Rotation-invariant texture retrieval with Gaussianized steerable pyramids

George Tzagkarakis, Baltasar Beferull-Lozano, *Associate Member, IEEE*, and Panagiotis Tsakalides, *Member, IEEE*

***Abstract*— This paper presents a novel rotation-invariant image retrieval scheme based on a transformation of the texture information via a steerable pyramid. First, we fit the distribution of the subband coefficients using a joint alpha-stable sub-Gaussian model to capture their non-Gaussian behavior. Then, we apply a normalization process in order to Gaussianize the coefficients. As a result, the feature extraction step consists of estimating the covariances between the normalized pyramid coefficients. The similarity between two distinct texture images is measured by minimizing a rotation-invariant version of the Kullback-Leibler Divergence between their corresponding multivariate Gaussian distributions, where the minimization is performed over a set of rotation angles.**

***Index Terms*— Statistical image retrieval, rotation-invariant Kullback-Leibler Divergence, steerable model, Fractional Lower-Order Moments, sub-Gaussian distribution.**

I. Introduction

DURING the last decades, information is being gathered and stored at an impressive rate on large digital databases. Examples include multimedia databases containing audio, images and video. The search of large digital multimedia libraries, unlike the search of conventional text-based digital databases, cannot be realized by simply searching text annotations. Because of the amount of details in multimedia data, it is difficult to provide automatic annotation without human support. The design of completely automatic mechanisms that extract meaning from this data and characterize the information content in a compact and meaningful way is a challenging task.

Content-based Image Retrieval (CBIR) is a set of techniques for retrieving relevant images from a database on the basis of automatically-derived features, which accurately specify the information content of each image. We can distinguish two major tasks, namely Feature Extraction (FE) and Similarity Measurement (SM). In the FE step, a set of features constituting the so-called image signature is generated after a preprocessing step (image transformation), to accurately represent the content of a given image. This set has to be much smaller in size than the original image, while capturing as much as possible of the image information. During the SM step, a distance function is employed which measures how close each image in the database is to a query image, by comparing their signatures.

Typical low-level image features, such as color [1], shape [2] and texture [3], are commonly used in CBIR applications. In this work, we focus on the use of *texture information* for image retrieval. Loosely speaking, the class of images that we commonly call *texture images* includes images that are spatially homogeneous and consist of repeated elements, often subject to some randomization in their location, size, color and orientation. Previously developed texture extraction methods include multi-orientation filter-banks and spatial Gabor filters [4]. The basic assumption for these approaches is that the energy distribution in the frequency domain identifies a texture. These retrieval systems use simple norm-based distances (e.g. Euclidean distance) on the extracted image signatures, as a similarity measure.

In this work, we consider the tasks of FE and SM in a joint statistical framework. Thus, in our approach, the FE step becomes a Maximum Likelihood (ML) estimator of the model parameters fitting the given image data, while the SM step employs a statistical measure of similarity, such as the Kullback-Leibler Divergence (KLD) [5], between probability density functions having different model parameters. In this setting, optimal retrieval is asymptotically achieved. Using this statistical approach, a simple extension of the energy-based methods for texture retrieval is to model each texture by the marginal densities of the transform coefficients. This is motivated by the results of resent physiological research on human texture perception, which suggest that two homogeneous textures are often difficult to discriminate if they produce similar marginal distributions of responses from a filter-bank [6].

The development of retrieval models in a transform-domain is based on the observation that often a linear, invertible transform restructures the image, resulting in a set of transform coefficients whose structure is simpler to model. Real-world images are characterized by a set of "features", such as textures, edges, ridges and lines. For such images, the 2-dimensional (2-D) wavelet transform has been shown to be a powerful modeling tool, providing a natural arrangement of the wavelet coefficients into multiscale and oriented subbands representing the horizontal, vertical and diagonal edges [7]. Texture information is modeled using the first or second order statistics of the coefficients obtained via a Gabor wavelet transform [8], or an overcomplete wavelet decomposition constituting a tight frame [9].

On the other hand, considering a statistical framework, texture is modeled by joint probability densities of wavelet

This work was supported by the Greek General Secretariat for Research and Technology under Program ΕΠΑΝ, Code ΗΠΑ-011.

G. Tzagkarakis and P. Tsakalides are with the Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH-ICS) and the Department of Computer Science, University of Crete, Heraklion, Greece. Emails: {gtzag, tsakalid}@ics.forth.gr

B. Beferull-Lozano is with the School of Computer and Comm. Sciences, Audiovisual Communications Laboratory (EPFL), Lausanne, Switzerland. Email: Baltasar.Beferull@epfl.ch

subband coefficients. Until recently, wavelet coefficients have been modeled either as independent Gaussian variables or as jointly Gaussian vectors [10]. However, it has been pointed out that the wavelet transforms of real-world images tend to be sparse, resulting in a large number of small coefficients and a small number of large coefficients [11]. This property is in conflict with the Gaussian assumption, giving rise to peaky and heavy-tailed *non-Gaussian* marginal distributions of the wavelet subband coefficients [11], [12].

Experimental results have proven that the *generalized Gaussian density* (GGD) is a suitable member of the class of non-Gaussian distributions for modeling the marginal behavior of the wavelet coefficients [11], [13]. Computationally tractable image retrieval mechanisms based on a combination of overcomplete wavelet-based texture and color features are described in [14], where the similarity measure between the GGD models is based on the Bhattacharrya distance. Recently, the GGD models have been also introduced in a statistical framework for texture retrieval in CBIR applications, by jointly considering the two problems of FE and SM [15].

In recent work, we showed that successful image processing algorithms can achieve both superior noise reduction and feature preservation if they take into consideration the actual heavy-tailed behavior of the signal and noise densities [16], [17]. We demonstrated that successful modeling of subband decompositions of many texture images is achieved by means of *symmetric alpha-stable* ($S\alpha S$) distributions [18], [19], which very often provide a better fit of the non-Gaussian heavy-tailed distributions, than the generalized Gaussian distribution (GGD), thus motivating their use in our CBIR model. After extracting the $S\alpha S$ model parameters, we analytically derived the KLD between two $S\alpha S$ distributions. Our formulation improved the retrieval performance, resulting in a decreased probability error rate for images with distinct non-Gaussian statistics [20], compared with the GGD model.

However, the majority of current approaches does not take into account the important interdependencies between different subbands of a given image, which can be employed in order to provide a more accurate representation of the texture image profile. Huang studied the correlation properties of wavelet transform coefficients at different subbands and resolution levels, applying these properties on an image coding scheme based on neural networks [21]. Portilla and Simoncelli developed an algorithm for synthesizing texture images by setting different constraints on the correlation between the transform coefficients and their magnitudes [22].

The theory of Markov random fields has enabled a new generation of statistical texture models, in which the full model is characterized by statistical interactions within local neighborhoods [23]. Recently, a new framework for statistical signal processing based on wavelet-domain hidden Markov models has been proposed [24], [25]. It provides an attractive modeling of both the non-Gaussian statistics and the property of persistence across scales in a wavelet decomposition.

In this paper, we proceed by grouping the wavelet subband coefficients and considering them as samples of a multivariate *sub-Gaussian* random process, which is characterized by the associated *fractional lower-order statistics*. Within the framework of sub-Gaussian processes, we use the notion of *covariation* instead of the second-order covariance, in order to extract possible interdependencies between wavelet coefficients at different image orientations and scales. The joint sub-Gaussian modeling preserves the heavy-tailed behavior of the marginal distributions, as well as the strong statistical dependence across orientations and scales.

A desirable property in a CBIR system is rotation invariance. This is a topic that has been previously pursued by various researchers. Greenspan $et.\ al.$ [26] and Haley and Manjunath [27], [28] employed rotation-invariant structural features, using autocorrelation and DFT magnitudes, obtained via multiresolution Gabor filtering. Recently, a rotation-invariant image retrieval system based on steerable pyramids was proposed by Beferull-Lozano *et al.* [29]. In this system, the correlation matrices between several basic orientation subbands at each level of a wavelet pyramid are chosen as the energy-based texture features. Mao and Jain [30] presented a multiresolution simultaneous autoregressive (MR-SAR) model where a multivariate rotation-invariant SAR (RISAR) model is introduced, which is based on the circular autoregressive (CAR) model.

A second category of methods achieving rotation invariance includes the implementation of a Hidden Markov Model (HMM) on the subband coefficients of the transformed image. Do and Vetterli [25] derived a steerable rotation-invariant statistical model by enhancing a recently introduced technique on wavelet-domain HMM [24]. Liu and Picard [31] exploited the effectiveness of the 2-D Wold decomposition of homogeneous random fields, in order to extract features that represent perceptual properties described as "periodicity", "directionality" and "randomness".

The above mentioned rotation-invariant CBIR techniques can be classified in two classes. The first class includes techniques where the FE step consists of computing rotation-invariant texture features, while the SM step consists of applying a common similarity function, such as the Euclidean distance and the KLD. The second class includes techniques where the FE step consists of estimating the parameters of a so-called *steerable model* and then applying a rotation-invariant version of a common similarity function (e.g. KLD), during the SM step.

In this paper, we describe a novel technique belonging to the second class. First, we design a new steerable model, which is based on the joint sub-Gaussian modeling of the coefficients of a *steerable pyramid* incorporating dependence across orientations and scales. Then, we apply a Gaussianization procedure on the steerable pyramid coefficients, by jointly considering them as samples of a multivariate sub-Gaussian distribution, viewed as a special case of a Gaussian Scale Mixture (GSM). After the Gaussianization step, we derive an analytical expression for a rotation-invariant version of the KLD between multivariate Gaussian densities (including the rotation angle between textures), avoiding the use of a Monte-Carlo method, usually employed to approximate the KLD in the non-Gaussian case [25].

Our system has several advantages with respect to the HMM-based methods. First, HMMs require the use of an

Expectation-Maximization (EM) algorithm, which in some cases may not converge, for the estimation of the model parameters (hidden state variables and statistics of a Gaussian mixture). On the other hand, our proposed method incorporates dependence across space, orientations and scales, combined in an efficient way of estimating the multipliers of the multivariate sub-Gaussian model, which are necessary to perform the Gaussianization. Besides, by exploiting the statistical dependencies between subbands at adjacent scales, we insert the same first-order Markovian dependence as in HMMs, but in a simpler way. Also, for the heavy-tailed modeling we use $S\alpha S$ distributions, which are often better than GGDs.

The rest of the paper is organized as follows: in section II, we briefly review the probabilistic setting for a CBIR problem. In section III, we justify the choice of the multivariate sub-Gaussian model for the joint modeling of the wavelet coefficients. In section IV, we develop a rotation-invariant CBIR system by applying a Gaussianization procedure on the coefficients of a steerable pyramid. In section V, we apply our scheme to a set of textures and evaluate the retrieval performance. Finally, in section VI, we provide conclusions and directions for future research.

II. Statistical CBIR

Let $\mathcal{F}$ denote the feature space and $\vec{X} = \{\vec{x}_1, \ldots, \vec{x}_N \,|\, \vec{x}_i \in \mathcal{F}\ ,\ i = 1, \ldots, N\}$ be a set of N independent feature vectors associated to a query. Also, let $\mathcal{S} = \{1, \ldots, K\}$ be the set of class indicators associated with the image classes in the database. Denote the probability density function (PDF) of the query feature vector space by $p_q(\vec{x})$ and the PDF of class $i \in \mathcal{S}$ by $p_i(\vec{x})$. The design of a retrieval system in a probabilistic framework, consists of finding an appropriate map $g : \mathcal{F} \mapsto \mathcal{S}$. These maps constitute the set of similarity functions.

The goal of a probabilistic CBIR system is the *minimization of the probability of retrieval error*, that is, the probability $P(g(\vec{X}) \neq s)$. Hence, if we provide the system with a set of feature vectors $\vec{X}$ drawn from class s, we want to minimize the probability that the system will return images from a class $g(\vec{X})$ different from s. It can be shown [32] that the optimal similarity function, that is, the one minimizing $P(g(\vec{X}) \neq s)$, is the Bayes or maximum a-posteriory (MAP) classifier

$$\begin{aligned} g^*(\vec{X}) &= arg \max_i P(s = i|\vec{X}) \\ &= arg \max_i p(\vec{X}|s = i)P(s = i), \end{aligned} \tag{1}$$

where $p(\vec{X}|s = i)$ is the likelihood for the i-th class and $P(s = i)$ its prior probability. Under the assumption that all classes are a-priori equally likely, the MAP classifier reduces to the ML classifier:

$$\begin{aligned} g^*(\vec{X}) \ &= \ arg \max_i p(\vec{X}|s = i) \\ &\overset{i.i.d.}{=} arg \max_i \frac{1}{N} \sum_{j=1}^{N} \log p(\vec{x}_j|s = i) \ . \end{aligned} \tag{2}$$

When the number N of feature vectors is large, application of the Weak Law of Large Numbers [33] to (2) results in the following equation:

$$g^*(\vec{X}) = arg \min_i \underbrace{\int p_q(\vec{x}) \log \frac{p_q(\vec{x})}{p_i(\vec{x})} \, \mathrm{d}\vec{x}}_{D(p_q \| p_i)}, \tag{3}$$

where $D(p_q \| p_i)$ denotes the *Kullback-Leibler divergence* or *relative entropy* between the two densities, $p_q(\cdot)$ and $p_i(\cdot)$.

The problem of retrieving the top M images similar to a given query image, can be formulated as a *multiple hypothesis problem*. The query image I_q is represented by a feature data set, $\vec{X} = \{x_1, \ldots, x_N\}$, obtained after a transformation step, and each image in the database, I_i $(i = 1, \ldots, C)$, is assigned with a hypothesis H_i. Therefore, the problem of retrieving the top M images consists of selecting the M images in the database that are closer in terms of best hypotheses to the data $\vec{X}$ of the given query image.

Under the assumption that all hypotheses are a-priori equally likely, the optimum rule resulting in the minimum probability of retrieval error, is to select the hypotheses with the highest likelihoods among the C. Thus, the top M matches correspond to the M hypotheses, $H_{i_1}, H_{i_2}, \ldots, H_{i_M}$ for which $p(\vec{X}|H_{i_1}) \geq \cdots \geq p(\vec{X}|H_{i_M}) \geq p(\vec{X}|H_i),\ i \notin \{i_1, \ldots, i_M\}$.

A computationally efficient implementation of this setting is to adopt a *parametric* approach. Then, each conditional PDF, $p(\vec{X}|H_i)$ is modeled by a member of a family of PDFs, denoted by $p(\vec{X}; \boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ is a set of model parameters to be specified. In this framework, the extracted signature for the image I_i is the *estimated model parameter* $\hat{\boldsymbol{\theta}}_i$, computed in the FE step. Then, implementation of (3) gives the optimal rule for retrieving the top M similar images to the given query image I_q:

1. Compute the KLDs between the query density $p(\vec{X}; \boldsymbol{\theta}_q)$ and the density $p(\vec{X}; \boldsymbol{\theta}_i)$ associated with image I_i in the database, $\forall\, i = 1, \ldots, C$:
$$D(p(\vec{X}; \boldsymbol{\theta}_q) \| p(\vec{X}; \boldsymbol{\theta}_i)) = \int p(x; \boldsymbol{\theta}_q) \log \frac{p(x; \boldsymbol{\theta}_q)}{p(x; \boldsymbol{\theta}_i)} \, \mathrm{d}x \ . \tag{4}$$
2. Retrieve the M images corresponding to the M smallest values of the KLD.

The KLD in (4) can be computed using consistent estimators $\hat{\boldsymbol{\theta}}_q$ and $\hat{\boldsymbol{\theta}}_i$, for the model parameters. The ML estimator is a consistent estimator [5] and for the query image it gives:

$$\hat{\boldsymbol{\theta}}_q = arg \max_{\boldsymbol{\theta}} \log p(\vec{X}; \boldsymbol{\theta}) \ . \tag{5}$$

We can also apply a *chain rule* [33], in order to combine the KLDs from multiple data sets. This rule states that the KLD between two joint PDFs, $p(\vec{X}, \vec{Y})$ and $q(\vec{X}, \vec{Y})$, where $\vec{X}, \vec{Y}$ are assumed to be independent data sets, is given by

$$D(p(\vec{X}, \vec{Y}) \| q(\vec{X}, \vec{Y})) = D(p(\vec{X}) \| q(\vec{X})) + D(p(\vec{Y}) \| q(\vec{Y})). \tag{6}$$

III. Statistical Modeling of Wavelet Subband Coefficients via Joint sub-Gaussian Distributions

In this section, we introduce the family of multivariate sub-Gaussian distributions justifying this choice in terms of an accurate approximation of the marginal and joint densities of the transform coefficients.

A. The family of multivariate sub-Gaussian distributions

We first give the definition for the family of univariate symmetric alpha-stable ($S\alpha S$) distributions, before introducing the family of multivariate sub-Gaussian distributions. The $S\alpha S$ distribution is best defined by its characteristic function [34]:

$$\phi(t) = \exp(\imath\delta t - \gamma^{\alpha}|t|^{\alpha}), \tag{7}$$

where α is the *characteristic exponent*, taking values $0 < \alpha \leq 2$, δ ($-\infty < \delta < \infty$) is the *location parameter*, and γ ($\gamma > 0$) is the *dispersion* of the distribution. The characteristic exponent is a shape parameter, which controls the "thickness" of the tails of the density function. The smaller the α, the heavier the tails of the $S\alpha S$ density function. The dispersion parameter determines the spread of the distribution around its location parameter, similar to the variance of the Gaussian. A $S\alpha S$ distribution is called *standard* if $\delta = 0$ and $\gamma = 1$. The notation $X \sim f_{\alpha}(\gamma, \delta)$ means that the random variable X follows a $S\alpha S$ distribution with parameters α, γ, δ.

In general, no closed-form expressions exist for most $S\alpha S$ density and distribution functions. Two important special cases of $S\alpha S$ densities with closed-form expressions are the Gaussian ($\alpha = 2$) and the Cauchy ($\alpha = 1$). Unlike the Gaussian density which has exponential tails, stable densities have tails following an algebraic rate of decay ($P(X > x) \sim Cx^{-\alpha}$, as $x \to \infty$, where C is a constant depending on the model parameters), hence random variables following $S\alpha S$ distributions with small α values are highly impulsive.

An important characteristic of non-Gaussian $S\alpha S$ distributions is the non-existence of second-order moments. Instead, all moments of order p less than α do exist and are called the *Fractional Lower Order Moments* (FLOM's). In particular, the FLOM's of a $S\alpha S$ random variable $X \sim f_{\alpha}(\gamma, \delta = 0)$, are given by [18]:

$$E\{|X|^p\} = \big(C(p,\alpha)\cdot\gamma\big)^p \ , \quad 0 < p < \alpha, \tag{8}$$

where

$$\big(C(p,\alpha)\big)^p = \frac{2^{p+1}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(-\frac{p}{\alpha}\right)}{\alpha\sqrt{\pi}\,\Gamma\left(-\frac{p}{2}\right)} = \frac{\Gamma\big(1-\frac{p}{\alpha}\big)}{\cos\big(\frac{\pi}{2}p\big)\Gamma(1-p)}. \tag{9}$$

The $S\alpha S$ model parameters (α, γ) can be estimated using the consistent Maximum Likelihood (ML) method described by Nolan [35], which gives reliable estimates and provides the tightest confidence intervals.

Extending the $S\alpha S$ model to heavy-tailed random vectors leads to the *multivariate sub-Gaussian* $S\alpha S$ distribution[1] [18].

Definition 1 *Any vector $\vec{X}$ distributed as $\vec{X} = A^{1/2}\vec{G}$, where A is a positive $\frac{\alpha}{2}$-stable random variable and $\vec{G} = (G_1, G_2, \ldots, G_n)$ is a zero-mean Gaussian random vector, independent of A, with covariance matrix $\boldsymbol{R}$, is called a sub-Gaussian $S\alpha S$ random vector (in $\mathbb{R}^n$) with underlying Gaussian vector $\vec{G}$.*

[1]In the following, instead of saying sub-Gaussian $S\alpha S$ variable / vector / distribution, we simply use the term sub-Gaussian variable / vector / distribution.

A multivariate sub-Gaussian distribution, with underlying covariance matrix $\mathbf{R}$, is often denoted by α-SG($\mathbf{R}$). In this work, the transform coefficients at different subbands are tied up in vectors and are assumed to be samples of an α-SG($\mathbf{R}$) distribution, which can be viewed as a variance mixture of Gaussian processes [36].

It is important to note that covariances do not exist for the family of $S\alpha S$ random variables, due to the lack of finite variance. Instead, we measure correlation between transform coefficients using a quantity called *covariation* [18], which plays an analogous role for $S\alpha S$ random variables to the one played by covariance for Gaussian random variables. Let X and Y be jointly $S\alpha S$ random variables with $1 < \alpha \leq 2$, zero location parameters and dispersions γ_X and γ_Y respectively. Then, for all $1 < p < \alpha$, the covariation of X with Y is given by

$$[X,Y]_{\alpha} = \frac{E\{XY^{<p-1>}\}}{E\{|Y|^p\}}\gamma_Y^{\alpha}, \tag{10}$$

where for any real number z and $a \geq 0$ we use the notation

$$z^{<a>} = \begin{cases} z^a, & z > 0 \\ 0, & z = 0 \\ -(-z)^a, & z < 0. \end{cases}$$

The *covariation coefficient* of X with Y, is defined by

$$\lambda_{XY} = \frac{[X,Y]_{\alpha}}{[Y,Y]_{\alpha}} = \frac{E\{XY^{<p-1>}\}}{E\{|Y|^p\}}. \tag{11}$$

Note the asymmetric nature of the covariation and the covariation coefficient, as opposed to the usual second-order moments.

Consider the sub-Gaussian random vector $\vec{X} = A^{1/2}\vec{G}$, where $\vec{G} = (G_1, G_2, \ldots, G_n)$ the underlying Gaussian vector with covariance matrix $\mathbf{R}$. Then, the covariations between the components of $\vec{X}$, $[X_i, X_j]_{\alpha}$ $i, j = 1, \ldots, n$, are given by [18]:

$$\mathbf{c}_{ij} = [X_i, X_j]_{\alpha} = 2^{-\frac{\alpha}{2}}\,[\mathbf{R}]_{ij}\,[\mathbf{R}]_{jj}^{\frac{(\alpha-2)}{2}}. \tag{12}$$

Note that $\mathbf{c}_{ij} = \mathbf{c}_{ji}$ only if $[\mathbf{R}]_{ii} = [\mathbf{R}]_{jj}$. During the FE step, it is necessary to estimate the covariations from the transform coefficients of the images. In the next section, we describe how this estimation is performed.

B. Estimation of covariations

By applying (8) on Y we have

$$\gamma_Y = \frac{\big(E\{|Y|^p\}\big)^{1/p}}{C(p,\alpha)} \ . \tag{13}$$

Let the vectors $\{\vec{X}^1, \vec{X}^2, \ldots, \vec{X}^N\}$ constitute a set of N independent realizations of an α-SG($\mathbf{R}$) distribution, where $\vec{X}^k = (X_1^k, X_2^k, \ldots, X_n^k)$, $k = 1, \ldots, N$. Now, observe that we can find an estimation of $[X,Y]_{\alpha}$ by multiplying an estimated value of λ_{XY} and the ML estimation of γ_Y. The value of λ_{XY} is estimated via the *Fractional Lower Order Moment (FLOM) Estimator* [37], which is very simple and computationally efficient, in addition to being unbiased and consistent. For two jointly $S\alpha S$ random variables X, Y with $\alpha > 1$, and

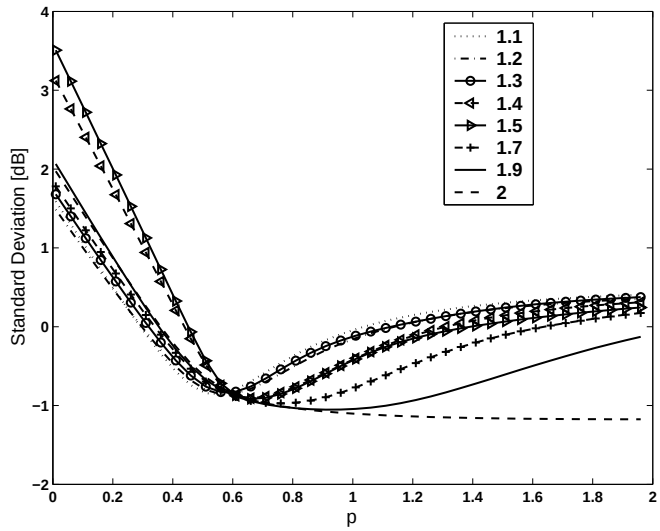


Fig. 1. Curves representing the standard deviation of the covariation estimation as a function of the parameter p for the $\hat{\mathbf{c}}^{FLOM}$ estimator.

TABLE I

PERFORMANCE OF THE COVARIATION ESTIMATOR.

α	$\hat{\mathbf{c}}^{FLOM}(p)$		True $[X,Y]_\alpha$
1.1	-1.9158 (0.1397)	(p=0.56)	-1.916
1.2	-1.8203 (0.1390)	(p=0.57)	-1.8254
1.3	-1.755 (0.1481)	(p=0.59)	-1.7476
1.4	-1.6814 (0.1234)	(p=0.66)	-1.6821
1.5	-1.6201 (0.1218)	(p=0.68)	-1.6285
1.7	-1.5533 (0.1062)	(p=0.76)	-1.5561
1.9	-1.5109 (0.0886)	(p=0.91)	-1.5288
2	-1.5301 (0.0668)	(p=2)	-1.532

a set of *independent* observations $(X_1, Y_1), \ldots, (X_n, Y_n)$, the FLOM estimator is defined as follows:

$$\hat{\lambda}_{FLOM} = \frac{\sum_{i=1}^{n} X_i |Y_i|^{p-1}\, sign(Y_i)}{\sum_{i=1}^{n} |Y_i|^p}. \tag{14}$$

Thus, the covariation estimator between the components of a sub-Gaussian vector $\vec{X}$ is given by:

$$\hat{\mathbf{c}}_{ij}^{FLOM} = \frac{\sum_{k=1}^{N} \vec{X_i^k}\, |\vec{X_j^k}|^{p-1}\, sign(\vec{X_j^k})}{\sum_{k=1}^{N} |\vec{X_j^k}|^p}\, \gamma_{X_j}^{\alpha}, \tag{15}$$

where the dispersion γ_{X_j} can be estimated using the ML estimator, described in [35].

We define the *covariation matrix* $\mathbf{C}$, the matrix having as elements the covariations. Then, the estimated covariation matrix $\hat{\mathbf{C}}$, is the matrix with elements $[\hat{\mathbf{C}}]_{ij} = \hat{\mathbf{c}}_{ij}^{FLOM}$. Once these covariations are estimated from the data, we can estimate the elements $[\mathbf{R}]_{ij}$ of the underlying covariance matrix, $\mathbf{R}$, using (12):

$$[\hat{\mathbf{R}}]_{jj} = \left(2^{\frac{\alpha}{2}}\, [\hat{\mathbf{C}}]_{jj}\right)^{\frac{2}{\alpha}}, \quad [\hat{\mathbf{R}}]_{ij} = 2^{\frac{\alpha}{2}}\, \frac{[\hat{\mathbf{C}}]_{ij}}{[\hat{\mathbf{R}}]_{jj}^{\frac{(\alpha-2)}{2}}}, \tag{16}$$

which are consistent and asymptotically normal, that is, the distribution of the above estimators tends to a normal distribution, as the number of observations N tends to infinity.

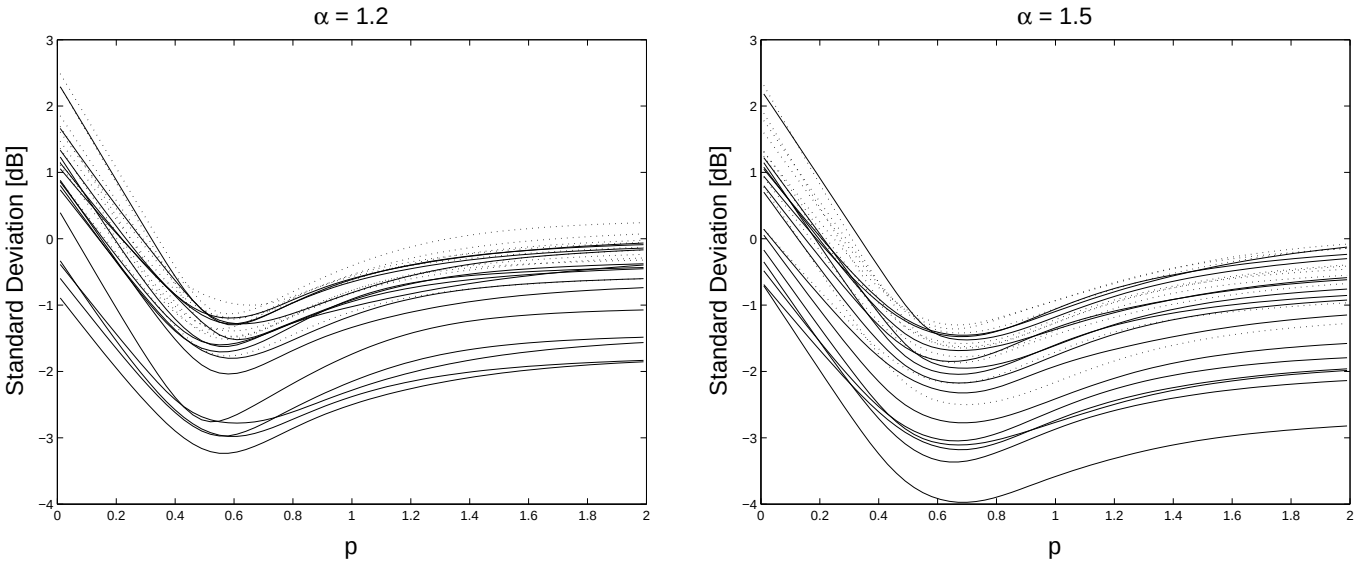
Notice that the estimation of covariations and consequently the estimation of the covariation matrices, requires the specification of the parameter p. We compute the optimal p as a function of the characteristic exponent α, by finding the value of p that minimizes the standard deviation of the estimator, for different values of $\alpha > 1$. For this purpose, we studied the influence of the parameter p on the performance of the covariation estimator given by (15) via Monte-Carlo simulations.

We generated two real $S\alpha S$ $(1 < \alpha \leq 2)$ random variables, $X = a_1 U + b_1 V$, $Y = a_2 U + b_2 V$, where U and V are independent, standard $S\alpha S$ random variables and $\{a_i,\, b_i,\, i = 1, 2\}$ are real coefficients. The true covariation of X with Y is $[X,Y]_\alpha = a_1 a_2^{<\alpha-1>} + b_1 b_2^{<\alpha-1>}$. We generated $N = 5000$ independent samples of U and V and calculated the covariation estimator by means of (15) for different values of p in the range $(0, 2]$. We ran $K = 1000$ Monte-Carlo simulations for different values of $\alpha \in (1, 2]$. We randomly selected, without loss of generality, the coefficient values to be equal to $a_1 = 0.32$, $a_2 = -2.45$, $b_1 = -1.7$, $b_2 = 0.44$. Fig. 1 displays the standard deviation of the estimator $\hat{\mathbf{c}}^{FLOM}(p)$ as a function of the parameter p and for different values of α.

Table I shows results on the performance of the estimator. We include the mean of the estimator, the standard deviation in parentheses and the value of p for which the smallest standard deviation is achieved by the estimator. We also note that we obtained similar experimental results for different values of the coefficients $\{a_i,\, b_i,\, i = 1, 2\}$.

In our proposed CBIR system, we need to estimate the covariations between the components of the sub-Gaussian vectors, which are special cases of $S\alpha S$ random variables. We repeated the above Monte-Carlo simulations using two sub-Gaussian random variables, $X = A^{1/2} G_X$, $Y = A^{1/2} G_Y$. By definition, X and Y can be viewed as $S\alpha S$ random variables with dispersion γ_X and γ_Y, respectively. We generate a sample of a sub-Gaussian random variable by first generating a sample A drawn from a $S_{\alpha/2}((\cos\frac{\pi\alpha}{4})^{2/\alpha},\, 1,\, 0)$ distribution and then by generating a sample G drawn from a zero-mean Gaussian distribution with variance $2\gamma^2$, which is viewed as a $S_2(\gamma,\, 0,\, 0)$ variable (with $\gamma = \gamma_X$ or $\gamma = \gamma_Y$ depending on whether the Gaussian part G corresponds to the variable X or Y, respectively).

Fig. 2 displays the curves representing the standard deviation of the $FLOM$ covariation estimator as a function of p, for two values of α and 25 pairs of dispersions $(\gamma_X,\, \gamma_Y)$, with the dispersions ranging in the interval $(0,\, 3.5)$, which corresponds to the dispersions estimated from the wavelet subbands of some selected images used in our experiments (obtained from



(a) (b)

Fig. 2. Curves representing the standard deviation of the covariation estimation as a function of the parameter p for $\alpha = 1.2$, 1.5 and 25 dispersion pairs (γ_X, γ_Y), using the $\hat{\mathbf{c}}^{FLOM}$ estimator.

the USC, SIPI database[2], cf. Fig. 7).

For each α, we observe that all the curves are minimized in a common interval on the p-axis, and actually the optimal values for p are close to each other. We repeated the procedure for $\alpha = 1 : 0.05 : 2$ and for a given α_i, we defined the optimal p_i as the mean of the optimal p values of its corresponding 25 curves, corresponding to the 25 pairs of dispersions. In section V, we use the values, shown in Table II, for the optimal p as a function of α. This table is used as a lookup table in order to find the optimal p for every $1 < \alpha \leq 2$ by linearly interpolating these values.

TABLE II
OPTIMAL p PARAMETER AS A FUNCTION OF THE CHARACTERISTIC EXPONENT α.

α	Optimal p	α	Optimal p
1	0.52	1.5	0.69
1.05	0.54	1.55	0.71
1.1	0.56	1.6	0.72
1.15	0.57	1.65	0.74
1.2	0.58	1.7	0.76
1.25	0.59	1.75	0.79
1.3	0.61	1.8	0.81
1.35	0.62	1.85	0.84
1.4	0.64	1.9	0.88
1.45	0.66	1.95	0.93
		2	0.8

C. Joint sub-Gaussian modeling of wavelet coefficients

In this section, we justify the selection of the family of sub-Gaussian distributions as a statistical modeling tool for the wavelet coefficients of texture images and, as an example, we show results on modeling data obtained by applying standard 2-D, orthogonal, discrete wavelet transform (DWT) on real texture images. Similar results are obtained when using other types of wavelet transforms, such as a steerable wavelet transform, which is more convenient to achieve rotation invariance.

The 2-D orthogonal DWT expands an image using a certain basis, whose elements are scaled and translated versions of a single prototype filter. In particular, the DWT decomposes images in dyadic scales, providing at each resolution level one low-pass subband approximation and three spatially oriented wavelet subbands. There are interesting properties of the wavelet transform [7] that justify its use in CBIR systems: *Locality* (image content is localized in both space and frequency), *multiresolution* (image is decomposed at a nested set of dyadic scales), and *edge detection* (wavelet filters operate as local edge detectors). Because of these properties, the wavelet transforms of real-world images tend to be sparse, resulting in a large number of small magnitude coefficients and a small number of large magnitude coefficients. In our modeling, we employ all the subbands except the low-pass residual, since it does not present this sparsity behavior, but an average of the original image. Importantly, this property is in conflict with the Gaussian assumption, giving rise to peaky and heavy-tailed *non-Gaussian* marginal distributions of the wavelet subband coefficients, which leads us to use joint sub-Gaussian distributions.

In our proposed retrieval scheme, we proceed by using a statistical model that captures both wavelet subband marginal distributions and inter-subband correlations. Various experimental results have shown the importance of the cross-correlation of each subband with other orientations at the same decomposition level in characterizing the texture information [38].

Our joint modeling is performed by tying up the wavelet

[2] `http://sipi.usc.edu/services/database`

coefficients at the same or adjacent spatial locations, levels and subbands, to form a sub-Gaussian vector. This modeling of the subband coefficients preserves the heavy-tailed behavior of their marginal distributions. Notice that the components of a sub-Gaussian vector are highly dependent, as illustrated in [18], and this makes the joint sub-Gaussian model appropriate for capturing the cross-dependencies between different subbands, since around features, such as edges and lines, the wavelet coefficients at all subbands are dependent in the sense that they have high probability of being significant.

Next, we assess the effectiveness of a $S\alpha S$ density function for the approximation of the empirical density of the subband coefficients, near the mode and on the tails. In our data modeling, the statistical fitting proceeds in two steps: first, we assess whether the data deviate from the normal distribution and we determine if they have heavy tails by employing normal probability plots [39]. Then, we check if the data is in the stable domain of attraction by estimating the characteristic exponent α directly from the data and by providing the related confidence intervals. As a further stability diagnostics, we employ the amplitude probability density (APD) curves $(P(|X| > x))$ that give a good indication of whether the $S\alpha S$ fit matches the data near the mode and on the tails of the distribution.

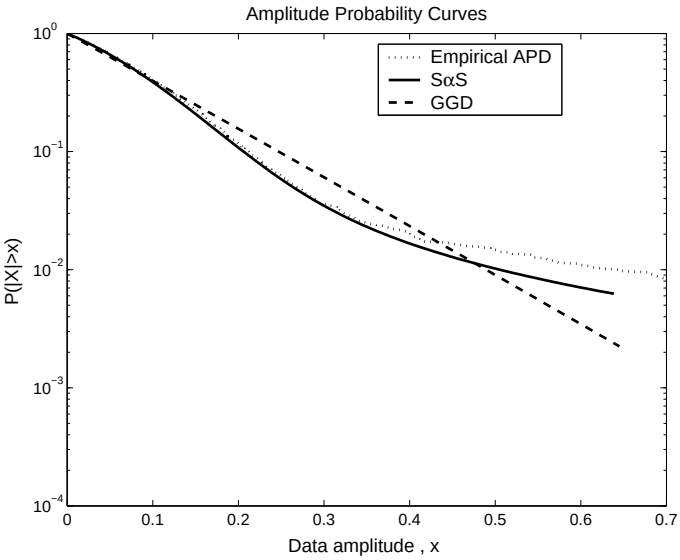


Fig. 3. Modeling of the horizontal subband at the first level of decomposition of the *Flowers.6* image with the $S\alpha S$ and the GGD depicted in solid and dashed lines, respectively. The estimated parameters for the $S\alpha S$ distribution have the values $\alpha = 1.76$, $\gamma = 0.08$ while the GGD has parameters $\alpha = 0.11$ and $\beta = 1.02$. The dotted line denotes the empirical APD.

Fig. 3 compares the $S\alpha S$ and GGD fits for a selected subband of a certain image. Clearly, the $S\alpha S$ density is superior to the GGD, following more closely both the mode and the tail of the empirical APD, than the exponentially decaying GGD. Table III shows the ML estimates of the characteristic exponent α together with the corresponding 95% confidence intervals, for a set of 10 textures (real-world 512×512 natural scene images) obtained from the MIT Vision Texture (VisTex) database, decomposed in 3-levels using Daubechies' 4 ('db4') filters [40]. It can be observed that the confidence intervals depend on the decomposition level. In particular, they become wider as the level increases since the number of samples used for estimating the $S\alpha S$ parameters gets smaller because of the subsampling that takes place between scales. This table also demonstrates that the coefficients of different subbands and decomposition levels exhibit various degrees of non-Gaussianity, with values of α varying between 0.9 (close to Cauchy) and 2 (close to Gaussian).

TABLE III

$S\alpha S$ MODELING OF WAVELET SUBBAND COEFFICIENTS OF TEXTURE IMAGES FROM THE VISTEX DATABASE, USING DAUBECHIES' 4 FILTER AND 3 DECOMPOSITION LEVELS. ML PARAMETER ESTIMATES AND 95% CONFIDENCE INTERVALS FOR THE CHARACTERISTIC EXPONENT α.

IMAGE	Image Subbands		
	Horizontal	Vertical	Diagonal
	Level 1		
Bark.10	1.601 ± 0.061	1.684 ± 0.058	1.681 ± 0.057
Brick.1	1.614 ± 0.056	1.577 ± 0.056	1.896 ± 0.031
Buildings.4	1.681 ± 0.039	1.684 ± 0.047	1.601 ± 0.046
Fabric.0	2.000 ± 0.007	1.270 ± 0.053	1.322 ± 0.055
Fabric.10	1.229 ± 0.057	1.367 ± 0.058	1.175 ± 0.054
Flowers.6	1.76 ± 0.041	1.701 ± 0.044	1.986 ± 0.025
Food.9	1.626 ± 0.061	1.339 ± 0.055	1.386 ± 0.056
Grass.1	1.879 ± 0.047	1.853 ± 0.053	1.791 ± 0.055
Metal.4	1.320 ± 0.057	1.228 ± 0.054	1.402 ± 0.059
Stone.3	1.591 ± 0.054	1.688 ± 0.049	1.746 ± 0.050
	Level 2		
Bark.10	1.855 ± 0.097	1.858 ± 0.110	1.850 ± 0.107
Brick.1	1.311 ± 0.105	1.539 ± 0.103	1.850 ± 0.101
Buildings.4	0.921 ± 0.089	1.186 ± 0.119	1.151 ± 0.098
Fabric.0	2.000 ± 0.006	1.171 ± 0.099	1.349 ± 0.114
Fabric.10	1.677 ± 0.120	1.611 ± 0.115	1.557 ± 0.114
Flowers.6	1.334 ± 0.097	1.424 ± 0.100	1.900 ± 0.057
Food.9	1.990 ± 7.7e-8	1.465 ± 0.107	1.750 ± 0.101
Grass.1	1.921 ± 0.073	1.858 ± 0.099	1.990 ± 0.073
Metal.4	1.680 ± 0.121	1.505 ± 0.119	1.690 ± 0.119
Stone.3	1.869 ± 0.083	1.509 ± 0.103	1.658 ± 0.117
	Level 3		
Bark.10	1.723 ± 0.225	2.000 ± 0.162	1.792 ± 0.248
Brick.1	1.227 ± 0.242	1.368 ± 0.204	1.990 ± 0.219
Buildings.4	1.220 ± 0.201	2.000 ± 0.109	1.014 ± 0.181
Fabric.0	2.000 ± 0.097	1.498 ± 0.246	1.874 ± 0.167
Fabric.10	1.865 ± 0.205	2.000 ± 0.103	1.904 ± 0.184
Flowers.6	1.643 ± 0.222	1.573 ± 0.210	1.851 ± 0.211
Food.9	2.000 ± 0.005	1.677 ± 0.222	1.990 ± 0.245
Grass.1	2.000 ± 0.375	1.862 ± 0.177	2.000 ± 0.241
Metal.4	1.787 ± 0.217	1.888 ± 0.145	1.863 ± 0.182
Stone.3	2.000 ± 0.096	1.278 ± 0.223	1.500 ± 0.226

IV. ROTATION-INVARIANT CBIR WITH GAUSSIANIZED STEERABLE PYRAMIDS

The property of rotation invariance is very desirable in a texture retrieval system. An important problem with the standard wavelet transform is that it lacks the translation and rotation invariant properties. This results in a mismatch of the retrieval process when the image orientation varies. In fact, the

wavelet coefficients of the rotated image will be completely different, in the sense that they will not be simply rotated versions of the wavelet coefficients of its original version.

A way to overcome this problem is to replace the standard wavelet transform with a steerable pyramid [41], [42], which is a linear multi-scale, multi-orientation image decomposition produced by a set of orientation filters, generated by a set of basis functions (directional derivative operators). Steerable pyramids are overcomplete and possess the desired properties of rotation invariance and (approximate) translation invariance.

In this section, we design a rotation-invariant CBIR technique, which is based on the joint sub-Gaussian modeling of a steerable pyramid coefficients, incorporating dependence across space, angles and scales. In particular, we construct a steerable model, relating the fractional lower-order statistics of a rotated image with that of its original version, and then apply a Gaussianization process on the steerable model by employing the local statistical behavior of the coefficients, which are grouped into appropriate spatial neighborhoods. The similarity measurement between two images is performed by deriving a rotation-invariant similarity function, which effectively performs angular alignment between the images.

A. Steerability of the pyramid subband coefficients

In the case of a database containing images along with rotated versions of them, we are interested in finding features which are as "steerable" as possible, that is, given the features of an image oriented at an angle ϕ, we should be able to obtain the features corresponding to the same image rotated at an angle θ, without having to re-extract the features from the rotated image[3].

Let $c^l(x_k,\phi)$ represent the value of a transform coefficient at a spatial location x_k $(k=1,\ldots,N)$, orientation ϕ and level l $(l=1,\ldots,L)$. In a steerable pyramid with J *basic* orientations (subbands), at each level l, given the J basic coefficients $X_k^l=[c^l(x_k,\phi_1),c^l(x_k,\phi_2),...,c^l(x_k,\phi_J)]^T$, the transform coefficient $c^l(x_k,\phi)$ for any angle ϕ is given by [29]:

$$c^l(x_k,\phi)=\sum_{i=1}^{J} f_i(\phi)c^l(x_k,\phi_i) \quad \forall\,\phi, \quad l=1,...,L \tag{17}$$

where $\{f_1(\phi),f_2(\phi),...,f_J(\phi)\}$ is the set of J steering functions.

Let $\mathbf{R}^l$ and $\mathbf{R}^l_\theta$ denote the sampled correlation matrices, with elements given by the correlations between pairs of subbands (at a given decomposition level l) of the original image I and its rotated version I_θ, respectively. The following proposition [29] establishes the relation between $\mathbf{R}^l$ and $\mathbf{R}^l_\theta$.

Proposition 1 ([29]) *The matrices $\mathbf{R}^l_\theta$ and $\mathbf{R}^l$ are related as follows:*

$$\mathbf{R}^l_\theta=\mathbf{F}(\theta)\mathbf{R}^l\mathbf{F}^T(\theta), \tag{18}$$

where

$$\mathbf{F}(\theta)=\begin{bmatrix} f_1(\phi_1-\theta) & f_2(\phi_1-\theta) & \cdots & f_J(\phi_1-\theta)\\ f_1(\phi_2-\theta) & f_2(\phi_2-\theta) & \cdots & f_J(\phi_2-\theta)\\ \vdots & \vdots & \vdots & \vdots\\ f_1(\phi_J-\theta) & f_2(\phi_J-\theta) & \cdots & f_J(\phi_J-\theta)\end{bmatrix}. \tag{19}$$

Proof: The proof of Proposition 1 follows easily by direct computation and making use of the properties of the steering functions $\{f_i(\theta)\}$. ■

In our work, the J basic angles are taken to be equispaced, which makes $\mathbf{F}(\theta)$ an orthogonal matrix for any θ, i.e., $\mathbf{F}^T(\theta)=\mathbf{F}^{-1}(\theta)$ $(=\mathbf{F}(-\theta))$, and thus, in this case, $\mathbf{R}^l$ and $\mathbf{R}^l_\theta$ become orthogonally equivalent.

Under a joint sub-Gaussian assumption, the coefficients of the J basic orientations (subbands) at a given level l are modeled as joint sub-Gaussian vectors α-SG($\mathbf{R}^l$), with $\mathbf{R}^l$ denoting the underlying covariance matrix corresponding to the subbands at the lth-level.

The pyramid coefficients at a given subband are assumed to follow a sub-Gaussian marginal distribution. So, the coefficients corresponding to the basic orientation ϕ_i at level l can be expressed as:

$$c^l(x_k,\phi_i)=\sqrt{A}\,c^l_G(x_k,\phi_i), \quad i=1,...,J, \tag{20}$$

where $c^l_G(x_k,\phi_i)$ is the Gaussian part of the α-SG($\mathbf{R}^l$) vector. From (17), the transform coefficient at level l, at any angle ϕ is:

$$\begin{aligned} c^l(x_k,\phi) &= \sum_{i=1}^{J} f_i(\phi)\big(\sqrt{A}\,c^l_G(x_k,\phi_i)\big)\\ &= \sqrt{A}\sum_{i=1}^{J} f_i(\phi)c^l_G(x_k,\phi_i)=\sqrt{A}\,c^l_G(x_k,\phi) \end{aligned} \tag{21}$$

Notice that (21) shows that the pyramid subband coefficients of a rotated image at an angle ϕ, are also sub-Gaussian random variables with the same characteristic exponent as that one of the corresponding subbands of the original (non-rotated) image, and with a Gaussian part which is the rotated version of the original Gaussian part at the same angle ϕ. Therefore, it can be seen that if one is able to estimate accurately the multiplier $\sqrt{A}$, it would be possible to normalize the coefficients $c^l(x_k,\phi)$ dividing them by $\sqrt{A}$, and work with the Gaussianized coefficients $c^l_G(x_k,\phi)$. This is convenient because, as it will become clearer later, it is easier to use appropriate and simple (analytical) similarity functions with the Gaussianized coefficients.

In order to accurately estimate the multiplier A, we consider dependence across orientations, scales and space, which results in an improved statistical model for natural images. We achieve this by defining an appropriate neighborhood for each coefficient, which is then modeled as a sub-Gaussian random vector. This joint sub-Gaussian modeling is followed by a Gaussianization procedure, which results in a steerable pyramid whose coefficients are jointly Gaussian (Gaussianized steerable pyramid).

There are several reasons that justify the Gaussianization step:

[3]Through the next sections, we consider counter-clockwise rotation.

a) the normalized transform domain can be well modeled statistically, using only second-order covariances between pairs of subbands,
b) the similarity measurement can be performed using an analytical expression for the KLD between two multivariate Gaussian distributions, avoiding computationally complex methods, such as the Monte-Carlo method,
c) the normalized pyramid allows to perform easily steerability in the feature space.

B. Variance-adaptive local modeling using multivariate sub-Gaussian distributions

The dependencies between the coefficients forming a certain neighborhood, including in general coefficients located at a small spatial region and at different orientations and scales, can be modeled using a homogeneous random field with a spatially changing variance. This requirement can be realized by modulating the vector of coefficients constituting the neighborhood (node of the field) with a hidden scaling random variable (multiplier), as follows:

$$\vec{X} \stackrel{d}{=} \sqrt{A}\,\vec{G} \tag{22}$$

where $\vec{G}$ is a zero-mean Gaussian random vector and A a positive scalar variable independent of $\vec{G}$ ($\stackrel{d}{=}$ denotes equality in distribution). A vector $\vec{X}$ that can be written like this, is said to follow a Gaussian Scale Mixture (GSM) distribution [43]. Notice that when the multiplier A is drawn from a $S\alpha S$ distribution, this is exactly the case of a multivariate α-sub-Gaussian model.

Two basic assumptions are made in order to reduce the dimensionality of these models: (i) the probability structure is defined *locally*. In particular, the probability density of a coefficient when conditioned on the rest of neighbors, is independent of the coefficients outside the neighborhood, (ii) all such neighborhoods obey the same distribution (*spatial homogeneity*).

The construction of a global probabilistic model for images, based on these local descriptions, needs the specification of a *neighborhood structure* for each subband coefficient, and the *distribution of the multipliers*, which we have already specified that it is a member of the family of $S\alpha S$ distributions. We extract the interdependencies between coefficients at different orientations, levels and spatial positions, by utilizing their joint α-sub-Gaussian statistics: Let $\vec{X}_k^{l,\,i}$ denote a generic P-dimensional neighborhood of the coefficient $c^l(x_k, \phi_i)$ at the spatial position x_k ($k = 1, \ldots, K$), orientation ϕ_i ($i = 1, \ldots, J$) and level l ($l = 1, \ldots, L$). This neighborhood is supposed to be drawn of an α-SG($\mathbf{R}^l$) random vector.

C. Gaussianization of the multivariate sub-Gaussian model

An important property of a GSM model is that the probability density of a P-dimensional GSM vector $\vec{X}$ is Gaussian when conditioned on A. Combining this property with (22), it is clear that the normalized vector $\vec{X}/\sqrt{A}$ follows a joint Gaussian distribution. The probability density of $\vec{X}$ conditioned on A is given by:

$$p(\vec{X}|A) = \frac{\exp(-\vec{X}^T (A\mathbf{R})^{-1} \vec{X}/2)}{(2\pi)^{P/2}\ |A\mathbf{R}|^{1/2}}\ . \tag{23}$$

From (23), it can be seen that the ML estimator for the multiplier A is

$$\hat{A}(\vec{X}) = \frac{\vec{X}^T \mathbf{R}^{-1} \vec{X}}{P}, \tag{24}$$

where the estimator is explicitly written as a function of $\vec{X}$ to emphasize the assumption of locality. This simplifies the computational procedure for the Gaussianization of the steerable pyramid subband coefficients, as we assume that the multipliers associated with different neighborhoods are estimated independently, even though the neighborhoods are overlapping.

In our implementation, we estimate, as explained in section III-B, the underlying covariance matrix $\mathbf{R}^{l,\,i}$, corresponding to the basic orientation ϕ_i at the lth-level, by employing the neighborhoods of all coefficients (or a subset of them, which is computationally efficient, at the cost of a reduced estimation accuracy) at the given orientation ($\vec{X}_k^{l,\,i}$, $k = 1, \ldots, N$). This procedure has the advantage of resulting in a computationally efficient way to estimate the hidden multiplier A and normalize the subband coefficients. Also, our technique avoids the use of a Gaussian Mixture Model (GMM), as in other approaches [25], which requires complicated Expectation-Maximization algorithms to estimate the multipliers, nested in a Markovian manner. We must also note that the multipliers in [25] are discrete, whereas in our model they vary in a continuous fashion.

Summarizing, the steps of our Gaussianization method are:

1. Decompose the given image into L levels and J orientations per level, via a steerable pyramid.
2. For each decomposition level l, $l = 1, \ldots, L$:
   For each orientation ϕ_i, $i = 1, \ldots, J$, at the l-th level:
   i) Estimate the covariance matrix $\mathbf{R}^{l,\,i}$ using (16).
   ii) For each coefficient $c^l(x_k, \phi_i)$, $k = 1, \ldots, K$:
   - Construct the corresponding neighborhood $\vec{X}_k^{l,\,i}$.
   - Estimate the multiplier $\hat{A}_k^{l,\,i}(\vec{X}_k^{l,\,i})$ using (24).
   - Compute the normalized coefficient $\tilde{c}^l(x_k, \phi_i) = c^l(x_k, \phi_i)/\sqrt{\hat{A}_k^{l,\,i}}$.

From (24), it is obvious that the estimation accuracy for the multiplier depends on the accurate estimation of the underlying covariance matrix $\mathbf{R}^{l,\,i}$ and the neighborhood structure.

D. Computation of inter-level covariations

The multiplier estimation, as well as the construction of an image signature which we describe later on, may require the involvement of coefficients or the computation of covariations between subbands at different levels. Using the standard pyramid decomposition, we move from level l to the next coarser level $(l+1)$ by subsampling the output of a low-pass filter. As a result, the subbands at the $(l+1)$-th level are $1/4$ in size than those of the l-th level (since we are dealing with images), which is undesired since the covariation estimation includes

TABLE IV

NEIGHBORHOOD SHAPES USED IN THE MEASUREMENT OF THE GAUSSIANIZATION PERFORMANCE.

Index	Neighborhood Structure for a given $c^l(x_k, \phi_i)$	Size (P)
1	(l, ϕ_i): (3x3) $\cup$ $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1 $\cup$ $(l+1, \phi_i)$: 1	$J+9$
2	(l, ϕ_i): (3x3) $\cup$ $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1	$J+8$
3	(l, ϕ_i): (3x3)	9
4	(l, ϕ_i): 4 (cross shape (c.s.))	5
5	(l, ϕ_i): 4 (c.s.) $\cup$ $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1 $\cup$ $(l+1, \phi_i)$: 1	$J+5$
6	(l, ϕ_i): 4 (c.s.) $\cup$ $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1	$J+4$
7	(l, ϕ_i): 4 (c.s.) $\cup$ $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 4 (c.s.) $\cup$ $(l+1, \phi_i)$: 4 (c.s.)	$5(J+1)$
8	(l, ϕ_i): (5x5) $\cup$ $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1 $\cup$ $(l+1, \phi_i)$: 1	$J+25$
9	(l, ϕ_i): (5x5) $\cup$ $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1	$J+24$
10	(l, ϕ_i): (5x5)	25

1) (l, ϕ_i): X, means that X is the neighborhood at the level l and orientation ϕ_i, centered at the coefficient $c^l(x_k, \phi_i)$.
2) $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: Y, means that for each one of the rest $(J-1)$ orientations at level l, Y is the neighborhood centered at the coefficient $c^l(x_k, \phi_j), \mathrm{J} = 1, \ldots, J,\ j \neq i$.
3) $(l+1, \phi_i)$: Z, means that Z is the neighborhood at the next lower level $l+1$ and orientation ϕ_i centered at the corresponding spatial location.

summations between vectors of equal length. Subsampling is good for compression and for saving memory and complexity when performing the decomposition. However, subsampling introduces some aliasing, which is not good for extracting features. There are two ways to avoid this:

a) Instead of subsampling the output of the filters, we can upsample the filters and perform the filtering without subsampling their outputs. In addition to avoiding aliasing, it allows us also to keep the same number of coefficients across scales, which favors the computation of covariation matrices across scales.

b) Using Fourier domain filtering, simply by multiplying the DFT of the image with the DFT of the filter and then taking the inverse DFT to obtain the subband coefficients. We follow the frequency-domain approach since, although both implementations should give approximately the same values, the frequency-domain implementation is more exact as it does not suffer from the finite-length constraint that is imposed on the filters for the convolution.

Besides, the estimation of covariation assumes two jointly sub-Gaussian random variables, i.e., with equal characteristic exponent values. The values of α, estimated from the different orientation subbands using a steerable pyramid without subsampling, are close to each other but not equal. This problem is overcome by making use of the asymmetry in the definition of the covariation:

- from (15), we observe that the free p parameter affects the second variable (as an exponent),
- the intuitive idea is to use the estimated characteristic exponent, α, corresponding to the second variable (subband) in order to estimate the covariation: for instance, in order to estimate $[X, Y]_\alpha$ we first estimate α from Y and then assume that X follows a distribution with the same α, while in order to estimate $[Y, X]_\alpha$ we estimate α from X and then assume that Y also follows a distribution with that α. This procedure exploits the differences between the subbands, regarding their distribution.

We also use this approach for the estimation of the covariations between subbands at the *same* decomposition level.

E. Neighborhood construction

There is a trade-off between the computational complexity and the neighborhood size. This can be seen from (24), where the estimated stable multiplier depends on the inverse of the underlying covariance matrix $\mathbf{R}$. The computational complexity increases as the neighborhood size P increases, since the complexity to estimate $\mathbf{R}$ and to calculate its inverse, $\mathbf{R}^{-1}$, depends on its dimension ($P \times P$). It is clear that is not computationally feasible to construct all possible neighborhoods for each subband coefficient in order to select the optimal neighborhood, because of the large amount of combinations.

In this section, we examine the performance of our Gaussianization procedure with respect to different neighborhoods, taking also into consideration the computational limitations. For this purpose, we implement the Gaussianization process using the neighborhoods shown in Table IV.

For a given subband coefficient $c^l(x_k, \phi_i)$, the formation of some of the above neighborhoods requires the inclusion of coefficients at the same spatial location of the subbands but at the next coarser scale. In order to associate coefficients between adjacent levels, we use the frequency-domain implementation of the steerable pyramid, without subsampling the output of the filters, as described in section IV-D. In this case, it results in subbands with equal size at all decomposition levels. Then, the coefficient at the corresponding spatial location of the next coarser subband is simply $c^{(l+1)}(x_k, \phi_i)$.

We tested the performance of our Gaussianization procedure with respect to the neighborhood structure, by implementing it on a set of 5, randomly selected textures of size 512×512 from the Brodatz database with the following code numbers: **1.** 1.1.01, **2.** 1.1.04, **3.** 1.1.08, **4.** 1.2.08, **5.** 1.5.04, along with their rotations at 30, 60, 90 and 120 degrees. We applied a 3-level pyramid decomposition with 4 orientations per level, thus

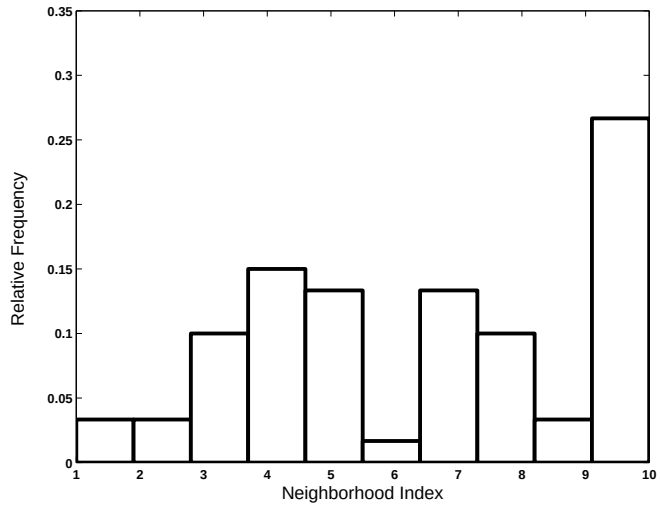


Fig. 4. Histogram of the neighborhood indices in terms of resulting in the best Gaussianization performance for a set of 300 subbands.

obtaining a total of 300 subbands. After the Gaussianization, at each subband, we calculated the relative entropy (ΔH) between the histogram (with 256 bins) and the Gaussian PDF fitting the normalized subband coefficients, as a fraction of the histogram entropy (H):

$$\frac{\Delta H}{H} = \frac{\sum_{k=1}^{256} h(x_k)\ \log\big(\frac{h(x_k)}{g(x_k)}\big)}{-\sum_{k=1}^{256} h(x_k)\ \log(h(x_k))}, \tag{25}$$

where $h(x_k)$ is the probability density of the center x_k of the k-th bin, as estimated from the histogram, and $g(x_k)$ is the corresponding value of the fitting Gaussian PDF with parameters estimated from the normalized coefficients. The best choice for the neighborhood structure corresponds to the smallest fraction $\Delta H/H$.

Fig. 4 displays the histogram of the neighborhood indices (cf. Table IV), for the 300 subbands. The vertical axis is the relative frequency of each neighborhood shape (horizontal axis), whose selection resulted in the smallest fraction (25) for the above subbands. For the given set of textures, the choice of the 10-th neighborhood shape results in the best Gaussianization performance for most of the pyramid subbands.

F. Feature Extraction

After the normalization procedure, the marginal and joint statistics of the coefficients at adjacent spatial positions, orientations and levels are close to the Gaussian distribution. Then, to extract the features, we simply compute the $J \times J$ covariance matrix at each decomposition level.

Thus, for a given image I, decomposed in L levels, a possible signature $\mathcal{S}^{\mathcal{G}}$ is given by the set of the L covariance matrices:

$$I \mapsto \mathcal{S}^{\mathcal{G}} = \{\mathbf{\Sigma}_I^1, \mathbf{\Sigma}_I^2, \ldots, \mathbf{\Sigma}_I^L\}, \tag{26}$$

where $\mathbf{\Sigma}_I^l$ is the covariance matrix of the l-th decomposition level. Due to the symmetric property of the covariance matrix, the total size of the above signature equals: $size(\mathcal{S}^{\mathcal{G}}) = \frac{J(J+1)}{2} \cdot L$. The signature $\mathcal{S}^{\mathcal{G}}$ contains only the across orientation, second-order dependence at a given decomposition level. Because of the strong dependence across scales and orientations, we may obtain a better (more complete) signature by considering in addition the across levels dependence as expressed by the covariance matrices between consecutive levels. In this case, the signature of an image I is the following:

$$I \mapsto \mathcal{S}_{\mathcal{E}}^{\mathcal{G}} = \{\mathbf{\Sigma}_I^1, \ldots, \mathbf{\Sigma}_I^L, \mathbf{\Sigma}_I^{1\to 2}, \ldots, \mathbf{\Sigma}_I^{(L-1)\to L}\}, \tag{27}$$

where $\mathbf{\Sigma}_I^{l\to(l+1)}$ denotes the (in general *asymmetric*) covariance matrix corresponding to the subbands at levels l and $(l+1)$. In particular, the element $[\mathbf{\Sigma}_I^{l\to(l+1)}]_{ij}$ is equal to the covariance of the i-th subband at level l with the j-th subband at the next lower level $(l+1)$. The enhanced signature $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$ contains more texture-specific information than the signature $\mathcal{S}^{\mathcal{G}}$ at the cost of an increased computational complexity, since its size equals

$$size(\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}) = \frac{J(J+1)}{2} \cdot L + \frac{J(J+1)}{2} \cdot (L-1)\ .$$

Notice that, although neighborhood 10, which does not include dependencies across levels, was best for Gaussianization, regarding the design of a retrieval scheme, the dependence across orientations and levels is very useful in extracting a more accurate profile of the texture information.

G. Similarity Measurement

In recent work [29], a Gaussian assumption for the marginal and joint distributions of the steerable pyramid coefficients results in a deterministic rotation-invariant similarity measure, between two images I and Q. If we use the signature $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$, this measure takes the following form:

$$D(I,Q) = \min_{\theta} \Bigg[\sum_{l=1}^{L} \|\mathbf{\Sigma}_I^l - \mathbf{F}(-\theta)\mathbf{\Sigma}_Q^l\mathbf{F}^T(-\theta)\| + \\ + \sum_{l=1}^{L-1} \|\mathbf{\Sigma}_I^{l\to l+1} - \mathbf{F}(-\theta)\mathbf{\Sigma}_Q^{l\to l+1}\mathbf{F}^T(-\theta)\| \Bigg], \tag{28}$$

where $\|\cdot\|$ denotes any of the common matrix norms (however, a good choice is the Frobenius norm, which gives an indication of the "matrix amplitude"). If the texture information is represented by the signature $\mathcal{S}^{\mathcal{G}}$, the similarity function is modified by omitting the second sum, which corresponds to inter-level dependencies.

In our method, we construct and employ a novel statistical rotation-invariant similarity function. After the Gaussianization procedure has been applied, we model the distribution of each decomposition level in the case of using $\mathcal{S}^{\mathcal{G}}$, as well as the joint distribution between consecutive levels in the case of using $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$, using a multivariate Gaussian density (MvGD). The similarity between two images is measured by employing the KLD between MvGDs. Consider the case in which the texture information of each image is expressed using the signature $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$, that is, each decomposition level, as well as each pair of adjacent decomposition levels, is associated with a covariance matrix.

Given two images I and Q, let I^l, Q^l be the set of orientation subbands at the l-th decomposition level and $I^{l,\,l+1}$,

$Q^{l,\,l+1}$ be the set of orientation subbands at two adjacent levels l and $l+1$, following zero-mean MvGDs with covariance matrices $\mathbf{\Sigma}_I^l$, $\mathbf{\Sigma}_Q^l$, and $\mathbf{\Sigma}_I^{l\to l+1}$, $\mathbf{\Sigma}_Q^{l\to l+1}$, respectively. The KLD between two corresponding levels is given by [5]:

$$D(I^l\|Q^l) = \frac{1}{2}\left(tr(\mathbf{\Sigma}_I^l(\mathbf{\Sigma}_Q^l)^{-1} - I) - \ln|\mathbf{\Sigma}_I^l(\mathbf{\Sigma}_Q^l)^{-1}|\right). \quad (29)$$

In the same way, the KLD between two corresponding pairs of adjacent levels is given by:

$$D(I^{l,\,l+1}\|Q^{l,\,l+1}) = \frac{1}{2}\Big(tr(\mathbf{\Sigma}_I^{l\to l+1}(\mathbf{\Sigma}_Q^{l\to l+1})^{-1} - I) - \\ - \ln|\mathbf{\Sigma}_I^{l\to l+1}(\mathbf{\Sigma}_Q^{l\to l+1})^{-1}|\Big). \quad (30)$$

We define the overall KLD between images I, Q to be equal to the following sum:

$$D(I\|Q) = \sum_{l=1}^{L} D(I^l\|Q^l) + \sum_{l=1}^{L-1} D(I^{l\to l+1}\|Q^{l\to l+1}). \quad (31)$$

In our problem, we deal with image databases which may contain rotated versions of a given image. Notice that from (18), it follows that:

$$\mathbf{\Sigma}_{Q_\theta}^l = \mathbf{F}(\theta)\mathbf{\Sigma}_Q^l\mathbf{F}^T(\theta)\,,\ \mathbf{\Sigma}_{Q_\theta}^{l\to l+1} = \mathbf{F}(\theta)\mathbf{\Sigma}_Q^{l\to l+1}\mathbf{F}^T(\theta), \quad (32)$$

where $\mathbf{\Sigma}_{Q_\theta}^l$ is the l-th level covariance matrix and $\mathbf{\Sigma}_{Q_\theta}^{l\to l+1}$ is the covariance matrix between levels l and $l+1$, of a rotated version of image Q at an angle θ, Q_θ.

Consider I to be the query image and $\tilde{Q} = Q_\phi$ to be a counter-clockwise rotation, by an angle ϕ, of the original image Q in the database. In a real application, of course, the value of ϕ is unknown. Thus, the distance between the l-th levels of I and $\tilde{Q}$ (I^l and $\tilde{Q}^l$, respectively) is defined as the minimum KLD between I^l and $\tilde{Q}^l_{-\theta}$, where the minimization is over a set of possible rotations Θ, and thus, it is necessary to perform an angular alignment by finding the optimum θ^*. The notation $\tilde{Q}_{-\theta}$ means a clockwise rotation of image $\tilde{Q}$. By noticing that $\mathbf{\Sigma}_{\tilde{Q}_{-\theta}}^l = \mathbf{F}(-\theta)\mathbf{\Sigma}_{\tilde{Q}}^l\mathbf{F}^T(-\theta)$ and substituting (32) into (29), we obtain that the KLD between the l-th level of an image I and a clockwise rotation of image $\tilde{Q}$ by an angle θ (I^l and $\tilde{Q}^l_{-\theta}$, respectively), is given by:

$$D(I^l\|\tilde{Q}^l_{-\theta}) = \frac{1}{2}\Big[tr\big(\mathbf{\Sigma}_I^l\mathbf{F}^T(\theta)(\mathbf{\Sigma}_{\tilde{Q}}^l)^{-1}\mathbf{F}(\theta) - I\big) - \\ - \ln(|\mathbf{\Sigma}_I^l||(\mathbf{\Sigma}_{\tilde{Q}}^l)^{-1}|)\Big]. \quad (33)$$

Similarly, we obtain the KLD between two corresponding pairs of adjacent levels, $D(I^{l\to l+1}\|\tilde{Q}^{l\to l+1}_{-\theta})$, by replacing the covariance matrices $\mathbf{\Sigma}_I^l$, $\mathbf{\Sigma}_{\tilde{Q}}^l$ with the covariance matrices $\mathbf{\Sigma}_I^{l\to l+1}$, $\mathbf{\Sigma}_{\tilde{Q}}^{l\to l+1}$, respectively. Finally, the overall KLD between I and $\tilde{Q}$ is defined as:

$$D(I\|\tilde{Q}) = \min_{\theta\in\Theta}\left(\sum_{l=1}^{L} D(I^l\|\tilde{Q}^l_{-\theta}) + \sum_{l=1}^{L-1} D(I^{l\to l+1}\|\tilde{Q}^{l\to l+1}_{-\theta})\right), \quad (34)$$

which results in the following proposition:

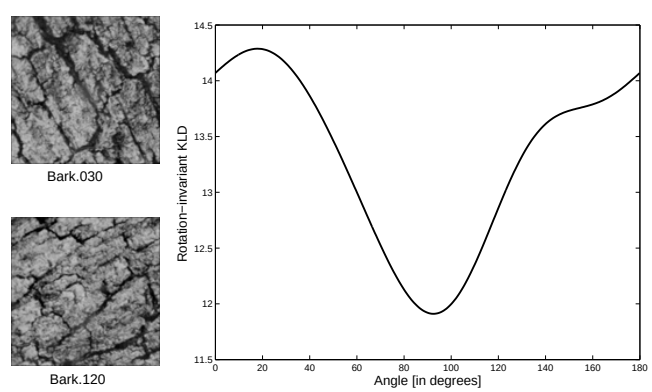
Proposition 2 *Let $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}(I)$ and $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}(\tilde{Q})$ be the signatures corresponding to the normalized coefficients of the steerable pyramids for two given homogeneous textures I and $\tilde{Q}$, respectively. The rotation-invariant KLD between the two textures takes the following form:*

$$\begin{aligned} D &= D(I\|\tilde{Q}) \\ &= \min_{\theta\in\Theta}\frac{1}{2}\Bigg[\sum_{l=1}^{L} tr\big(\mathbf{\Sigma}_I^l\mathbf{F}^T(\theta)(\mathbf{\Sigma}_{\tilde{Q}}^l)^{-1}\mathbf{F}(\theta)\big) + \\ &\quad + \sum_{l=1}^{L-1} tr\big(\mathbf{\Sigma}_I^{l\to l+1}\mathbf{F}^T(\theta)(\mathbf{\Sigma}_{\tilde{Q}}^{l\to l+1})^{-1}\mathbf{F}(\theta)\big)\Bigg] - J(L-\tfrac{1}{2}) - \\ &\quad - \frac{1}{2}\Bigg[\sum_{l=1}^{L}\ln(|\mathbf{\Sigma}_I^l(\mathbf{\Sigma}_{\tilde{Q}}^l)^{-1}|) + \sum_{l=1}^{L-1}\ln(|\mathbf{\Sigma}_I^{l\to l+1}(\mathbf{\Sigma}_{\tilde{Q}}^{l\to l+1})^{-1}|)\Bigg]. \end{aligned} \quad (35)$$

Proof: The proof of Proposition 2 follows by direct computation, after the substitution of (33) in (34). ■

If we employ the signature $\mathcal{S}^{\mathcal{G}}$ instead of $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$, the derivation of the KLD between two distinct textures I and $\tilde{Q}$ is straightforward by implementing the above proposition, omitting the terms which contain the covariance matrices $\mathbf{\Sigma}_I^{l\to l+1}$, $\mathbf{\Sigma}_{\tilde{Q}}^{l\to l+1}$ and replacing $(L-\frac{1}{2})$ with $\frac{L}{2}$.

Notice that when I and $\tilde{Q}$ are two rotated versions of the same image, the angle θ^* for which the minimum is achieved in (35) should be close to the exact relative angle between I and $\tilde{Q}$, that is, the angle one needs to clockwise rotate I in order to get $\tilde{Q}$. Thus, a way to evaluate the performance of the above rotation-invariant KLD, is to verify whether the estimated angle θ^* is actually close to the real relative angle between two physically rotated versions of the same image. Besides, it may also be useful on its own for many practical applications to find out approximately this relative angle. Fig. 5(b) illustrates this by showing the function $D(I\|Q)(\theta)$ given by (35) for the *Bark* texture sample obtained from the Brodatz database.



(a) (b)

Fig. 5. (a) *Bark* physically rotated at 30 and 120 degrees, (b) $D(I\|Q)(\theta)$ for $J=4$. Notice that the minimum is achieved for $\theta^*=90$ degrees, which is the exact relative angle between the two texture samples.

It is also important to note that, in our implementation the steering functions have only odd harmonics, which oscillate at some finite speed. Thus, the number of local minima of (35), as a function of θ, can be at most equal to twice the number of independent harmonics (which happens to be equal to the

number of basic harmonics). In addition, the distance between any two consecutive local minima is lower bounded making it possible to search for them in a few non-overlapping angular intervals [44], which is useful in order to speed up the search for the optimal angle θ^*.

V. Experimental Results

In order to evaluate the efficiency of our overall CBIR system, we apply it on a set of 13, 512×512 texture images obtained from USC SIPI database (cf. Fig. 7). Each of them was physically rotated at 30, 60, 90 and 120 degrees, resulting in a set of 65 texture samples. Then, the texture image dataset is formed by dividing each image into 4, 256×256 non-overlapping subimages constructing a database with a total of $4 \times 65 = 260$ textures. We implemented a 3-level steerable pyramid decomposition with $J = 4$ basic orientations, $\phi_1 = 0$, $\phi_2 = \pi/4$, $\phi_3 = \pi/2$, $\phi_4 = 3\pi/4$, which means that the steering functions are [42]:

$$f_1(\theta) = \frac{1}{2}[\cos(\theta) + \cos(3\theta)] \ , \ f_2(\theta) = f_1\left(\frac{\pi}{4} - \theta\right) \ ,$$

$$f_3(\theta) = f_1\left(\frac{\pi}{2} - \theta\right) \ , \ f_4(\theta) = f_1\left(\frac{3\pi}{4} - \theta\right) \ .$$

The histogram of the estimated characteristic exponent values for the 260 textures is shown in Fig. 6. We observe that only 18% of the textures exhibit Gaussian statistics.

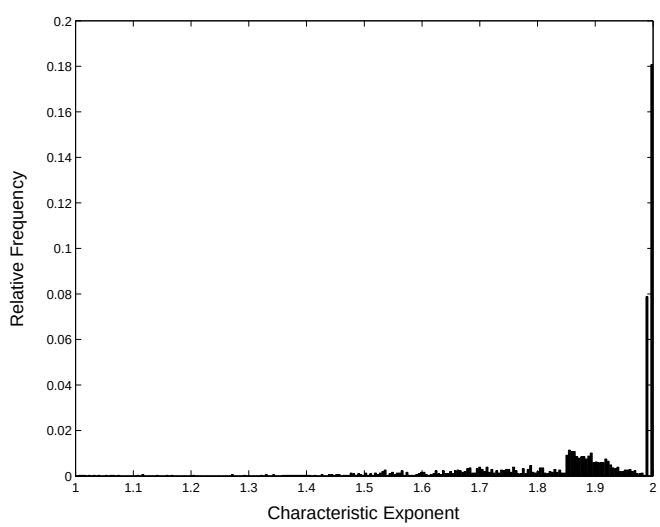


Fig. 6. Histogram of the estimated values for the characteristic exponent, α, for the set of 260 texture images of size 256×256.

In the following illustration, the query is anyone of the non-overlapping 256×256 subimages corresponding to the original, non-rotated set of images. The relevant images for each query are defined as the other 16 subimages corresponding to the rotated versions of the same original image.

We evaluate the performance of the retrieval scheme which employs the signatures $\mathcal{S}^{\mathcal{G}}$ and $\mathcal{S}^{\mathcal{G}}_{\mathcal{E}}$ as the set of extracted features, containing intra- and inter-scale dependencies and the rotation-invariant KLD as the similarity measure. We compare the performance of this retrieval scheme with that

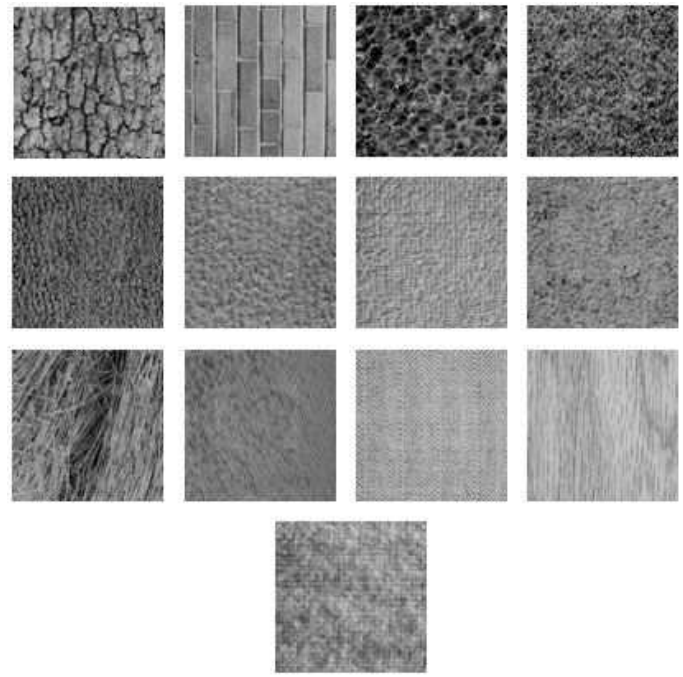


Fig. 7. Texture images from the VisTex database, from left to right and top to bottom: 1) Bark, 2) Brick, 3) Bubbles, 4) Grass, 5) Leather, 6) Pigskin, 7) Raffia, 8) Sand, 9) Straw, 10) Water, 11) Weave, 12) Wood, 13) Wool.

obtained by minimizing the Frobenius norm of the differences between the corresponding covariance matrices ([29]) after the Gaussianization procedure, as well as with the performance obtained by minimizing the corresponding Frobenius norm between sample correlation matrices (Gaussian assumption) without applying the Gaussianization step.

Table V shows the performance in average percentages of retrieving relevant images in the top 16 matches. Comparing

TABLE V
Average retrieval rate (%) in the top 16 matches.

Neighborhood Index	Methods		
	Non-Gaussianized & Frobenius (28)	$\mathcal{S}^{\mathcal{G}}$ & Frobenius (28)	$\mathcal{S}^{\mathcal{G}}$ & KLD (35)
1	88.62	89.26	94.01
Neighborhood Index	Methods		
	Non-Gaussianized & Frobenius (28)	$\mathcal{S}^{\mathcal{G}}_{\mathcal{E}}$ & Frobenius (28)	$\mathcal{S}^{\mathcal{G}}_{\mathcal{E}}$ & KLD (35)
10	88.62	95.40	95.62

the average retrieval rates corresponding to the first two methods of the table, we conclude that the fractional lower-order statistics provide better approximations of the joint statistics between coefficients at adjacent orientations and scales, than the second order moments. Of course, both methods employ the covariance matrices between pairs of subbands, but in the first scheme (Non-Gaussianized & Frobenius) we estimate the sample covariances using the raw subband coefficients without Gaussianization, while in the second scheme ($\mathcal{S}^{\mathcal{G}}$

(or $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$) & Frobenius) the covariances are estimated after the implementation of the Gaussianization procedure, which exploits lower-order moments, since the estimation is based on covariations. The comparison between the retrieval rates of the second and third methods, verifies the fact that a statistical similarity function (KLD) is preferable than a deterministic one (Frobenius norm), for the same set of extracted features.

As shown in Table V, the use of the 10-th neighborhood type combined with the enhanced signature $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$, results in an improved retrieval performance with respect to the other combinations. This is consistent with the above analysis, since the 10-th neighborhood type results in the best Gaussianization performance for most of the 300 constructed subbands and since the enhanced signature contains more texture-specific information than the $\mathcal{S}^{\mathcal{G}}$ signature. We can also observe that by choosing the 10-th neighborhood type, the average retrieval rate using the rotation-invariant Frobenius norm (28) is very close to the rate corresponding to the rotation-invariant KLD. This is due to the improved Gaussianization performance, compared with that corresponding to the first neighborhood type, which results in an increased performance of the Frobenius norm that is best suited for Gaussian distributions. However, the statistical similarity function (KLD) remains superior than the deterministic Frobenius norm.

VI. Conclusions and Future work

In this paper, we studied the design of a rotation-invariant CBIR system based on a multivariate sub-Gaussian ($S\alpha S$) modeling of the coefficients of a steerable wavelet decomposition. We exploited the variance adaptation of the coefficients in small regions at different orientation subbands and levels, by applying a Gaussianization procedure. Then, the FE step consists of simply estimating second-order moments between orientation subbands at the same and at adjacent levels. This process takes also into account the actual heavy-tailed behavior of the coefficients, represented by the fractional lower-order statistics (covariations) between pairs of subbands. We achieve rotation invariance by constructing an appropriate rotation-invariant version of the KLD between zero-mean multivariate Gaussian densities. The experimental results showed an increased average retrieval performance in comparison with the performance of previous methods based on second-order statistics estimated directly from the original subband coefficients, without implementing the Gaussianization. We also conclude that a statistical similarity function, such as KLD, is preferable than the deterministic Frobenius norm.

Future research directions, which could further result in an improved retrieval system with decreased probability of retrieval error, are the following: first of all, the main assumption throughout the present work was the stationary behavior of texture content. That is, we assumed that the distribution of the subband coefficients, which is closely related with the texture-specific information, is invariable within each subband. Instead, we could consider a *non-stationary* approach by permitting a *locally adapted* distribution, that is, by spatially adapting the characteristic exponent and the dispersion parameters. This can also be used in segmentation applications, since the different "objects" contained in a picture can be viewed as local intensity variations.

Regarding the task of similarity measurement between two distinct images, we weighted the contribution of inter-level dependencies in the same way as the intra-level ones, resulting in an overall similarity function that is written as a sum of partial distances. We could further improve the power of the similarity measure by considering some kind of chain rule for the KLD between two images [5].

In section IV, we evaluated the performance of a Gaussianization procedure applied on the subband coefficients of a steerable pyramid. For this purpose, we preserved the same neighborhood type across subbands, for every image in the database. Obviously, it is computationally unfeasible to check the Gaussianization performance for all possible neighborhood formations. However, we expect an improved Gaussianization performance and consequently better retrieval performance, by performing some reduced complexity adaptation of the optimal neighborhood type across subbands and images.

References

[1] J. Ashley, M. Flickner, J. Hafner, D. Lee, W. Niblack, and D. Petkovic, "The query by image content (qbic) system," in *Proceedings of the 1995 ACM SIGMOD Interl. Conf. on Management of data*, (New York, U.S.), pp. 475–, ACM Press, 1995.

[2] S. Belongie, J. Malik, and J. Puzicha, "Matching shapes," *In International Conference on Computer Vision, Vancouver, Canada*, 2001.

[3] B. Manjunath, J. R. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits and Syst. Video Tech.*, vol. 11, pp. 703–715, June 2001.

[4] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms," *Journal of Optical Soc. Am.*, vol. 7, no. 5.

[5] S. Kullback, *Information Theory and Statistics*. Dover, 1997.

[6] J. R. Bergen and E. H. Adelson, "Theories of visual texture perception," in *Spatial Vision* (D. R. Ed., ed.), CRC press, 1991.

[7] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[8] P. Wu, B. S. Manjunath, S. Newsam, and H. D. Shin, "A texture descriptor for browsing and similarity retrieval," *Signal Processing: Image Communication*, vol. 16, pp. 33–43, 2000.

[9] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. on Image Processing*, vol. 4, pp. 1549–1560, November 1995.

[10] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, 1999.

[11] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–692, July 1989.

[12] E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *31st Asilomar Conf on Signals, Systems and Computers*, (Pacific Grove, CA), pp. 673–678, IEEE Computer Society, 1997.

[13] S. G. Chang, B. Yu, and M. Vetterli, "Lossy compression and wavelet thresholding for image denoising," *submitted to IEEE Trans. Image Processing*.

[14] S. Liapis and G. Tziritas, "Color and texture image retrieval using chromaticity histograms and wavelet frames," *IEEE Trans. on Multimedia*, vol. 6, pp. 676–686, Oct. 2004.

[15] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Processing*, vol. 11, pp. 146–158, Feb. 2002.

[16] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.*, vol. 20, pp. 772–783, Aug. 2001.

[17] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Trans. Geosc. and Rem. Sens.*, vol. 41, pp. 1773–1784, Aug. 2003.

[18] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. New York: Chapman and Hall, 1994.

[19] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. New York: John Wiley and Sons, 1995.

[20] G. Tzagkarakis and P. Tsakalides, "A statistical approach to texture image retrieval via alpha-stable modeling of wavelet decompositions," *5th International Workshop on Image Analysis for Multimedia Interactive Services, Lisboa, Portugal*, April 21-23 2004.

[21] J. Huang, "Study on the correlation properties of wavelet transform coefficients and the applications in a neural network-based hybrid image coding system," *CISST'03, Las Vegas, USA*, June 22-26 2003.

[22] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, pp. 49–71, Dec. 2000.

[23] G. Cross and A. Jain, "Markov random field texture models," *IEEE Trans Pattern Analysis and Machine Intelligence*, vol. 5, pp. 25–39, 1983.

[24] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans. on Signal Processing (Special issue on Wavelets and Filterbanks)*, pp. 886–902, April 1998.

[25] M. N. Do and M. Vetterli, "Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models," *Submitted to IEEE Trans. on Multimedia*, April 2001.

[26] H. Greenspan, S. Belongie, R. Goodman, and P. Perona, "Rotation invariant texture recognition using a steerable pyramid," *in Proc. Int. Conf. on Pattern Recognition*, pp. 162–167, 1994.

[27] G. M. Haley and B. S. Manjunath, "Rotation invariant texture classification using modified Gabor filters," *in Proc. IEEE Int. Conf. Image Processing, Washington, DC*, Oct 1995.

[28] G. M. Haley and B. S. Manjunath, "Rotation-invariant texture classification using a complete space-frequency model," *IEEE Trans. Image Processing*, vol. 8, pp. 255–269, Feb 1999.

[29] B. Beferull-Lozano, H. Xie, and A. Ortega, "Rotation-invariant features based on steerable transforms with an application to distributed image classification," in *IEEE International Conference on Image Processing, Barcelona, Spain*, 2003.

[30] J. Mao and A. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern Recognition*, vol. 25, pp. 173–188, Feb. 1992.

[31] F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 18, pp. 722–733, July 1996.

[32] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.

[33] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley, 1991.

[34] J. P. Nolan, "Parameterizations and modes of stable distributions," *Statistics & Probability Letters*, no. 38, pp. 187–195, 1998.

[35] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," *Commun. Statist.-Stochastic Models*, vol. 13, pp. 759–774, 1997.

[36] S. Cambanis and G. Miller, "Linear problems in pth order and stable processes," *SIAM J. Appl. Math.*, vol. 41, pp. 43–69, 1981.

[37] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE*, vol. 81, pp. 986–1010, 1993.

[38] E. P. Simoncelli and J. Portilla, "Texture characterization via joint statistics of wavelet coefficient magnitudes," *in Proc. of IEEE Int. Conf. on Image Processing*, 1998.

[39] J. Chambers, W. Cleveland, B. Kleiner, and P. Tukey, *Graphical Methods for Data Analysis*. Wadsworth, 1983.

[40] C. K. Chui, *An Introduction to Wavelets*. Academic Press, 1992.

[41] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, 1992.

[42] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, 1991.

[43] D. Andrews and C. Mallows, "Scale mixtures of normal distributions," *J. Royal Stat. Soc.*, vol. 36, pp. 99–, 1974.

[44] B. Beferull-Lozano, "Quantization design for structured overcomplete expansions," *Ph. D. thesis, University of Southern California*, 2002.