wishart distribution except that this distribution is only applicable in case of covariance matrices modeling. Thanks to this parametrization, we develop a closed-form expression of CSD between MoWbl models. An in-depth experimental process has been applied to the proposed model to assess its goodness-of-fit, its retrieval accuracy and its similarity measurement execution time. Texture retrieval experiments have been conducted on three popular texture datasets; VisTex [33], Brodatz [34] and ALOT [35] datasets.

To sum up, the main contribution of this paper is twofold: i) We propose a Weibull mixture model for magnitudes of complex wavelet coefficients which parameters are estimated using a new MLE based unsupervised learning. ii) By providing a canonical form of the proposed model, we compute an analytic expression of Cauchy-Schwarz divergence between two mixtures of Weibull distributions.

This paper is organized as follows. In Section II, we elaborate the parameters estimation algorithm for the MoWbl distribution model. The proposed closed-form expression of CSD is detailed in Section III while Section IV focuses on the results of several retrieval experiments on textured images datasets. Finally, Section V concludes the paper.

II. MIXTURE OF WEIBULL DISTRIBUTIONS MODEL

The Weibull distribution function which is a two-parameter function is expressed mathematically as:

$$p(x \mid \alpha, \beta) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-(\frac{x}{\alpha})^{\beta}} \tag{1}$$

where $\beta > 0$ is a shape parameter, $\alpha > 0$ is a scale parameter.

In statistics, a mixture density is a probability density function which is a convex linear combination of probability density functions. The mixture of Weibull distributions can be defined as a weighted sum of Weibull probability density functions:

$$p(x \mid \Theta) = \sum_{i=1}^{M} \pi_i p(x \mid \theta_i) \tag{2}$$

where π_i represents the weight of the *ith* Weibull component within the mixture, also called the mixing weight, and

$\theta_i = (\alpha_i, \beta_i)$ corresponds to the parameters of that component; the weights sum to 1, $\sum_{i=1}^{M} \pi_i = 1$. Our finite mixture model is characterized by a set of $3 \times M$ parameters: $\Theta = \{\pi_1, \alpha_1, \beta_1, \ldots, \pi_M, \alpha_M, \beta_M\}$, where M is the number of mixture components.

The feature extraction step aims to obtain the optimal estimate of the set of parameters Θ for a given set of sub-band coefficient magnitudes.

A. MAXIMUM-LIKELIHOOD BASED PARAMETERS ESTIMATION

Given a set of N coefficient magnitudes $\mathcal{X} = (x_1, \ldots, x_N)^T$ from a specific wavelet details sub-band, the feature extraction procedure consists of finding the optimal set of $3 \times M$ parameters, $\hat{\Theta}$, that best fits the histogram of coefficients magnitudes. We use the maximum-likelihood estimator (ML) which implies maximizing the log-likelihood function, that is:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}}\ \mathcal{L}(\mathcal{X}|\Theta) \tag{3}$$

with

$$\mathcal{L}(\mathcal{X} \mid \Theta) = \log p(\mathcal{X}|\Theta) = \log \prod_{i=1}^{N} \sum_{j=1}^{M} \pi_j p(x_i|\theta_j) \tag{4}$$

$$\mathcal{L}(\mathcal{X} \mid \Theta) = \sum_{i=1}^{N} \log \left(\sum_{j=1}^{M} \pi_j p(x_i|\theta_j) \right) \tag{5}$$

To solve this maximization problem we use the EM algorithm [11] iteratively through two steps: i) an expectation step (E-step), where the expectancy of the complete log-likelihood function (5) is calculated with respect to the observed information $\mathcal{X}$ and the current state of the model parameters estimate $\hat{\Theta}^{(t)}$, and ii) a maximization step (M-step) where mixture parameters are updated so that to increase the expectation of the complete log-likelihood. EM algorithm interprets the observed data $\mathcal{X}$ as incomplete data with the missing information being a corresponding set of labels $\mathcal{Z} = \{z_{ij} | i = 1, \ldots, N \text{ and } j = 1, \ldots, M\}$ where $z_{ij} = 1$ indicates that the coefficient x_i is produced by the j-*th* component of the mixture and $z_{ij} = 0$ otherwise. Hence, $\mathcal{X}$ is augmented by $\mathcal{Z}$ to form a complete data set. The resulting complete-data log-likelihood is given by:

$$\mathcal{L}(\mathcal{X} \mid \Theta) = \sum_{i=1}^{N} \sum_{j=1}^{M} z_{ij} \log \left(\pi_j p(x_i \mid \theta_j) \right) \tag{6}$$

The following equations correspond to the tasks performed in each step:

1) **E-step:** Compute the conditional expectation of the complete log-likelihood:

$$Q(\Theta, \hat{\Theta}^{(t)}) = \mathbb{E}[\log (p(\mathcal{X}, \mathcal{Z} \mid \Theta)) \mid \mathcal{X}, \hat{\Theta}^{(t)}] \tag{7}$$

given x and the current estimate $\hat{\Theta}^{(t)}$.

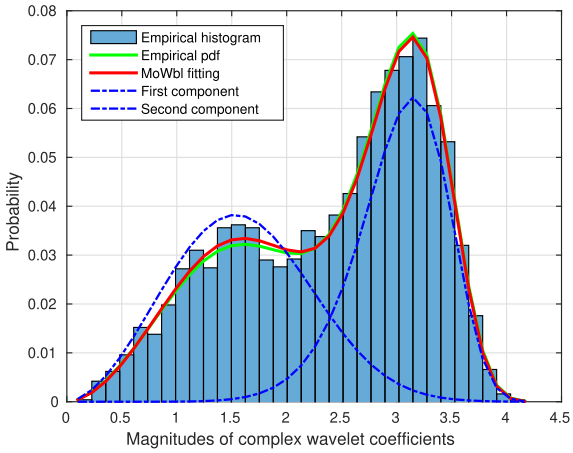


FIGURE 1. Fitting of a mixture of two Weibull distributions: $\{\pi_1 = 0.59, \alpha_1 = 3.2, \beta_1 = 8.61\}$ and $\{\pi_2 = 0.41, \alpha_2 = 1.8, \beta_2 = 2.78\}$.

2) **M-step:** Update the parameter estimates according to:

$$\hat{\Theta}^{(t+1)} = \underset{\Theta}{\operatorname{argmax}}\ Q(\Theta, \hat{\Theta}^{(t)}) \tag{8}$$

The two steps are iteratively processed until convergence is achieved.

In our model, which concerns the mixture of Weibull distributions, the updating equations are as follows:

$$\hat{\pi}_j^{(t+1)} = \frac{\sum_{i=1}^{N} p(z_{ij} \mid x_i, \Theta^{(t)})}{\sum_{i=1}^{N} \sum_{j=1}^{M} p(z_{ij} \mid x_i, \Theta^{(t)})} = \frac{1}{N} \sum_{i=1}^{N} p(z_{ij} \mid x_i, \Theta^{(t)}) \tag{9}$$

$$\hat{\alpha}_j^{(t+1)} = \left(\frac{\sum_{i=1}^{N} p(z_{ij} \mid x_i, \Theta^{(t)}) x_i^{\beta_j^{(t)}}}{\sum_{i=1}^{N} p(z_{ij} \mid x_i, \Theta^{(t)})} \right)^{\frac{1}{\beta_j^{(t)}}} \tag{10}$$

The equation for the shape parameter β_j is solved with Newton-Raphson method. The corresponding equation is as follows:

$$\hat{\beta}_j^{(t+1)} \simeq \beta_j^{(t)} - \frac{\partial Q(\Theta \mid \Theta^{(t)})}{\partial \beta_j^{(t)}} \left(\frac{\partial^2 Q(\Theta \mid \Theta^{(t)})}{\partial {\beta_j^{(t)}}^2} \right)^{-1} \tag{11}$$

Θ designates here the set of parameters of the mixture as defined in (2). The posterior probability $p(z_{ij}|x_i, \Theta^{(t)})$ is given by:

$$p(z_{ij} \mid x_i, \Theta^{(t)}) = \frac{\pi_j^{(t)} p(x_i \mid \theta_j^{(t)})}{\sum_{j=1}^{M} \pi_j^{(t)} p(x_i \mid \theta_j^{(t)})} \tag{12}$$

Fig. 1 illustrates an example of a mixture of two Weibull distributions for which we estimate parameters using the EM algorithm. The estimated probability density function is very similar to the original one.

B. UNSUPERVISED LEARNING OF MOWBL

In real image retrieval applications, the exact number of mixture components is unknown in advance. A mixture model of an inaccurate number of components tends to overfit or

underfit the data. This problem is often resolved with a model selection criterion such as Akaike's information criterion (AIC) [36], Bayesian information criterion (BIC) [37], minimum message length (MML) [38], minimum description length (MDL) [39] or integrated completed likelihood (ICL) [40]. All of these criteria assume that the number of components is fixed for each model and they choose the best model by minimizing the following equation:

$$\hat{\Theta}_M = \underset{M}{\operatorname{argmin}} \left\{ \mathcal{C}(\hat{\Theta}_M), M = M_{min}, \ldots, M_{max} \right\} \quad (13)$$

where a set of parameters $\hat{\Theta}_M$ is estimated for each number of components M which varies from M_{min} to M_{max}, then the best model is selected according to the underlying selection criteria $\mathcal{C}(\hat{\Theta}_M)$. Such an algorithm is obviously computationally expensive and is not straightforward as it implies a large number of EM executions for each number of components, which takes an important execution time to converge. To overcome this issue, Figueiredo and Jain [32] have proposed an effective algorithm which aims to find directly the "best" overall model in the entire set of available models. Rather than using EM to compute a set of candidate models, their algorithm implements the MML criterion's equation using a variant of EM. In our work, we adopted the following criterion from Figueiredo and Jain [32]:

$$\hat{\Theta}_M = \underset{M}{\operatorname{argmin}} \Big\{ -\log(h(\Theta)) - \log(p(\mathcal{X}|\theta)) + \frac{1}{2}\log(|F(\Theta)|) + \frac{c}{2}(1 + \log(\frac{1}{12})) \Big\} \quad (14)$$

where $F(\Theta)$ is the expected Fisher Information (FI) matrix, $|F(\Theta)|$ denotes its determinant, and $c = 3M$ is the number of free parameters in MoWbl distribution.

Fisher Information matrix is the Hessian matrix of the logarithm of minus the likelihood of the mixture. In our case, we deal with a $2M \times 2M$ Hessian matrix defined by the following elements:

$$H_{l_1\, l_2} = -\frac{\partial^2}{\partial\theta_{l_1}\theta_{l_2}} \log(p(\mathcal{X} \mid \Theta)) \quad (15)$$

Such a Hessian matrix makes difficult to derive an analytic form of the MML criterion for mixtures, so we follow the approximation proposed by Baxter and Oliver [41]. The determinant of the FI matrix is approximated by the product of the determinant of the FI matrix for each component, times the determinant of the FI matrix of the mixture weights π_j:

$$|F(\Theta)| = |F(\pi)| \prod_{j=1}^{M} |F_j(\alpha_j, \beta_j)| \quad (16)$$

where the FI matrix for a single Weibull distribution with scale parameter α_j and shape parameter β_j is [42]:

$$F_j(\alpha_j, \beta_j) = \begin{bmatrix} N(\frac{\beta_j}{\alpha_j})^2 & -b\frac{N}{\alpha_j} \\ -b\frac{N}{\alpha_j} & a\frac{N}{\beta_j^2} \end{bmatrix} \quad (17)$$

where $a = 1.823600$ and $b = 0.422784$ are two constants. The mixing weights can be viewed as being the parameters of a multinomial distribution. In this case, the determinant of the FI matrix of the mixing weights is $|F(\pi)| = \frac{N}{\prod_{j=1}^{M} \pi_j}$ ([41]), where N is the number of data elements.

To obtain an analytic form of the prior distribution $h(\Theta)$, we assume that the weights π_j and the components parameters are independent. Based on this independence, the prior is given as

$$h(\Theta) = h(\pi_1, \ldots, \pi_M) \prod_{j=1}^{M} h(\theta_j) \quad (18)$$

The two densities $h(\pi_1, \ldots, \pi_M)$ and $h(\theta_j)$ are obtained by using a Jeffrey's non-informative prior [43] as $h(\theta_j) = \sqrt{|F(\theta_j|}$ and $h(\pi_1, \ldots, \pi_M) = \sqrt{|F(\pi)|}$. By substituting (18) and (17) into (14), we finally obtain the following incomplete-data penalized log-likelihood function [44]:

$$\hat{\Theta}_M = \underset{M}{\operatorname{argmin}}\, \mathcal{L}_{MML}(\mathcal{X}, \Theta) \quad (19)$$

with

$$\mathcal{L}_{MML}(\mathcal{X}, \Theta) = \frac{1}{2}\sum_{j=1}^{M} log(\pi_j) + \frac{1 - K_{nz}}{2} log(a - b^2) + \frac{3K_{nz}}{2}\left(1 + log(\frac{1}{12})\right) - log(p(X \mid \Theta)) \quad (20)$$

where $K_{nz} = \sum_j [\pi_j > 0]$ denotes the number of non-zero probability components.

Now that we have defined a MML based criterion, we integrate it into the EM algorithm by replacing the Q function by the following one:

$$\begin{aligned} Q_{MML}(\Theta, \hat{\Theta}^{(t)}) = & \sum_{i=1}^{N}\sum_{j=1}^{M} p(z_{ij} \mid x_i, \Theta^{(t)}) \log(\pi_j) \\ & + \sum_{i=1}^{N}\sum_{j=1}^{M} p(z_{ij} \mid x_i, \Theta^{(t)}) \log(p(x_i \mid \theta_j)) \\ & - \frac{1}{2}\sum_{j=1}^{M} log(\pi_j) - \frac{1 - K_{nz}}{2} log(a - b^2) \\ & - \frac{3K_{nz}}{2}\left(1 + log(\frac{1}{12})\right) \end{aligned} \quad (21)$$

Based on this integration of the MML criterion on the Q function, we derive the new ML estimators of the mixture parameters in both the E- and M-step. From (21), we note that none of the new added terms depends on the component parameters θ_j and only the term $-(1/2)\sum_{j=1}^{M} log(\pi_j)$ depends on the mixing weights π_j. Thus, no modifications are required for the estimators of α_j and β_j in (9) and (11) respectively. By proceeding to the derivation of (21), we obtain the new

updated estimate of π_j:

$$\hat{\pi}_j^{(t+1)} = \frac{max\left\{0, \sum_{i=1}^{N} p(z_{ij} \mid x_i, \Theta^{(t)}) - \frac{1}{2}\right\}}{\sum_{j=1}^{M} max\left\{0, \sum_{i=1}^{N} p(z_{ij} \mid x_i, \Theta^{(t)}) - \frac{1}{2}\right\}} \quad (22)$$

The *max* operator in this expression can make weak components irrelevant since any component for which $\pi_j = 0$ does not contribute to the log-likelihood. This operation, called *component annihilation* [32], reduces the number of non-zero components K_{nz} during the parameter estimation process which can avoid approaching the boundary of the parameter space. The unsupervised learning algorithm for the MoWbl is summarized in Algorithm 1. Complexity of this algorithm is $O(t_{max} M_{max}^2)$, where M_{max} is the maximum number of components, t_{max} is the number of maximum EM iterations.

Algorithm 1 EM-FJ-MoWbl

Require: $\mathcal{X}, N, M_{min}, M_{max}, \epsilon, t_{max}$
Ensure: Mixture model in $\hat{\Theta}_{best}$
Initialization
$t \leftarrow 1,\ K_{nz} \leftarrow M_{max}, \mathcal{L}_{min} \leftarrow +\infty$
Calculate $\hat{\pi}^{(0)} = \frac{1}{M_{max}}$ for $i = 1, \ldots, K_{nz}$
Set random initial scales and shapes: $\{\hat{\theta}_1^{(0)}, \ldots, \hat{\theta}_{M_{max}}^{(0)}\}$
Calculate $p(z_{im} \mid x_i, \hat{\Theta}^{(0)})$ using (12)
Main loop
while $K_{nz} \geq M_{min}$ **do**
  repeat
    for $m = 1$ to K_{nz} **do**
      Calculate $p(z_{im} \mid x_i, \hat{\Theta}^{(t)})$ using (18)
      Calculate $\hat{\pi}_m^{(t)}$ using (12), for $i = 1, \ldots, N$
      $\{\hat{\pi}_1, \ldots, \hat{\pi}_{K_{nz}}\} \leftarrow \{\hat{\pi}_1, \ldots, \hat{\pi}_{K_{nz}}\} \times \left(\sum_{j=1}^{K_{nz}} \hat{\pi}_j\right)^{-1}$
      if $\hat{\pi}_m^{(t)} > 0$ **then**
        Update $\hat{\theta}_m^{(t)}$ using (11) and (9) successively
        Calculate $p(x_i|\hat{\theta}_m)$ for $i = 1, \ldots, N$
      else
        $K_{nz} \leftarrow K_{nz} - 1$
      end if
    end for
    $\hat{\Theta}^{(t)} \leftarrow \{\hat{\pi}_1^{(t)}, \hat{\theta}_1^{(t)}, \ldots, \hat{\pi}_{K_{nz}}^{(t)}, \hat{\theta}_{K_{nz}}^{(t)}\}$
    Calculate $\mathcal{L}_{MML}(\mathcal{X}, \hat{\Theta}^{(t)})$ using (20)
    $t \leftarrow t + 1$
  until $\mathcal{L}_{MML}(\mathcal{X}, \hat{\Theta}^{(t-1)}) - \mathcal{L}_{MML}^{(t)}(\mathcal{X}, \hat{\Theta}^{(t)}) \leq |\mathcal{L}_{MML}(\mathcal{X}, \Theta^{(t-1)})| \times \epsilon$ and $t \leq t_{max}$
  if $\mathcal{L}_{MML}(\mathcal{X}, \Theta^{(t)}) \leq \mathcal{L}_{min}$ **then**
    $\mathcal{L}_{min} \leftarrow \mathcal{L}_{MML}(\mathcal{X}, \Theta^{(t)})$
    $\hat{\Theta}_{best} \leftarrow \hat{\Theta}(t)$
  end if
  $m^* \leftarrow \underset{m}{\text{argmin}}\{\hat{\pi}_m > 0\}, \hat{\pi}_{m^*} \leftarrow 0, K_{nz} \leftarrow K_{nz} - 1$
end while

III. CAUCHY-SCHWARZ DIVERGENCE

Given two MoWbl models $p(x \mid \Theta^{(1)}) = \sum_{j=1}^{K} \pi_j p(x \mid \theta_j^{(1)})$ and $q(x \mid \Theta^{(2)}) = \sum_{j=1}^{M} \omega_j p(x \mid \theta_j^{(2)})$, the KLD between p and q is defined as:

$$\begin{aligned} \mathrm{KLD}(p||q) &= \int p(x \mid \Theta^{(1)}) \log\left(\frac{p(x \mid \Theta^{(1)})}{q(x \mid \Theta^{(2)})}\right) dx \\ &= \int \sum_{i=1}^{K} \pi_i p(x \mid \theta_i^{(1)}) \log\left(\frac{\sum_{j=1}^{K} \pi_j p(x \mid \theta_j^{(1)})}{\sum_{j=1}^{M} \omega_j p(x \mid \theta_j^{(2)})}\right) dx \end{aligned} \quad (23)$$

Note that in this expression of KLD, the logarithm prevents the marginalization of x out by the integral because it separates the integral and the summations over mixture components. Hence, it prevents the development of an analytic expression of this divergence between finite mixture distributions. The same problem has been pointed out in case of the mixture of Gaussians [45] and the mixture of generalized Gaussians [17]. This particularity prevents the mathematical development from obtaining a closed-form expressions for KLD in case of finite mixture models. Several approximation techniques have been used like Monte-Carlo integration [12], variational approximation [46] and matching based approximation [47]. Recently, Nielson and Sun [48] proposed an algorithm to build a closed-form formula that guarantees a lower and upper bound on the KLD for mixtures of distributions. The main issue with these approximation techniques is the increase in computation time of the similarity measurement which is caused by the evaluation of KLD integrals. In what follows, we introduce an analytic expression of Cauchy-Schwarz divergence.

Cauchy-Schwarz divergence (CSD) is based on the Cauchy-Schwarz inequality of density functions: $0 < (\int p(x \mid \Theta^{(1)}) q(x \mid \Theta^{(2)}) dx)^2 \leq \int p^2(x \mid \Theta^{(1)}) dx \int q^2(x \mid \Theta^{(2)}) dx$. For two pdfs $p(x \mid \Theta^{(1)})$ and $q(x \mid \Theta^{(2)})$ it is defined as follows [18]:

$$\mathrm{CSD}(p||q) = -\log \frac{\int p(x \mid \Theta^{(1)}) q(x \mid \Theta^{(2)}) dx}{\sqrt{\int p(x \mid \Theta^{(1)})^2 dx \int q(x \mid \Theta^{(2)})^2 dx}} \quad (24)$$

Compared to KLD, the CSD has the advantage of being symmetric in its arguments and verifies the triangle inequality. It is also always positive such that $0 \leq CSD(p||q) < \infty$ for any two pdfs p and q, and vanishes if and only if $p(x \mid \Theta^{(1)}) = q(x \mid \Theta^{(2)})$.

A. MONTE-CARLO APPROXIMATION

One can not calculate the integral in formula (1) because of the shape parameter which is a power in the expression. Stochastic integration with Monte-Carlo sampling is a judicious solution to overcome this issue. It consists of generating a sufficiently large number of samples $\{x_1, x_2, \ldots, x_N\}$ from the distribution p in order to approximate the integral $\int p(x \mid \theta^{(1)}) q(x \mid \theta^{(2)}) dx$ by $E_p[q(x_i|\theta^{(2)})]$. By applying this approximation to the CSD, the integral in the first term in (24) can be approximated by:

$$\int p(x \mid \theta^{(1)}) q(x \mid \theta^{(2)}) dx \approx \frac{1}{N} \sum_{i=1}^{N} q(x_i|\theta^{(2)}) \quad (25)$$

where $x_i, (i = 1, \ldots, N)$ is a set of random samples generated from the mixture $p(x \mid \theta^{(1)})$.

By substituting (25) into (24), our approximated CSD is written as follows:

$$\mathrm{CSD}_{mc} = -\log \frac{\sum_{i=1}^{N} q(x_i|\Theta^{(2)})}{\sqrt{\sum_{i=1}^{N} p(x_i|\Theta^{(1)}) \sum_{i=1}^{N} q(y_i|\Theta^{(2)})}} \quad (26)$$

B. A CLOSED-FORM EXPRESSION FOR CAUCHY-SCHWARZ DIVERGENCE BETWEEN TWO MOWBL DISTRIBUTIONS

In [45], an analytic expression of CSD for mixtures of Gaussians has been derived. Since the Gaussian distribution is a particular case of the exponential family [49], a more general formula for the closed-form expression for this family of distributions has been proposed in [19] and various examples of some popular distributions were given. Hence, our proposed closed-form expression of the CSD between Mowbl distributions is derived from the general case.

An exponential family is a set of parametric probability distributions $\{p(x \mid \theta) \mid \theta \in \Theta\}$ whose probability density can be decomposed canonically as:

$$p(x \mid \theta) = e^{\langle T(x), \eta \rangle - A(\eta) + h(x)} \quad (27)$$

where $T(x)$ denotes the sufficient statistics, η the natural parameter, $A(\eta)$ the log-normalizer, and $h(x)$ the auxiliary carrier measure. $\langle x, y \rangle = x^T y$ denotes the inner product of vectors.

The natural sufficient statistic must be a function of a single variable, which is our random variable x, but the Weibull distribution, as presented in (1), is a challenging case where x is powered by β: x^β. Nevertheless, the Weibull distribution might also be written in the canonical form. To this end, we introduce a Weibull distribution with a known shape parameter β. That is, the parameter β is not anymore a member of the source parametrization and is instead a parameter of the distribution. Hence, the Weibull distribution belongs to the exponential family by adopting the following parametrization:

$$p(x \mid \theta) = p(x \mid \alpha) = e^{-\frac{x^\beta}{\alpha^\beta} - \beta \log(\alpha) + \log(\beta) + \log x^{\beta-1}} \quad (28)$$

which is decomposed into the following parts:

- *The natural sufficient statistic:* $T(x) = \frac{x^\beta}{\beta}$
- *The natural parameter:* $\eta = -\frac{\beta}{\alpha^\beta}$
- *The log-normalizer:* $A(\eta) = \log(-\frac{1}{\eta})$
- *The auxiliary carrier measure:* $h(x) = \log(x^{\beta-1})$

When written in the canonical form of the exponential family, the CSD between MoWbl distributions can be computed in a closed-form expression by using the following equation for each integral in (24), [19]:

$$\begin{aligned}\int p(x \mid \Theta_1) q(x \mid \Theta_2) dx &= \sum_{i=1}^{M} \sum_{j=1}^{K} \pi_i \pi'_j e^{(A(\eta_i + \eta'_j) - (A(\eta_i) + A(\eta'_j)))} \\ &= \sum_{i=1}^{M} \sum_{j=1}^{K} \pi_i \pi'_j \frac{\beta_i \beta'_j}{\beta_i \alpha_j'^{\beta'_j} + \beta'_j \alpha_i'^{\beta'_i}} \end{aligned} \quad (29)$$

In the same vein, the two remaining integrals are computed using the natural parameters:

$$\begin{aligned}\int p(x \mid \Theta_1)^2 dx &= \sum_{i=1}^{M} \sum_{j=1}^{M} \pi_i \pi_j e^{(A(\eta_i + \eta_j) - (A(\eta_i) + A(\eta_j)))} \\ &= \sum_{i=1}^{M} \sum_{j=1}^{M} \pi_i \pi_j \frac{\beta_i \beta_j}{\beta_i \alpha_j^{\beta_j} + \beta_j \alpha_i^{\beta_i}} \end{aligned} \quad (30)$$

$$\begin{aligned}\int q(x \mid \Theta_2)^2 dx &= \sum_{i=1}^{K} \sum_{j=1}^{K} \pi'_i \pi'_j e^{(A(\eta_i + \eta'_j) - (A(\eta_i) + A(\eta'_j)))} \\ &= \sum_{i=1}^{K} \sum_{j=1}^{K} \pi'_i \pi'_j \frac{\beta'_i \beta'_j}{\beta'_i \alpha_j'^{\beta'_j} + \beta'_j \alpha_i'^{\beta'_i}} \end{aligned} \quad (31)$$

Finally, the closed-form expression of the CSD is given as follows:

$$\begin{aligned}\mathrm{CSD}(p||q) = &-\log\left(\sum_{i=1}^{M} \sum_{j=1}^{K} \pi_i \pi'_j \frac{\beta_i \beta'_j}{\beta_i \alpha_j'^{\beta'_j} + \beta'_j \alpha_i'^{\beta'_i}}\right) \\ &+ \frac{1}{2} \log\left(\sum_{i=1}^{M} \sum_{j=1}^{M} \pi_i \pi_j \frac{\beta_i \beta_j}{\beta_i \alpha_j^{\beta_j} + \beta_j \alpha_i^{\beta_i}}\right) \\ &+ \frac{1}{2} \log\left(\sum_{i=1}^{K} \sum_{j=1}^{K} \pi'_i \pi'_j \frac{\beta'_i \beta'_j}{\beta'_i \alpha_j'^{\beta'_j} + \beta'_j \alpha_i'^{\beta'_i}}\right) \end{aligned} \quad (32)$$

The final closed-form expression is entirely expressed by the finite mixture parameters and the similarity between the query image I_r and the target image I_c is the sum of all the CSDs across sub-bands:

$$\mathrm{D_{cs}}(I_r, I_c) = \sum_{l=1}^{L} \sum_{j=1}^{J} (\mathrm{D_{cs}}(p_r^{(l,j)} | p_c^{(l,j)}) \quad (33)$$

where p_r and p_c represent MoWbl models of the two images I_r and I_c respectively for the sub-bands of the l-th level and orientation $j = \{1, 2, \ldots, J\}$.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL SETTING

To test the performance of the proposed model in texture retrieval and the potential of the proposed similarity measure, we achieve texture retrieval experiments using three popular datasets, Brodatz [34], Vistex [33] and ALOT [35]. Each one of these datasets represents a challenge for classification or retrieval tasks either by its size or the nature of its textures. Fig. 2 shows some textures from the three datasets. The following datasets are used:

- DS1: For this dataset, we adopt the same experimental benchmark as in [50] and [17] where 60 texture images of dimension 640×640 are taken from Brodatz [34]. By dividing each image into 16 160×160 non-overlapping sub-images, we obtain a dataset of 960 image. This is a large set of textures where some are similar to each other.

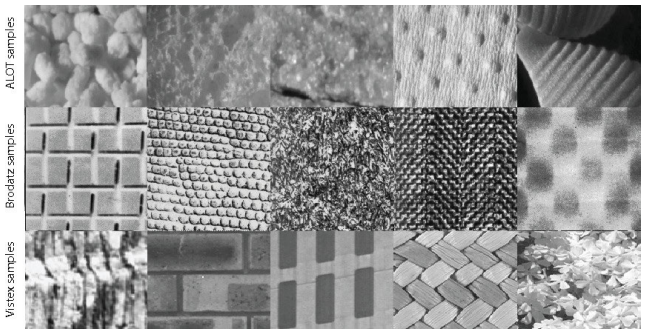


FIGURE 2. Some samples of gray-scale textures taken from the three considered datasets: (top) ALOT [35], (middle) Brodatz [34], (bottom) Vistex [33].

- DS2: This dataset is based on textured images from Vistex dataset [33]. With the same process of subdivision as in DS1, 40 reference textures have been chosen from the overall dataset. Each one of them is divided into 16 128×128 non-overlapping sub-images creating a dataset of 640 image.
- DS3: 250 texture classes from Amsterdam Library of Textures (ALOT) [35]. We selected the gray-scale version under the C1L1 capture condition. As for DS1 and DS2, each reference image has been split into 16 non-overlapping sub-images which gives us a set of 4000 images.

We apply a normalization to the luminance of database images in order to reduce the bias in our retrieval results: $I_i = ((I_i-\mu_i)/\sigma_i)\times\sigma_M+\mu_M$, where σ_i and μ_i are the standard deviation and the mean of I_i. σ_M and μ_M are the standard deviations and the medians of the means of the 16 sub-images taken from the same reference image.

In this retrieval framework, images are decomposed into various wavelet coefficient sub-bands at different decomposition scales using two wavelet transforms: The discrete wavelet transform (DWT) and the dual-tree complex wavelet transform (DTCWT). Our choice of the DTCWT has been driven by two reasons. First, it overcomes two shortcomings of the DWT: The lack of shift-invariance and the lack of directional selectivity. These properties are especially relevant for image analysis purposes. The second reason is that empirical histograms of the magnitudes of complex wavelet coefficients may be modeled by means of two-parameter Weibull distribution since the peak of the histogram is slightly shifted to the right [51]. Although the DWT does not provide complex wavelet coefficients, we use in each level the horizontal and vertical coefficient sub-bands to compute the magnitudes and the angles as in [14] with the following equations:

$$\text{mag} = \sqrt{\text{H}^2 + \text{V}^2} \tag{34}$$

$$\phi = \begin{cases} \arctan\left(\frac{V}{H}\right) & \text{if } H \neq 0 \\ \frac{\pi}{2} & \text{otherwise} \end{cases} \tag{35}$$

where H and V represent the horizontal and the vertical coefficient sub-bands respectively. *mag* is the resulting magnitude matrix. α are the angles of discrete wavelet coefficients.

For each wavelet coefficients sub-band is estimated a MoWbl. Finally, the image signature is the set of estimated parameters. That is, each image is characterized by $N_p = \sum_{l=1}^{L} 3 \times \sum_{j=1}^{J} K_j$ parameters, where L is the number of decomposition scales, J is the number of sub-bands and K_j is the number of mixture components that model the j-th sub-band.

During the evaluation process, each image in the database is considered a query image for which we retrieve the most similar images ($N_s = 16$). We measure the retrieval performance by averaging the obtained retrieval rates of all queries. This measure refers to the average retrieval rate (ARR):

$$\text{ARR} = \frac{\sum_{i=1}^{N} N_i}{N_s \times N} \tag{36}$$

where N denotes the number of images in the dataset (e.g. $N = 960$ in case of DS1), N_s is the number of retrieved images ($N_s = 16$ in all our retrieval experiments), and N_i is the number of correctly retrieved samples at the i^{th} query test.

According to the dataset construction, a sub-image is considered relevant if it is part of the same original image as the query sub-image.

The reminder of the experimental results section is organized as follows. A model validation test is presented in the next subsection. Then, a dedicated experiment is designed to highlight the performance of the proposed similarity measurement. Finally, our texture retrieval experiment has been divided in two parts. In the first part we study the impact of multi-modal distributions on modeling only the magnitudes of complex coefficients, whereas the second part concerns the influence of the information from angles.

B. MODEL VALIDATION AND GOODNESS-OF-FIT

In this model validation test, we investigate whether the magnitude of complex wavelet coefficients can be distributed according to a MoWbl distribution. As shown in Fig. 3 and Fig. 4, some normalized histograms of the magnitudes of the wavelet coefficients are multimodal. These multimodal shapes are well fitted with a MoWbl in comparison with Weibull and generalized Gamma distributions which fail at fitting more than one mode in the histogram shape. Even the mixture of Gaussian distributions (MoG) is slightly inacurate compared to the MoWbl.

A goodness-of-fit (GoF) Chi-Square (χ^2) test has been applied to different texture images from all datasets in order to justify the use of the MoWbl distribution model. Table 1 presents the percentages of rejected null-hypothesis of the Chi-Square GoF tests when using either a Weibull, GΓ or MoWbl distribution to model the empirical histogram of magnitudes of the complex wavelet coefficients (at a 5% significance level). The null-hypothesis percentages drawn from this quantitative GoF tests confirm that the MoWbl is suitable to model the magnitudes of complex wavelet

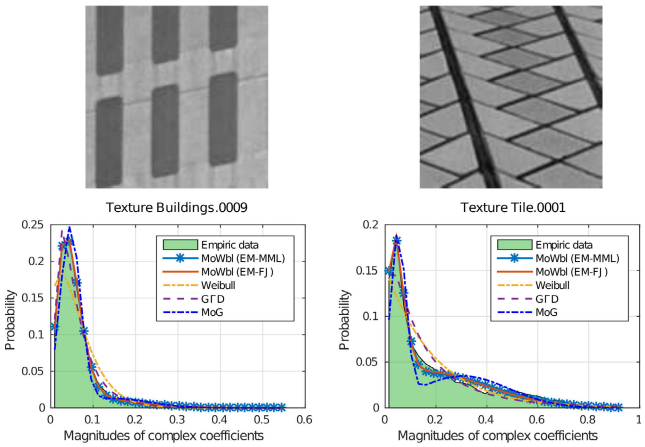


FIGURE 3. Fitting of the empirical normalized histograms of DTCWT wavelets coefficients magnitudes at the first decomposition level using Weibull, generalized Gamma (*G*Γ) mixture of Gaussians (MoG) and the proposed mixture of Weibull (MoWbl) distributions. The corresponding texture images are taken from the dataset DS2 (VisTex).

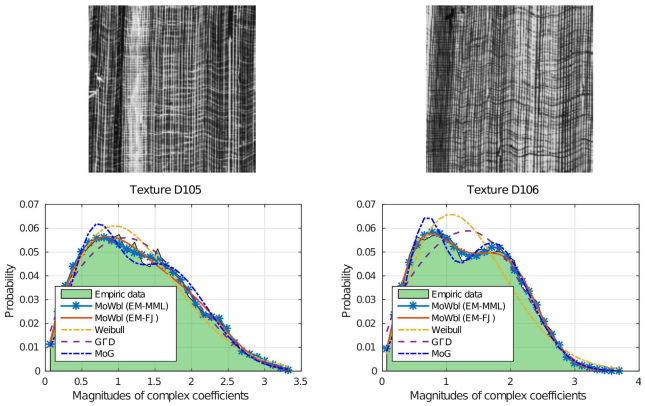


FIGURE 4. Fitting of the empirical normalized histograms of orthogonal wavelets coefficients magnitudes at the first decomposition level using Weibull, generalized Gamma (*G*Γ), mixture of Gaussians (MoG) and the proposed mixture of Weibull (MoWbl) distribution. The corresponding texture images are taken from the dataset DS1 (Brodatz).

coefficients, especially the coefficients issued from the DTCWT. These rejection rates go in line with our assumption about the existence of multimodal magnitude histograms.

The proposed unsupervised expectation-maximization based algorithm (EM-FJ-MoWbl) for finite mixtures of Weibull distributions demonstrated good fitting accuracy in a highly reduced execution time and few refinement iterations compared to the supervised version (EM-MML-MoWbl). The BIC criterion [38] has also been considered in this comparison. Both EM-BIC-MoWbl and EM-MML-MoWbl have almost the same results expect some cases where MML provides the best number of components. This performance evaluation has been conducted on some textures from Brodatz and Vistex datasets. Results of this experiment are summarized in Table 2. Quantitative measures used in this evaluation are the Kullback-Leibler divergence (KLD), the Kolmogorov-Smirnov distance (KS), the execution time (t), the number of estimated mixture components (M*) and the number of EM iterations (*iter*). The unsupervised learning algorithm

TABLE 1. Percentage of rejected null-hypotheses, at a significance level of 0.05, over the used datasets, for Weibull, GΓD, MoG and MoWbl distributions that model magnitudes of both DTCWT and DWT coefficients.

Datasets	Dist.	Rejection rate per level (%)					
		Level 1		Level 2		Level 3	
		DTCWT	DWT	DTCWT	DWT	DTCWT	DWT
DS1	Weibull	55.52	52.40	45.52	41.04	38.44	29.58
	GΓD	50.94	45.31	47.50	41.72	39.53	33.75
	MoG	69.46	70.23	57.27	59.17	38.92	31.87
	MoWbl	**41.74**	**43.39**	**36.58**	**39.28**	**36.73**	**25.53**
DS2	Weibull	27.40	29.33	29.00	25.16	18.23	9.22
	GΓD	23.15	25.38	22.75	16.51	14.22	9.97
	MoG	55.83	59.22	51.30	50.62	15.70	12.19
	MoWbl	**18.54**	**20.78**	**19.95**	**12.5**	**12.58**	**6.41**
DS3	Weibull	47.16	57.5	41.34	48.52	30.28	22.05
	GΓD	27.11	35.58	23.93	25.12	14.97	11.77
	MoG	51.26	57.12	44.10	49.86	40.15	38.50
	MoWbl	**24.57**	**33.17**	**22.60**	**22.10**	**12.55**	**10.08**

EM-FJ-MoWbl runs 96% faster than the supervised version EM-MML-MoWbl and gives approximatively the same fitting accuracy in case of we choose $M_{max} = 20$ and $M_{min} = 2$. By starting with a large number of components, we guarantee the robustness of the algorithm against initialization issues. Local maxima of the likelihood arise when there are too many components in one region of the space, and few in another. This problem is avoided by starting with a large number of components then the algorithm removes the weakest components iteratively by proceeding to components annihilation [32]. After some experiments on M_{max}, 20 is found to be the best value to start with.

C. SIMILARITY MEASUREMENT PERFORMANCE

The improvement in runtime speed is huge compared to the approximation of KLD. Besides the computational time, CSD has also improved the retrieval rates. This is demonstrated by comparing results of the proposed analytic expression of CSD and the Monte-Carlo approximation of KLD using various number of random samples. The average retrieval rates and their corresponding average runtime execution per query are illustrated in Fig. 5. The average retrieval rate using the KLD increases as the number of samples used for Monte-Carlo integration increases. Even when $1e^6$ random samples are generated for KLD, the CSD still outperforms the KLD approximation. Based on our retrieval experiment on DS1 textures, the CSD based approach reaches 90.89% in ARR with an average runtime of $0.42sec$ per query, while it takes $12sec$ in average per query for the KLD based approach to reach only 89.80% in ARR. Monte-Carlo approximation of CSD takes longer than KLD especially when the number of the set of random samples is large. This is due to the second integral of CSD which requires another set of random samples to be generated. Experiments were performed in MATLAB on a workstation with 2.6 *GHz* processor and 4 *Gb* of RAM.

D. TEXTURE RETRIEVAL PERFORMANCE

In this experiment, we execute our method with two types of wavelet transforms, the discrete wavelet transform (DWT)

TABLE 2. Quantitative assessment of the parameters estimation when using the supervised or the unsupervised learning algorithm for the mixture of Weibull distributions (MoWbl). Complex coefficient magnitudes of the first level of decomposition using the DTCWT are considered.

Texture	EM-BIC-MoWbl					EM-MML-MoWbl					EM-FJ-MoWbl				
	KLD	KS	t	iter	M^*	KLD	KS	t	iter	M^*	KLD	KS	t	iter	M^*
D096	0.0025	0.7213	288.4s	11507	3	**0.0022**	0.7240	286s	11504	3	0.0052	**0.9476**	**11.84s**	**691**	2
D105	0.0039	0.0091	291.5s	16432	3	0.0040	0.0084	290.15s	16430	3	**0.0038**	**0.8238**	**7.19s**	**537**	**2**
D106	0.0051	0.5113	217.2s	12365	2	0.0043	**0.5060**	219.11s	12367	2	**0.0025**	0.3879	**6.57s**	**409**	2
D103	0.0082	0.220	167s	9080	2	0.0088	0.216	166.2s	9077	2	0.0088	**0.230**	**12.88s**	**945**	2
Metal.0002	0.0045	0.60	280.15s	16155	2	0.0042	**0.6030**	278.70s	16151	2	**0.0029**	0.5083	**6.83s**	**345**	2
Tile.0001	0.0050	0.5510	196s	12715	3	**0.0050**	0.5300	195.14s	12712	3	0.0055	**0.8808**	**12s**	**1178**	3
Terrain.0010	0.0052	0.3970	98s	16053	3	**0.0057**	0.3964	99.50s	16062	2	0.0051	**0.7689**	**3.37s**	**474**	2
Wood.0002	0.0041	0.7240	61.2s	10073	3	0.0039	0.6216	59.40s	10062	2	**0.0040**	**0.8035**	**3.36s**	**676**	2

TABLE 3. Average retrieval rates (%) in the top 16 images retrieved with different methods compared with our proposed method on all datasets by modeling only magnitudes of coefficients from DWT and DTCWT.

Dataset	Transform	GGD +KLD [9]	Weibull +KLD [24]	GΓD +KLD [50]	MoG +CSD [45]	MoGG +KLD [12]	MoGG +CSD [17]	MoWbl +CSD (proposed)
DS1	DWT	59.30	64.93	67.82	60.47	65.83	67.16	71.11
	DTCWT	81.23	83.85	84.51	85.51	86.90	87.77	**90.89**
DS2	DWT	49.90	79.77	83.35	53.17	58.73	59.41	88.12
	DTCWT	76.57	81.24	84.50	84.16	87.55	88.36	**91.10**
DS3	DWT	23.77	30.36	34.52	26.47	29.50	30.80	36.47
	DTCWT	31.81	37.97	43.48	42.67	43.65	44.12	**46.17**

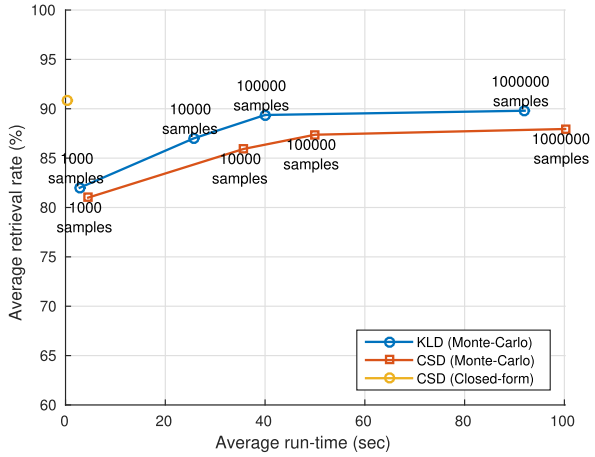
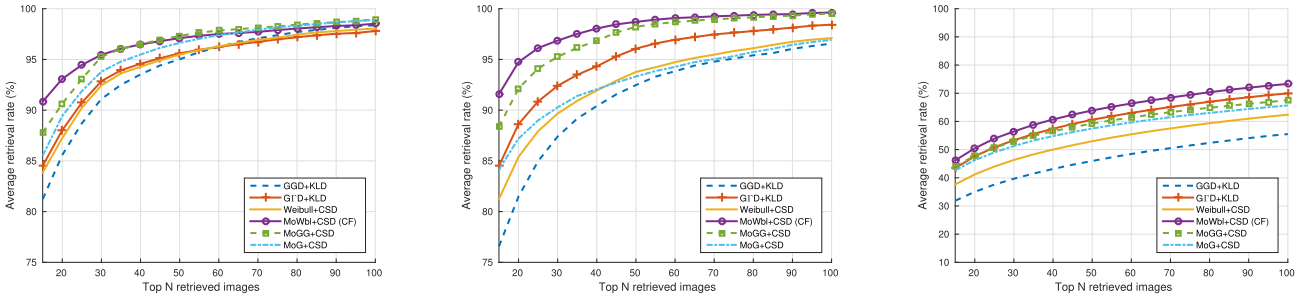


FIGURE 5. The average retrieval rate *vs* the averaged run-time per query (in seconds) using the proposed MoWbl model. Even though KLD reaches $1e^6$ random samples, its curve is below that of CSD, which indicates that the proposed closed-form expression of CSD outperforms the approximation of KLD.

with daubchies 4 filter and the dual-tree complex wavelet transform (DTCWT).

We compared our method to some state-of-the-art statistical models in texture retrieval, namely: i) GGD+KLD: The generalized Gaussian distribution (GGD) with KLD proposed initially by Do and Vetterli [9] to characterize histograms of real valued coefficients of DWT, ii) GΓD+KLD: The generalized Gamma distribution (GΓD) with KLD proposed by Choy and Tong [50] for characterizing histograms of positive DWT coefficients and also used by de Ves *et al.* [23], [25] to characterize histograms of magnitudes of complex wavelet coefficients. iii) Weibull+KLD: The Weibull distribution with KLD which has also been successfully used to characterize histograms of magnitudes of the complex wavelet coefficients by Kwitt and Uhl [24]. Additionally, in the context of finite mixture models, two mixture distributions have been used for comparison too: The mixture of Gaussian distributions (MoG+CSD) with an analytic expression of CSD proposed by Kampa *et al.* [45], the mixture of generalized Gaussian distributions with KLD (MoGG+KLD) by Allili [12] and the mixture of generalized Gaussian distribution with CSD (MoGG+CSD) that we previously proposed in [17]. We recall that all of these methods have been used in the current experiment to characterize histograms of magnitudes of complex wavelet coefficients from both DTCWT and DWT.

Table 3 shows the obtained results in the three datasets DS1, DS2 and DS3 respectively using three decomposition levels of different wavelet transforms. The obtained ARRs, confirm that Weibull and GΓD are very suitable for magnitude modeling in case of complex wavelet coefficients. Results obtained with the MoWbl distribution confirm our assumption about the existence of multi-modal histograms. There is a significant improvement from a single Weibull distribution to a mixture of Weibull distribution, which is around 7% in DS1 and 10% in DS2. The same improvement is noticed in textured images of ALOT database where the use of a mixture distribution gives better retrieval results. Compared to MoGG, the proposed MoWbl model performs better by increasing the ARR by 3.12%, 2.74% and 2.05% on the the three datasets DS1, DS2 and DS3 respectively. This is mainly due to the support of the underline probability density function which handles positive and negative values whereas magnitudes of complex coefficients are only positive. Hence, the MoWbl is highly compatible with the modeled information. Another considerable reason is the



(a) Performance on DS1 (b) Performance on DS2 (c) Performance on DS3

FIGURE 6. Average retrieval rate as a function of the number of the top 16 matches using different statistical models for magnitudes compared to the proposed MoWbl+CSD model on (a) DS1, (b) DS2 and (c) DS3. (a) Performance on DS1. (b) Performance on DS2. (c) Performance on DS3.

fixed shape parameter imposed in the MoGG in order to obtain an analytic expression of CSD. The obtained low ARRs in DS3 are due to the high number of textures of the dataset (4000 textures), which makes difficult to discriminate between similar texture classes.

Fig. 6 illustrates the average retrieval rates (ARR) in function of the query size for different model configurations. These curves provide an additional justification of the improvement in texture retrieval using a mixture of Weibull distribution (MoWbl) for magnitudes.

E. INFLUENCE OF PHASE INFORMATION

Although our main contribution concerns magnitudes of complex wavelet coefficients, another experiment on influence of the information from angles has been conducted on the three datasets. To capture this information, we use von Mises distribution to characterize histograms of the relative phase as in [22], [23], and [25]. This joint modeling requires some changes in the similarity measurement. Consequently, the new similarity measurement is a convex combination of the CSD between the MoWbl distributions and KLD between von Mises distributions as defined by:

$$D(I_r, I_c) = \alpha \sum_{l=1}^{L} \sum_{j=1}^{J} D_{cs}(p_r^{(l,j)} \mid p_c^{(l,j)}) + (1-\alpha) \sum_{l=1}^{L} \sum_{j=1}^{J} D_{KL}(q_r^{(l,j)} \mid q_c^{(l,j)}) \quad (37)$$

where p_r and p_c represent MoWbl models of the two images I_r and I_c respectively for the sub-bands of the l^{th} level and subband $j = \{1, 2, \ldots, J\}$. Statistical models of the relative phase part are represented by q_r and q_c, which are von Mises distributions. The weight α serves as a balance between the contribution of magnitudes and relative phase models. For $\alpha = 0$, we use only information from magnitudes, whereas when $\alpha = 1$ we use only information from angles. The optimal value of α has been found to be around 0.5 by running retrieval experiments on the three datasets by varying α from 0 to 1.

TABLE 4. Impact of the relative phase information modeling on average retrieval rates (%) in the top 16 images retrieved with various methods using 3 scales of DTCWT.

Method	Features	Average retrieval rate (%)		
		DS1	DS2	DS3
GGD_{re}+KLD [9]	36	83.12	79.71	36.07
GGD_{re}-VM+KLD [29]	72	83.45	81.11	41.00
GGD_{im}+KLD [9]	36	83.34	80.00	39.57
GGD_{im}-WC [29]	72	83.45	81.40	41.13
Weibull+KLD	36	83.85	81.24	39.67
Weibull-VM+KLD	72	83.90	82.58	41.48
GΓD+KLD	36	84.51	84.50	43.48
GΓD-VM+KLD [23]	72	85.20	85.28	44.73
MoWbl+CSD (proposed)	54	90.89	91.10	46.17
MoWbl-VM+CSD (proposed)	90	**92.15**	**93.35**	**47.60**

Methods that have been proposed specifically for complex wavelet coefficients modeling are: i) GGD_{re}-VM and GGD_{im}-WC by Vo and Oraintara [29] which models real and imaginary parts of a given complex wavelet coefficients respectively. ii) GΓD-VM by de Ves *et al.* [23] which uses a GΓD for magnitudes and a von Mises distribution for relative phase.

In Table 4, we show the average retrieval rates obtained when combining statistical models for both magnitudes and relative phase and the feature size of each method. As expected, involving information from angles increases the ARRs. For instance, the MoWbl model benefits from an improvement of 1.26%, 2.25% and 1.43% when we add von Mises models in DS1, DS2 and DS3 respectively, but at the expense of additional 36 features.

V. CONCLUSION AND FUTURE WORK

In this paper, a statistical model for texture retrieval is proposed to characterize textured images by using the information from magnitudes of complex wavelet coefficients. Each sub-band is modeled with a mixture of Weibull distributions. The advantage of using a finite mixture model compared to a single distribution model is that multi-modal behavior found in many empirical histograms is well captured with a finite mixture. To estimate the appropriate number of mixture components, an unsupervised learning algorithm is used for the MoWbl model based on the minimum message length criterion. Thanks to the canonical form of the

exponential family, we came through a closed-form expression for the CSD between two MoWbl distributions. Information from the phase of complex wavelet coefficients has also been exploited by characterizing the relative phase with von Mises distributions. This joint combination of MoWbl and von Mises for has led to an improvement of the average retrieval rates. In a future work, we will further explore properties of the exponential family distributions to promote the Cauchy-Schwarz divergence to develop closed-form expressions for divergences between finite mixture models for magnitudes and relative phase too.

REFERENCES

[1] W.-Y. Ma and B. S. Manjunath, "NeTra: A toolbox for navigating large image databases," *Multimedia Syst.*, vol. 7, no. 3, pp. 184–198, May 1999.
[2] W. Y. Ma and B. S. Manjunath, "Texture features and learning similarity," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1996, pp. 425–430.
[3] 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.
[4] M. Haindl and P. Vacha, "Illumination invariant texture retrieval," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3, Aug. 2006, pp. 276–279.
[5] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, no. 10, pp. 847–856, Jan. 1980.
[6] T. Randen and J. H. Husøy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 291–310, Apr. 1999.
[7] S. G. 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.
[8] G. V. D. Wouwer, P. Scheunders, and D. V. Dyck, "Statistical texture characterization from discrete wavelet representations," *IEEE Trans. Image Process.*, vol. 8, no. 4, pp. 592–598, Apr. 1999.
[9] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.
[10] 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. (ICASSP)*, vol. 3, May 2004, pp. 597–600.
[11] G. McLachlan and D. Peel, *Finite Mixture Models*. Hoboken, NJ, USA: Wiley, 2004.
[12] M. S. Allili, "Wavelet modeling using finite mixtures of generalized Gaussian distributions: Application to texture discrimination and retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1452–1464, Apr. 2012.
[13] M. S. Allili, N. Baaziz, and M. Mejri, "Texture modeling using contourlets and finite mixtures of generalized Gaussian distributions and applications," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 772–784, Apr. 2014.
[14] H.-C. Li, V. A. Krylov, P.-Z. Fan, J. Zerubia, and W. J. Emery, "Unsupervised learning of generalized gamma mixture model with application in statistical modeling of high-resolution Sar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2153–2170, Apr. 2016.
[15] W. Yang, X. Yang, T. Yan, H. Song, and G. S. Xia, "Region-based change detection for polarimetric SAR images using Wishart mixture models," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6746–6756, Nov. 2016.
[16] J. C. Seabra, F. Ciompi, O. Pujol, J. Mauri, P. Radeva, and J. Sanches, "Rayleigh mixture model for plaque characterization in intravascular ultrasound," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1314–1324, May 2011.
[17] H. Rami, L. Belmerhnia, A. D. E. Maliani, and M. E. Hassouni, "Texture retrieval using mixtures of generalized Gaussian distribution and Cauchy–Schwarz divergence in wavelet domain," *Signal Process., Image Commun.*, vol. 42, pp. 45–58, Mar. 2016.
[18] R. Jenssen, J. C. Principe, D. Erdogmus, and T. Eltoft, "The Cauchy–Schwarz divergence and Parzen windowing: Connections to graph theory and Mercer kernels," *J. Franklin Inst.*, vol. 343, no. 6, pp. 614–629, Sep. 2006.
[19] F. Nielsen, "Closed-form information-theoretic divergences for statistical mixtures," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 1723–1726.
[20] N. Kingsbury, "The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters," in *Proc. IEEE Digit. Signal Process. Workshop*, vol. 86, Aug. 1998, pp. 120–131.
[21] 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. 49–70, Oct. 2000.
[22] Y. Rakvongthai and S. Oraintara, "Statistical texture retrieval in noise using complex wavelets," *Signal Process., Image Commun.*, vol. 28, no. 10, pp. 1494–1505, Nov. 2013.
[23] E. De Ves, D. Acevedo, A. Ruedin, and X. Benavent, "A statistical model for magnitudes and angles of wavelet frame coefficients and its application to texture retrieval," *Pattern Recognit.*, vol. 47, no. 9, pp. 2925–2939, Sep. 2014.
[24] R. Kwitt and A. Uhl, "Lightweight probabilistic texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 241–253, Jan. 2010.
[25] E. de Ves, X. Benavent, A. Ruedin, D. Acevedo, and L. Seijas, "Wavelet-based texture retrieval modeling the magnitudes of wavelet detail coefficients with a generalized gamma distribution," in *Proc. 20th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2010, pp. 221–224.
[26] A. D. E. Maliani, M. E. Hassouni, Y. Berthoumieu, and D. Aboutajdine, "Color texture classification method based on a statistical multi-model and geodesic distance," *J. Vis. Commun. Image Represent.*, vol. 25, no. 7, pp. 1717–1725, Oct. 2014.
[27] N.-E. Lasmar and Y. Berthoumieu, "Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2246–2261, May 2014.
[28] A. Jalobeanu, L. Blanc-Feraud, and J. Zerubia, "Natural image modeling using complex wavelets," *Proc. SPIE*, vol. 5207, pp. 480–495, Nov. 2003.
[29] A. Vo and S. Oraintara, "A study of relative phase in complex wavelet domain: Property, statistics and applications in texture image retrieval and segmentation," *Signal Process., Image Commun.*, vol. 25, no. 1, pp. 28–46, Jan. 2010.
[30] H. Oulhaj, M. Rziza, A. Amine, H. Toumi, E. Lespessailles, M. E. Hassouni, and R. Jennane, "Anisotropic discrete dual-tree wavelet transform for improved classification of trabecular bone," *IEEE Trans. Med. Imag.*, vol. 36, no. 10, pp. 2077–2086, Oct. 2017.
[31] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.
[32] M. A. T. 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.
[33] *Mit Vision and Modelling Group. Texture Database*. Accessed: Apr. 2019. [Online]. Available: http://vismod.media.mit.edu
[34] *Brodatz Database*. Accessed: Apr. 2019. [Online]. Available: http://www.ux.uis.no/~tranden/brodatz.html
[35] *Amesterdam Library of Textures*. Accessed: Apr. 2019. [Online]. Available: http://staff.science.uva.nl/~aloi/public_alot
[36] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
[37] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.
[38] C. S. Wallace and D. M. Boulton, "An information measure for classification," *Comput. J.*, vol. 11, no. 2, pp. 185–194, 1968.
[39] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
[40] C. Biernacki, G. Celeux, and G. Govaert, "Assessing a mixture model for clustering with the integrated completed likelihood," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 7, pp. 719–725, Jul. 2000.
[41] R. A. Baxter and J. J. Oliver, "Finding overlapping components with MML," *Statist. Comput.*, vol. 10, no. 1, pp. 5–16, Jan. 2000.
[42] H. Rinne, *The Weibull Distribution: A Handbook*. Boca Raton, FL, USA: CRC Press, 2008.
[43] J. M. Bernard and A. F. Smith, *Bayesian Theory*. Hoboken, NJ, USA: Wiley, 1994.
[44] A. R. Barron, C. Huang, J. Q. Li, and X. Luo, "MDL, penalized likelihood, and statistical risk," in *Proc. IEEE Inf. Theory Workshop. (ITW)*, May 2008, pp. 247–257.
[45] K. Kampa, E. Hasanbelliu, and J. C. Principe, "Closed-form Cauchy–Schwarz PDF divergence for mixture of Gaussians," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul./Aug. 2011, pp. 2578–2585.

[46] 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. (ICASSP)*, vol. 4, Apr. 2007, pp. IV-317–IV-320.
[47] J. Goldberger, S. Gordon, and H. Greenspan, "An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 1, Oct. 2003, pp. 487–493.
[48] F. Nielsen and K. Sun, "Guaranteed bounds on the Kullback–Leibler divergence of univariate mixtures," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1543–1546, Nov. 2016.
[49] L. D. Brown, *Fundamentals of Statistical Exponential Families With Applications in Statistical Decision Theory* (Lecture Notes-Monograph Series), vol. 9. Beachwood, OH, USA: Institute of Mathematical Statistics, 1986, P. 279.
[50] 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.
[51] R. Kwitt and A. Uhl, "Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval," in *Proc. 15th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2008, pp. 933–936.

HASSAN RAMI received the master's degree from Mohammed V University, Rabat, Morocco, in 2013, where he is currently pursuing the Ph.D. degree in computer science. His current research interests include machine learning, image analysis, and computer vision, especially on statistical models for texture retrieval.

AHMED DRISSI EL MALIANI received the Ph.D. degree in engineering sciences from Mohammed V University, Rabat, Morocco, with a focus on the stochastic modeling of natural images with applications in content-based texture retrieval and classification, where he was a member of the LRIT Laboratory. He is currently an Associate Professor with the Faculty of Sciences, Mohammed V University. His current research interests include machine learning and medical image analysis. He was a member of numerous national and international conference committees. He also serves as a Reviewer for well-known journals in the field.

MOHAMMED EL HASSOUNI received the Habilitation degree from University Mohammed V-Agdal, in 2012, and the Ph.D. degree in image and video processing from the University of Burgundy, in 2005. Since 2006, he has been with Mohammed V University as an Assistant Professor, where he is currently a Full Professor, since 2018. He was a Visitor for several universities (Bordeaux I, Orleans, Dijon, and Konstanz). His research interests include image analysis, quality assessment, and mesh processing, especially on medical imaging, biometry, and QoE. He was a member of the IEEE Signal Processing Society.