Rotation-Invariant Texture Retrieval via Signature Alignment Based on Steerable Sub-Gaussian Modeling

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

***Abstract*—This paper addresses the construction of a novel efficient rotation-invariant texture retrieval method that is based on the alignment in angle of signatures obtained via a steerable sub-Gaussian model. In our proposed scheme, we first construct a steerable multivariate sub-Gaussian model, where the fractional lower-order moments of a given image are associated with those of its rotated versions. The feature extraction step consists of estimating the so-called covariations between the orientation subbands of the corresponding steerable pyramid at the same or at adjacent decomposition levels and building an appropriate signature that can be rotated directly without the need of rotating the image and recalculating the signature. The similarity measurement between two images is performed using a matrix-based norm that includes a signature alignment in angle between the images being compared, achieving in this way the desired rotation-invariance property. Our experimental results show how this retrieval scheme achieves a lower average retrieval error, as compared to previously proposed methods having a similar computational complexity, while at the same time being competitive with the best currently known state-of-the-art retrieval system. In conclusion, our retrieval method provides the best compromise between complexity and average retrieval performance.**

***Index Terms*—Fractional lower-order moments (FLOM), rotation-invariant texture retrieval, steerable multivariate sub-Gaussian model.**

I. Introduction

INCREASING accumulation of visual information in large digital databases has been prominent during the last decades. To improve the management of these collections, it is necessary to have effective and precise methods to search and interact with them. For this purpose, content-based image retrieval (CBIR) from unannotated image databases has gained an important interest in the research community. CBIR is a set of techniques for retrieving relevant images on the basis of automatically derived features, which accurately specify the information content of a given image.

There exist several practical examples where a CBIR plays a major role. Two typical examples are: a) retrieval of images (through the Internet) that are similar to a given query image and b) remote satellite sensing of images where only novel images have to be transmitted and where the novelty is checked on the earth by performing a comparison between the features of a query and the features of images contained in a database.

In a typical CBIR system 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 an image transformation, to accurately represent the content of a given image. In order to guarantee the computational efficiency of the retrieval process,[1] this set has to be much smaller in size than the original image, while capturing as much as possible the important information from the image, for the purpose of classification. During the SM step, a distance-like function is usually employed, which measures how close a database image is to a given query image, by measuring the distance between their corresponding 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 the description of image content. Loosely speaking, the class of *texture images* includes images that are spatially homogeneous and consist of repeated elements (*image texels*), often subject to some randomization in their location, size, color, and orientation. However, it is also important to mention that even though this assumption of homogeneity is done, our experimental results show that our method also tolerates certain degree of nonhomogeneity.

The design of retrieval systems in a transform-domain is justified by the observation that often a linear, invertible transform results in a set of coefficients whose structure is simpler to model than the structure of the original image in the pixel-domain. Features such as oriented edges, ridges, and lines are prominent in real-world images. For such images, the 2-D wavelet transform is a powerful modeling tool, providing a natural arrangement of the wavelet transform coefficients into multiple scales and oriented subbands representing the edges in different directions [4]. In previous studies, the texture information has been modeled using the first or second-order statistics of the coefficients obtained via a Gabor wavelet transform [5], or an overcomplete

Manuscript received November 29, 2006; revised March 18, 2006. This work was supported in part by the Greek General Secretariat for Research and Technology under Program ΠΕΝΕΔ, Code 03ΕΔ69, and in part by the Marie Curie TOK-DEV "ASPIRE" Grant (MTKD-CT-2005-029791) within the 6th European Community Framework Program. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Eli Saber.

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

B. Beferull-Lozano is with the Group of Information and Communication Systems, Instituto de Robótica-Escuela Técnica Superior de Ingeniería, Universidad de Valencia, Valencia, Spain (e-mail: baltasar.beferull@uv.es).

Digital Object Identifier 10.1109/TIP.2008.924390

[1] In other cases, where the encoder and decoder are physically separated, a reduction in the necessary transmission bandwidth is also a very important requirement.

wavelet decomposition constituting a tight frame [6]. Moreover, until recently, the wavelet coefficients have been modeled either as independent Gaussian variables, or as jointly Gaussian vectors [7].

In the texture retrieval scheme proposed in this work, the task of FE is considered in a statistical framework. In particular, the signature of a given image contains the estimations of *fractional lower-order moments* (FLOMs) between the orientation subbands at the same or at adjacent decomposition levels. This is justified by the observation that the wavelet transforms of real-world images tend to be sparse, that is, they result in a large number of small amplitude coefficients and a small number of large amplitude coefficients [8]. This property gives rise to peaky and heavy-tailed *non-Gaussian* marginal distributions of the wavelet subband coefficients, whose statistics are best described by lower than second-order moments [8], [9], as opposed to the traditional Gaussian modeling. Then, the SM step employs an appropriate norm-based distance function that exploits the heavy-tailed behavior of the distributions of the transform coefficients.

In a recent work [10], we have shown that an improved description of the texture information can be achieved if we take into consideration the actual heavy-tailed behavior of the distribution of transform coefficients, as well as their interdependencies across the subbands at the same or at adjacent decomposition levels. Our formulation improved the retrieval performance, resulting in a decreased average retrieval error for images with distinct non-Gaussian statistics, compared with the previously used GGD model. Our study in [10] motivated the use of a joint multivariate $S\alpha S$ distribution model in this work.

On the other hand, the theory of Markov random fields has enabled a new generation of statistical texture models, taking into account possible interdependencies between the transform coefficients, in which the full model is characterized by statistical interactions within local neighborhoods [11]. Similarly, a new framework for statistical signal processing based on wavelet-domain Hidden Markov Models was proposed in [12]. 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 exploit transform-domain interdependencies by tying up the subband coefficients at adjacent orientations and/or scales in vectors, which are then considered to be drawn from a multivariate *sub-Gaussian* distribution. Within the sub-Gaussian framework, we use the notion of *covariation*, that is, fractional lower-order statistics, instead of the second-order covariance, in order to describe the 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.

An important desirable property of a CBIR system is rotation invariance. Recently, a rotation-invariant texture retrieval system based on steerable pyramids was proposed by Beferull-Lozano *et al.* [13], where the correlation matrices between several basic orientation subbands at each level of a wavelet pyramid are chosen as the energy-based texture features. Do and Vetterli [14] derived a steerable rotation-invariant statistical model for texture description, by enhancing a technique based on a wavelet-domain hidden Markov model (HMM) [12]. In our recent work [10], we designed an efficient rotation-invariant texture retrieval system by applying a joint sub-Gaussian model on a steerable pyramid decomposition [15] of the images, followed by a Gaussianization process, resulting in a set of (modified) pyramid subbands whose coefficients are almost jointly Gaussian. Then, the SM was carried out by constructing a rotation-invariant version of the Kullback–Leibler divergence (KLD) between multivariate Gaussian density functions, describing statistically the Gaussianized subbands.

Although the method described in [10] results in an increased retrieval performance, compared with the previously developed methods based on second-order statistics, a main drawback is its computational complexity due to the Gaussianization procedure, since it involves the use of a sufficiently large neighborhood for each transform coefficient and repeating this procedure across all the pyramid. In this work, we propose a rotation-invariant texture retrieval system with reduced computational complexity, by constructing directly a *steerable multivariate sub-Gaussian* model and by applying a rotation-invariant deterministic similarity function based on matrix norms, similar to the one described in [13]. Our proposed method provides an average retrieval performance which is superior to the current state-of-the-art methods with similar complexity (e.g., [13] and [14]), while being competitive with the current best method [10], which has a much higher computational cost. Following the retrieval method presented in this paper, we avoid the computational burden introduced by the Gaussianization procedure during the FE step, while maintaining a very similar average retrieval error to the one in [10].

The rest of the paper is organized as follows. In Section II, we develop a rotation-invariant texture retrieval system based on a steerable multivariate sub-Gaussian model, which exploits the non-Gaussian behavior of the marginal and joint distributions of the subband coefficients obtained via a steerable pyramid decomposition. In Section III, we apply our scheme to a set of real-world textures in order to evaluate its retrieval performance and we compare it with the performance of other recently introduced texture retrieval techniques. Finally, we draw several conclusions and explore avenues of future research in Section IV.

II. Rotation-Invariance via a Steerable Multivariate Sub-Gaussian Model

A desirable property of any texture retrieval system is rotation invariance. Consider, for instance, the case of a content-based search over the Internet, where we are interested in retrieving images containing the same object with the one of a given query and the orientation of this object can vary among the images, or the case of remote satellite sensing, where there could be different rotated versions of the same area of interest. Many approaches for texture retrieval make use of transforms which suffer from rotation and translation variance. The standard wavelet transform, used in the previously developed retrieval methods, belongs in this class of transforms, that is, it lacks the properties of translation and rotation invariance. This results in a mismatch of the retrieval process when the image orientation varies. In fact, the wavelet coefficients of a rotated version 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 [15], [16], which is a linear, multiscale, multiorientation 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 texture retrieval technique that performs an angular alignment between texture images, thus achieving (indirectly) the rotation-invariance property. This technique is based on the joint sub-Gaussian modeling of several coefficients obtained from a steerable pyramid, defining features that incorporate dependence across orientations and scales.[2] In particular, we construct a steerable model, relating the fractional lower-order statistics of a rotated image with those of its original version and then we derive a rotation-invariant similarity function. The development of a steerable model implies extracting features which are "steerable," that is, given the features of an image, 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]

The use of the family of multivariate sub-Gaussian distributions, as an accurate tool for modeling the heavy-tailed behavior of the steerable pyramid coefficients, has been justified in [10], showing that it provides a statistical description that is more complete to both Gaussian and generalized Gaussian distributions, which have been used in previous texture retrieval techniques [14]. On the one hand, the oriented subbands of a steerable pyramid obtained from a texture image exhibit various degrees of non-Gaussianity both marginally and jointly. The joint modeling of coefficients at adjacent orientations and scales via a multivariate sub-Gaussian model is motivated by several previous studies on their correlation properties [17] and the fact that the components of a joint sub-Gaussian random vector, corresponding to these coefficients, are by definition highly dependent [18].

A. Sub-Gaussian Random Vectors

Before providing the definition of a sub-Gaussian random vector, we introduce first briefly the family of α-Stable distributions. The α-Stable distribution is best defined by its characteristic function [19]

$$\varphi(t)=\begin{cases}\exp\left(\imath\delta t-\gamma^{\alpha}|t|^{\alpha}\left(1-\imath\beta(\operatorname{sign}t)\tan\frac{\pi\alpha}{2}\right)\right), & \alpha\neq 1\\ \exp\left(\imath\delta t-\gamma|t|\left(1+\imath\beta\frac{2}{\pi}(\operatorname{sign}t)\ln|t|\right)\right), & \alpha=1\end{cases} \tag{1}$$

where $\operatorname{sign} t=\{1 \text{ if } t>0,\ 0 \text{ if } t=0,\ -1 \text{ if } t<0\}$. The α-Stable distribution is characterized by four parameters: α is the *characteristic exponent*, taking values $0<\alpha\leq 2$, $\delta(-\infty<\delta<\infty)$ is the *location parameter*, $\gamma(\gamma>0)$ is the *dispersion* of the distribution and $\beta(-1\leq\beta\leq 1)$ is the index of *skewness*. 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 α-Stable density function. The dispersion parameter determines the spread of the distribution around its location parameter, similar to the variance of the Gaussian. We will denote α-Stable distributions by $S_{\alpha}(\gamma,\beta,\delta)$ and write $X\sim S_{\alpha}(\gamma,\beta,\delta)$ to indicate that X is a random variable that follows an α-Stable distribution with parameters $(\alpha,\gamma,\beta,\delta)$.

A random variable X is said to follow a *Symmetric alpha-Stable* ($S\alpha S$) distribution if and only if $\beta=\delta=0$. A $S\alpha S$ distribution is called *standard* if $\gamma=1$ and $\delta=0$. A $S\alpha S$ distribution is best defined by its characteristic function given by the following expression:

$$\varphi(t)=\exp\left(-\gamma^{\alpha}|t|^{\alpha}\right). \tag{2}$$

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 exponentially decaying 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.

According to the following proposition [18], every $S\alpha S$ random variable is conditionally Gaussian.

Proposition 1 [18]: Let $G\sim S_{\alpha'}(\gamma,0,0)$ with $0<\alpha'\leq 2$ and let $0<\alpha<\alpha'$. Let A be an α/α'-stable random variable with characteristic function in Laplace form $\mathbb{E}[\exp(-kA)]=\exp(-k^{\alpha/\alpha'})$, $k>0$, i.e., $A\sim S_{\alpha/\alpha'}((\cos(\pi\alpha/2\alpha'))^{\alpha'/\alpha},1,0)$, and assume G and A to be independent. Then $X=A^{1/\alpha'}G\sim S_{\alpha}(\gamma,0,0)$, i.e., the random variable X follows a $S\alpha S$ distribution.

Proof: In order to prove that the random variable X follows a $S\alpha S$ distribution, we have to show that its characteristic function has the form of (2). For any real t, we have that the characteristic function of X is given by

$$\begin{aligned}\mathbb{E}\left[\exp(\imath tX)\right]&=\mathbb{E}\left[\exp\left(\imath tA^{1/\alpha'}G\right)\right]\\&=\mathbb{E}\left[\mathbb{E}\left[\exp\left(\imath tA^{1/\alpha'}G\right)\right]|A\right]\\&\overset{(*)}{=}\mathbb{E}\left[\exp\left(-\gamma^{\alpha'}|t|^{\alpha'}A\right)\right]\\&\overset{(**)}{=}\exp\left(-\left(\gamma^{\alpha'}|t|^{\alpha'}\right)^{\alpha/\alpha'}\right)\\&=\exp\left(-\gamma^{\alpha}|t|^{\alpha}\right)\end{aligned}$$

that is, the characteristic function of X is written in the form of (2), which means that X follows a $S\alpha S$ distribution with characteristic exponent α and dispersion parameter γ. The identity $(*)$ follows from the fact that $G\sim S_{\alpha'}(\gamma,0,0)$, and, thus, its characteristic function is written in the form of (2), while identity $(**)$ follows directly from the assumption about the characteristic function in Laplace form for A, by setting $k=\gamma^{\alpha'}|t|^{\alpha'}$. ■

[2] In the following, we will often refer to the "orientation" and "scale" with their equivalent terms "subband" and "level," respectively.

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

In particular, this implies that if G is a zero-mean Gaussian random variable (i.e., $G \sim S_2(\gamma, 0, 0)$) and if A is a positive $\alpha/2$-stable random variable independent of G, then $X = A^{1/2}G \sim S_\alpha(\gamma, 0, 0)$, i.e., X is a $S\alpha S$ random variable. Since $S_2(\gamma, 0, 0) = \mathcal{N}(0, 2\gamma^2)(= \mathcal{N}(\mathbb{E}[G], Var[G]))$ [18],[4] then any $S\alpha S$ random variable $X \sim S_\alpha(\gamma, 0, 0)$ can be generated as the product of an $\alpha/2$-stable random variable A, given by Proposition 1 for $\alpha' = 2$, and a zero-mean Gaussian random variable G with variance equal to $2\gamma^2$.

As mentioned above, in our method we are interested in extracting possible interdependencies between pyramid subband coefficients at the same or at adjacent decomposition level. For this purpose, we construct a joint (multivariate) statistical model for the different subbands. A natural way to design this model is to use the result of Proposition 1, that is, that every $S\alpha S$ random variable is conditionally Gaussian. Thus, extending the univariate $S\alpha S$ model (used in our previous work [20]) to a joint (multivariate) model with $S\alpha S$ components (viewed as conditionally Gaussian random variables), leads to the use of the so-called *sub-Gaussian $S\alpha S$ random vector*, defined as follows[5] [18].

Definition 1: A vector $\vec{X}$ is called a sub-Gaussian $S\alpha S$ random vector (in $\mathbb{R}^n$) with underlying Gaussian vector $\vec{G}$ iff it can be written in the form $\vec{X} = A^{1/2}\vec{G}$, where A is a positive $\alpha/2$-stable random variable with parameters $A \sim S_{\alpha/2}((\cos(\pi\alpha/4))^{2/\alpha}, 1, 0)$ and $\vec{G} = (G_1, G_2, \ldots, G_n)$ is a zero-mean Gaussian random vector, independent of A, with covariance matrix $\boldsymbol{R}$.

A multivariate sub-Gaussian distribution, with underlying covariance matrix $\mathbf{R}$, is often denoted by $\alpha - \mathrm{SG}(\mathbf{R})$, where the parameter α is the characteristic exponent, controlling the heaviness of the tails of the marginal sub-Gaussian distributions. According to Proposition 1, the marginal distributions of the components of a sub-Gaussian vector belong to the family of $S\alpha S$ distributions, namely, the ith component of a sub-Gaussian vector $\vec{X}$, $X_i \sim S_\alpha(\sqrt{Var(G_i)/2}, 0, 0)$ (where $Var(G_i)$ is the variance of the ith component of the underlying Gaussian vector $\vec{G}$). As it is described later in this section, following our approach, the subband coefficients at adjacent orientations and/or scales, are tied up in vectors which are assumed to be samples of an $\alpha - \mathrm{SG}(\mathbf{R})$ distribution. This assumption is also justified in [10].

B. Estimation of Covariation

Since second-order moments do not exist for the family of sub-Gaussian random variables, a quantity called *covariation*, which plays an analogous role for sub-Gaussian random variables to the one played by covariance in Gaussian random variables, has been proposed [18]. Let X and Y be joint sub-Gaussian random variables (representing the coefficients belonging to two different subbands for instance) with $\alpha_X = \alpha_Y = \alpha (1 < \alpha \leq 2)$, zero location parameters and dispersions γ_X and γ_Y ($\gamma_X \neq \gamma_Y$ in general), respectively. Then, for all $1 < p < \alpha$, the covariation of X with Y is given by

$$[X, Y]_\alpha = \frac{E\left\{XY^{\langle p-1\rangle}\right\}}{E\left\{|Y|^p\right\}}\gamma_Y^\alpha. \tag{3}$$

In the above expression, we use the notation $z^{\langle a\rangle} = \{z^a \text{ for } z > 0, 0 \text{ for } z = 0, -(-z)^a \text{ for } z < 0\}$, for any real number z and $a \geq 0$. The *covariation coefficient* of X with Y is defined by

$$\lambda_{XY} = \frac{[X, Y]_\alpha}{[Y, Y]_\alpha} = \frac{E\left\{XY^{\langle p-1\rangle}\right\}}{E\left\{|Y|^p\right\}}. \tag{4}$$

From (4), we observe that we can find an estimation of $[X, Y]_\alpha$ by multiplying an estimated value of λ_{XY} and an estimated value of $[Y, Y]_\alpha$. Since $[Y, Y]_\alpha = \gamma_Y^\alpha$, an estimation of $[Y, Y]_\alpha$ can be obtained via the Maximum Likelihood (ML) estimation of the characteristic exponent α and the dispersion parameter γ_Y. In the following, we estimate the model parameters (α, γ) using the consistent ML Nolan's method described in [21], which gives reliable estimates (where the reliability can be evaluated through simulations) and provides the tightest possible confidence intervals. On the other hand, the value of λ_{XY} can be estimated directly from (4) by replacing the expectations with sample means, yielding the FLOM estimator [22]

$$\hat{\lambda}_{\mathrm{FLOM}} = \frac{\sum_{i=1}^n X_i |Y_i|^{p-1} sign(Y_i)}{\sum_{i=1}^n |Y_i|^p} \tag{5}$$

where (X_i, Y_i), $i = 1, \ldots, n$, is a set of *independent* observations. Our choice of the FLOM estimator is based on its robustness against changes of α and also on the fact that it is very simple and inexpensive to compute.

Let the vectors $\{\vec{X^1}, \vec{X^2}, \ldots, \vec{X^N}\}$ constitute a set of N independent realizations of an $\alpha - \mathrm{SG}(\mathbf{R})$ distribution, where $\vec{X^k} = (X_1^k, X_2^k, \ldots, X_n^k)$, $k = 1, \ldots, N$. The FLOM covariation estimator between the ith and jth components of the sub-Gaussian vector $\vec{X}$ is defined by

$$\hat{\mathbf{c}}_{ij}^{\mathrm{FLOM}} = \frac{\sum_{k=1}^N X_i^k \left|X_j^k\right|^{p-1} \mathrm{sign}\left(X_j^k\right)}{\sum_{k=1}^N \left|X_j^k\right|^p}\gamma_{X_j}^\alpha \tag{6}$$

where the characteristic exponent α and dispersion parameter γ_{X_j} are estimated from the set $\{X_j^k\}_{k=1,\ldots,N}$ using Nolan's ML estimator. We define the estimated *covariation matrix* $\hat{\mathbf{C}}$ to be the matrix with elements $[\hat{\mathbf{C}}]_{ij} = \hat{\mathbf{c}}_{ij}^{\mathrm{FLOM}}$. We can then estimate the elements $[\mathbf{R}]_{ij}$ of the underlying covariance matrix, $\mathbf{R}$, using the following estimators that we provide in [10]:

$$[\hat{\mathbf{R}}]_{jj} = \left(2^{\frac{\alpha}{2}}[\hat{\mathbf{C}}]_{jj}\right)^{\frac{2}{\alpha}}, \quad [\hat{\mathbf{R}}]_{ij} = 2^{\frac{\alpha}{2}}\frac{[\hat{\mathbf{C}}]_{ij}}{[\hat{\mathbf{R}}]_{jj}^{\frac{(\alpha-2)}{2}}} \tag{7}$$

which are both consistent and asymptotically normal.[6]

[4] $G \sim \mathcal{N}(\mu, \sigma^2)$ denotes that G is a Gaussian random variable with mean μ and variance σ^2.

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

[6] An estimator $\hat{A}(n)$ (where n is the sample size) is a consistent estimator for a parameter A if: i) $\hat{A}$ is asymptotically unbiased (i.e., $\lim_{n\to\infty} \mathbb{E}[\hat{A}(n)] = A$) and ii) its asymptotic variance goes to 0 (i.e., $\lim_{n\to\infty} Var[\hat{A}(n)] = 0$). An estimator $\hat{A}(n)$ is asymptotically normal if its asymptotic distribution is a Gaussian with mean $\mathbb{E}[\hat{A}(\infty)]$ and variance $Var[\hat{A}(\infty)]$ (i.e., $\lim_{n\to\infty} \hat{A}(n) \sim \mathcal{N}(\mathbb{E}[\hat{A}(\infty)], Var[\hat{A}(\infty)])$).

Notice that the estimation of covariations and consequently the estimation of the covariation matrices, requires the specification of the parameter p. In [10], we describe a procedure to compute the optimal p as a function of the (estimated) characteristic exponent α by finding the value of p that minimizes the standard deviation of the covariation estimator (6). For this purpose, we studied the influence of the parameter p on the performance of this estimator via Monte-Carlo simulations, for different values of $\alpha > 1$. As a general conclusion, the simulations showed that the FLOM covariation estimator approximates accurately the true covariation value and the corresponding value of the optimal p increases as the characteristic exponent α increases. For instance, when $\alpha = 1.1, 1.2, 1.3, 1.4, 1.5$, the simulations gave the optimal p values to be equal to 0.56, 0.57, 0.59, 0.66, and 0.68, respectively.

C. Design of the Steerable Sub-Gaussian Model

A steerable pyramid provides a representation that consists of L decomposition levels, where in each level there are J orientation subbands; thus, in total, there are $L \times J$ subbands. In the following, assume for convenience that the coefficients of an lth-level subband, with dimension $N \times N$, are arranged in a $1 \times N^2$ row vector by grouping them row-wise. Let $\mathbf{\Sigma}^l$ and $\mathbf{\Sigma}^l_\theta$ 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 at an angle θ, I_θ, respectively. That is, if we denote by $c^l(x_k, \phi_i), c^l(x_k, \phi_j)$, the coefficients of the subbands corresponding to the orientations $\phi_i, \phi_j (i, j = 1, \ldots, J)$, placed at the spatial location $x_k (k = 1, \ldots, N^2)$ and level $l = 1, \ldots, L$, then, $[\mathbf{\Sigma}^l]_{ij} = (1/N^2) \sum_{k=1}^{N^2} c^l(x_k, \phi_i) c^l(x_k, \phi_j)$. The elements of $\mathbf{\Sigma}^l_\theta$ are defined similarly.

The following proposition states that $\mathbf{\Sigma}^l$ and $\mathbf{\Sigma}^l_\theta$ are equivalent matrices.

Proposition 2 [10], [13]: The relationship between the sampled correlation matrices $\mathbf{\Sigma}^l$ and $\mathbf{\Sigma}^l_\theta$ is given by

$$\mathbf{\Sigma}^l_\theta = \mathbf{F}(\theta) \mathbf{\Sigma}^l \mathbf{F}^T(\theta) \tag{8}$$

where

$$\mathbf{F}(\theta) = \begin{bmatrix} f_1(\phi_1 - \theta) & f_2(\phi_1 - \theta) & \cdots & f_J(\phi_1 - \theta) \\ f_1(\phi_2 - \theta) & f_2(\phi_2 - \theta) & \cdots & f_J(\phi_2 - \theta) \\ \vdots & \vdots & \vdots & \vdots \\ f_1(\phi_J - \theta) & f_2(\phi_J - \theta) & \cdots & f_J(\phi_J - \theta) \end{bmatrix} \tag{9}$$

with $\{f_1(\cdot), f_2(\cdot), \ldots, f_J(\cdot)\}$ being a set of J steering functions. In our work, the J basic angles $\{\phi_i\}_{i=1,\ldots,J}$ are taken to be equispaced, which makes $\mathbf{F}(\theta)$ an orthogonal matrix for any angle θ, i.e., $\mathbf{F}^T(\theta) = \mathbf{F}^{-1}(\theta) (= \mathbf{F}(-\theta))$, and, thus, in this particular case, $\mathbf{\Sigma}^l$ and $\mathbf{\Sigma}^l_\theta$ become orthogonally equivalent.[7] It can be easily shown that the above proposition also holds if we consider the sample covariance matrices between the J subbands of a given level l_1 and the J subbands of another level l_2.

[7]Given a matrix $\mathbf{A}$ which depends on an angle θ, we use the notations $\mathbf{A}(\theta)$ and $\mathbf{A}_\theta$ interchangeably. Following this convention, we always have that $\mathbf{A} = \mathbf{A}(\theta = 0) = \mathbf{A}_{\theta=0}$.

Let $\vec{X}^l_k = [c^l(x_k, \phi_1), c^l(x_k, \phi_2), \ldots, c^l(x_k, \phi_J)]^T$ denote the vector containing the J basic coefficients at the spatial location $x_k (k = 1, \ldots, N^2)$ and level $l = 1, \ldots, L$. Under a joint sub-Gaussian assumption (this has been justified previously in [10], where it has been found that it gives the best statistical fitting of the marginal and joint distributions of the pyramid subband coefficients), the vectors $\vec{X}^l_k$ are modeled as joint sub-Gaussian vectors $\alpha - \mathrm{SG}(\mathbf{R}^l)$, with $\mathbf{R}^l$ denoting the underlying covariance matrix corresponding to the subbands at the lth level. Under this assumption, the pyramid coefficients at each subband follow a sub-Gaussian marginal distribution, and, thus, the lth-level coefficient at spatial location x_k and at any angle θ can be expressed as

$$\begin{aligned} c^l(x_k, \theta) &= \sum_{i=1}^{J} f_i(\theta) c^l(x_k, \phi_i) = \sum_{i=1}^{J} f_i(\theta) \left(\sqrt{A} c^l_G(x_k, \phi_i) \right) \\ &= \sqrt{A} \sum_{i=1}^{J} f_i(\theta) c^l_G(x_k, \phi_i) = \sqrt{A} c^l_G(x_k \theta) \end{aligned} \tag{10}$$

which means that the pyramid subband coefficients of a rotated image at an angle θ, are also sub-Gaussian random variables with the same characteristic exponent as that of the corresponding subbands of the original (nonrotated) image, and having a Gaussian part which is the rotated version of the original Gaussian part at the same angle θ. This is important because it motivates clearly the construction of a steerable statistical model.

Let $\mathbf{C}^l$ and $\mathbf{C}^l_\theta$ denote the covariation matrices with elements that are the estimated covariations between pairs of orientation subbands at the lth decomposition level of the original image I and its rotated version I_θ, respectively. For instance, the element $[\mathbf{C}^l]_{ij}$ (that is, the estimated covariation between the lth-level subbands corresponding to the orientations ϕ_i, ϕ_j) is obtained from (6) by replacing X^k_i, X^k_j with $c^l(x_k, \phi_i)$, $c^l(x_k, \phi_j)$, respectively, for $k = 1, \ldots, N^2$. The elements of $\mathbf{C}^l_\theta$ are computed similarly from the subband coefficients of the rotated image I_θ.

Notice that the relation between the covariation matrix $\mathbf{C}^l$ and the associated underlying covariance matrix $\mathbf{R}^l$ is given by (7). This relation is also valid between the elements of the covariation matrix $\mathbf{C}^l_\theta$ of the rotated image and the elements of the corresponding underlying covariance matrix $\mathbf{R}^l_\theta$. Thus, by inverting (7) and expressing the elements of $\mathbf{C}^l_\theta$ as a function of the elements of $\mathbf{R}^l_\theta$, we have

$$\left[\mathbf{C}^l_\theta\right]_{ij} = 2^{-\frac{\alpha}{2}} \left[\mathbf{R}^l_\theta\right]_{ij} \left[\mathbf{R}^l_\theta\right]^{(\alpha-2)/2}_{jj}. \tag{11}$$

Taking into account that Proposition 2 holds for each lth-level, the replacement of $\mathbf{\Sigma}^l$ and $\mathbf{\Sigma}^l_\theta$ with $\mathbf{R}^l$ and $\mathbf{R}^l_\theta$, respectively in (8), and its combination with (11), yields the following proposition.

Proposition 3: Suppose that $\mathcal{S} = \{\mathbf{C}^1, \mathbf{C}^2, \ldots, \mathbf{C}^L\}$ is the signature obtained from a multivariate sub-Gaussian modeling of a steerable pyramid for a given homogeneous texture image I. Then, the corresponding signature for the rotated version I_θ of that texture by an angle θ is $\mathcal{S}_\theta = \{\mathbf{C}^1_\theta, \mathbf{C}^2_\theta, \ldots, \mathbf{C}^L_\theta\}$ where

$$\mathbf{C}^l_\theta = \mathbf{M}^l_\theta . * D\left(\mathbf{M}^l_\theta\right), \quad \mathbf{M}^l_\theta = \mathbf{F}(\theta) \tilde{\mathbf{C}}^l \mathbf{F}^T(\theta), \tag{12}$$

$$\tilde{\mathbf{C}}^l = \begin{bmatrix} \frac{[\mathbf{C}^l]_{11}}{[\mathbf{C}^l]_{11}^{(\alpha-2)/\alpha}} & \frac{[\mathbf{C}^l]_{12}}{[\mathbf{C}^l]_{22}^{(\alpha-2)/\alpha}} & \cdots & \frac{[\mathbf{C}^l]_{1J}}{[\mathbf{C}^l]_{JJ}^{(\alpha-2)/\alpha}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{[\mathbf{C}^l]_{J1}}{[\mathbf{C}^l]_{11}^{(\alpha-2)/\alpha}} & \frac{[\mathbf{C}^l]_{J2}}{[\mathbf{C}^l]_{22}^{(\alpha-2)/\alpha}} & \cdots & \frac{[\mathbf{C}^l]_{JJ}}{[\mathbf{C}^l]_{JJ}^{(\alpha-2)/\alpha}} \end{bmatrix} = \frac{1}{2}\mathbf{R}^l, \tag{13}$$

$$D\left(\mathbf{M}_\theta^l\right) = \begin{bmatrix} \left\{diag\left(\mathbf{M}_\theta^l\right)\right\} .\wedge \left(\frac{\alpha-2}{2}\right) \\ \vdots \\ \left\{diag\left(\mathbf{M}_\theta^l\right)\right\} .\wedge \left(\frac{\alpha-2}{2}\right) \end{bmatrix} \tag{14}$$

where $\mathbf{R}^l$ is the lth-level underlying covariance matrix, corresponding to the multivariate sub-Gaussian modeling of the original image. The notations $(.*)$ and $(.\wedge)$ represent element-by-element multiplication and element-by-element involution,[8] respectively, while $diag(\mathbf{M}_\theta^l)$ is a row vector containing the main diagonal of the square matrix $\mathbf{M}_\theta^l$. The dimension of all the above matrices equals $J \times J$, where J is the number of basic orientations at each decomposition level.

Proof: The proof follows by replacing $\mathbf{\Sigma}^l$ and $\mathbf{\Sigma}_\theta^l$ with $\mathbf{R}^l$ and $\mathbf{R}_\theta^l$, respectively in (8), and combining the resulting expression with (11). ∎

The above proposition states that the signature $\mathcal{S}_\theta$ constitutes a valid set of steerable features, since the matrices $\mathbf{C}_\theta^l (l = 1, \ldots, L)$, corresponding to the rotated image I_θ, can be obtained directly from the corresponding matrices $\mathbf{C}^l$ of its original version. Even though we assume in Proposition 3 that the texture images are homogeneous, as our experimental results show in Section III, our model tolerates a certain degree of nonhomogeneity.[9]

D. Feature Extraction

Under the multivariate sub-Gaussian modeling of the steerable pyramid subbands, at a given decomposition level, the FE step consists of estimating the $J \times J$ covariation matrix at each decomposition level. Thus, for a given image I decomposed in L levels, we denote by $\mathcal{S}$ the signature given by the set of L estimated covariation matrices

$$I \mapsto \mathcal{S} = \left\{\mathbf{C}_I^1, \mathbf{C}_I^2, \ldots, \mathbf{C}_I^L\right\} \tag{15}$$

where $[\mathbf{C}_I^l]_{ij}$ is equal to the covariation between the ith and the jth subbands at the lth pyramid level (notice that a predetermined enumeration for the orientation subbands at each level is assumed, in our case from left to right). The total size of this signature $\mathcal{S}$ equals $L \cdot J^2$, due to the assymetric nature of the covariation (cf. (3)), that is, $[X, Y]_\alpha \neq [Y, X]_\alpha$ in general.

The signature $\mathcal{S}$ contains only the cross-orientation dependence at a given decomposition level. We may also consider the cross-level dependence by estimating the covariation matrices between consecutive levels (this is basically equivalent to a first-order Markovian dependence used in a HMM). In this case, the signature of an image I is the following:

$$I \mapsto \mathcal{S}_\mathcal{E} = \left\{\mathbf{C}_I^1, \mathbf{C}_I^2, \ldots, \mathbf{C}_I^L, \mathbf{C}_I^{1\to 2}, \ldots, \mathbf{C}_I^{(L-1)\to L}\right\} \tag{16}$$

where $\mathbf{C}_I^{l\to(l+1)}$ denotes the covariation matrix between the J subbands at level l and the J subbands at level $(l+1)$. In particular, the element $[\mathbf{C}_I^{l\to(l+1)}]_{ij}$ is equal to the covariation of the ith subband at level l with the jth subband at the next coarser level $(l+1)$, that is, using (6), we have

$$\left[\mathbf{C}_I^{l\to(l+1)}\right]_{ij} = \frac{\sum_{k=1}^{N^2} c^l(x_k, \phi_i)\left|c^{l+1}(x_k, \phi_j)\right|^{p-1} sign\left(c^{l+1}(x_k, \phi_j)\right)}{\sum_{k=1}^{N^2}\left|c^{l+1}(x_k, \phi_j)\right|^p}\gamma^\alpha \tag{17}$$

where the parameters (α, γ) are estimated from the set of the jth subband coefficients at level $(l+1)$, $\{c^{l+1}(x_k, \phi_j)\}_{k=1,\ldots,N^2}$, using Nolan's ML estimator.

Proposition 3 also holds for $\mathbf{C}_I^{l\to(l+1)}$, by replacing the matrix $\mathbf{R}^l$ by $\mathbf{R}^{l\to(l+1)}$, whose element $[\mathbf{R}_I^{l\to(l+1)}]_{ij}$ is equal to the underlying covariance of the ith subband at level l with the jth subband at the next coarser level $(l+1)$. Thus, we also have

$$\begin{aligned} \mathbf{C}_\theta^{l\to(l+1)} &= \mathbf{M}_\theta^{l\to(l+1)} .* D\left(\mathbf{M}_\theta^{l\to(l+1)}\right) \\ \mathbf{M}_\theta^{l\to(l+1)} &= \mathbf{F}(\theta)\tilde{\mathbf{C}}^{l\to(l+1)}\mathbf{F}^T(\theta) \end{aligned} \tag{18}$$

$$\tilde{\mathbf{C}}^{l\to(l+1)} = \begin{bmatrix} \frac{[\mathbf{C}^{l\to(l+1)}]_{11}}{[\mathbf{C}^{l\to(l+1)}]_{11}^{(\alpha-2)/\alpha}} & \cdots & \frac{[\mathbf{C}^{l\to(l+1)}]_{1J}}{[\mathbf{C}^{l\to(l+1)}]_{JJ}^{(\alpha-2)/\alpha}} \\ \vdots & \vdots & \vdots \\ \frac{[\mathbf{C}^{l\to(l+1)}]_{J1}}{[\mathbf{C}^{l\to(l+1)}]_{11}^{(\alpha-2)/\alpha}} & \cdots & \frac{[\mathbf{C}^{l\to(l+1)}]_{JJ}}{[\mathbf{C}^{l\to(l+1)}]_{JJ}^{(\alpha-2)/\alpha}} \end{bmatrix} = \frac{1}{2}\mathbf{R}^{l\to(l+1)} \tag{19}$$

and the matrix $D(\mathbf{M}_\theta^{l\to(l+1)})$ is defined as in (14).

Notice that the estimation of $[\mathbf{C}_I^{l\to(l+1)}]_{ij}$, between subbands at consecutive pyramid levels, requires that the two subbands have the same number of coefficients. On one hand, by applying the standard steerable pyramid decomposition [16] with subsampling between consecutive levels, the subband coefficients at adjacent levels are associated in terms of a quad-tree structure, where a coefficient at the coarser level is associated with four coefficients at the same subband of the previous finer level. On the other hand, a Fourier-domain implementation of the steerable pyramid can be used instead, simply by multiplying the DFT of the image with the DFT of the filter (without subsampling) and then taking the inverse DFT to obtain the subband coefficients, resulting in subbands with equal size at all decomposition levels (one could use also upsampling of filters without subsampling, but the DFT satisfies more exactly the perfect reconstruction property of the filter bank, since there is no aliasing at all). Then, for a given subband coefficient $c^l(x_k, \phi_i)$, the coefficient at the same spatial location of the next coarser level is simply denoted by $c^{l+1}(x_k, \phi_i)$. Because of these reasons, in the implementation of our texture retrieval scheme, we construct the

[8] $(x_1, \ldots, x_k).\wedge(\mu) = (x_1^\mu, \ldots, x_k^\mu)$.

[9] Further future work could involve using our technique locally together with some segmentation.

steerable pyramids using the Fourier-domain approach, without subsampling between adjacent levels.[10]

Obviously, the signature $\mathcal{S}_E$ contains more information than the signature $\mathcal{S}$, since it exploits not only the interorientation but also the interlevel dependencies. On the other hand, using $\mathcal{S}_E$ results in an increased computational complexity, since its size equals

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

E. Similarity Measurement

In this section, we describe the construction of a similarity function, which performs an angular alignment in the feature space when comparing two image textures. Although our construction has similarities with the one presented in [13], the main difference is that we use an accurate stochastic model which is much more complete than in [13]. Moreover, as we show in Sections II-F and III, our method provides the best compromise between average retrieval performance and computational complexity, as compared to other previously proposed approaches.

Given two images I and Q, let I^l and Q^l be the set of orientation subbands at the lth 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 a certain predetermined order, modeled by multivariate sub-Gaussian distributions with covariation matrices $\mathbf{C}_I^l$, $\mathbf{C}_Q^l$, and $\mathbf{C}_I^{l\rightarrow(l+1)}$, $\mathbf{C}_Q^{l\rightarrow(l+1)}$, respectively. We define the distance between the lth levels of the two images as

$$d(Q^l, I^l) = \left\| \mathbf{C}_Q^l - \mathbf{C}_I^l \right\|. \tag{20}$$

Similarly, the distance between two corresponding sets of subbands at adjacent levels is given by

$$d(Q^{l,l+1}, I^{l,l+1}) = \left\| \mathbf{C}_Q^{l\rightarrow(l+1)} - \mathbf{C}_I^{l\rightarrow(l+1)} \right\|. \tag{21}$$

In the above expressions, in principle, $\|\cdot\|$ could be any of the common matrix norms. However, in our method, we choose the Frobenius norm, which gives an indication of the "matrix amplitude."

We define the overall distance between images I and Q, decomposed in L levels, to be equal to the following sum:

$$d(Q, I) = \sum_{l=1}^{L} d(Q^l, I^l) + \sum_{l=1}^{L-1} d(Q^{l,l+1}, I^{l,l+1}). \tag{22}$$

Note that, if the texture information is represented by the signature $\mathcal{S}$, the overall distance is simply obtained from (22) by omitting the second sum, which corresponds to the interlevel dependencies. Thus, in the following, we assume that the texture information is represented by the enhanced signature $\mathcal{S}_E$, resulting in a more general expression, which includes the expression corresponding to the signature $\mathcal{S}$ as a special case. Notice that the above distance is not a rotation-invariant distance because it does not take into account any relative angle.

In our problem, we deal with databases that may contain rotated versions of a given image. Notice that Proposition 3 gives a relation of equivalence between $\mathbf{C}_{I_\theta}^l$ and $\mathbf{C}_I^l$, as well as between $\mathbf{C}_{I_\theta}^{l\rightarrow(l+1)}$ and $\mathbf{C}_I^{l\rightarrow(l+1)}$, where $\mathbf{C}_{I_\theta}^l$ is the lth-level covariation matrix and $\mathbf{C}_{I_\theta}^{l\rightarrow(l+1)}$ is the covariation matrix between levels l and $l+1$, of image I rotated counter clockwise by an angle θ, which we denote by I_θ. Next, we construct a rotation-invariant retrieval system by defining another distance that involves an angular alignment. We call this distance a "*rotation-invariant*" distance.

Consider Q to be the query image, which is supposed to be a potentially rotated version of an original image I in the database, $\tilde{I} = I_{\theta^*}$, where $\tilde{I}$ is a counter-clockwise rotation of I by an angle θ^*. In a real application, the value of the angle θ^* is unknown. Thus, the rotation-invariant distance between the lth levels of Q and I (Q^l and I^l, respectively) is defined as the minimum of $d(Q^l, I_{-\theta}^l)$, where the minimization is taken over a set Θ of possible rotations. It follows that it is necessary to perform an angular alignment by finding the optimum angle θ^* minimizing $d(Q^l, I_{-\theta}^l)$. The notation $I_{-\theta}^l$ means the lth level of a clockwise rotated version of I by an angle θ.

By using Proposition 3, replacing θ by $-\theta$, the overall rotation-invariant distance between Q and I is given by the following definition.

Definition 2: Let $\mathcal{S}_E(Q)$ and $\mathcal{S}_E(I)$ be the signatures of two given texture images Q and I, respectively. The rotation-invariant distance $\bar{d}(Q, I)$ between Q and I, where Q is the query, is defined as follows:

$$\bar{d}(Q, I) = \min_{\theta\in\Theta} \left[\sum_{l=1}^{L} \left\| \mathbf{C}_Q^l - \mathbf{C}_{I_{-\theta}}^l \right\| + \sum_{l=1}^{L-1} \left\| \mathbf{C}_Q^{l\rightarrow(l+1)} - \mathbf{C}_{I_{-\theta}}^{l\rightarrow(l+1)} \right\| \right] \tag{23}$$

where the matrix $\mathbf{C}_{I_{-\theta}}^l$ is given by the following proposition:

Proposition 4: Under the steerable multivariate sub-Gaussian model, the matrix $\mathbf{C}_{I_{-\theta}}^l$ is equal to

$$\mathbf{C}_{I_{-\theta}}^l = 2^{-\alpha/2} \cdot \left(\mathbf{R}_{I_{-\theta}}^l . * D\left(\mathbf{R}_{I_{-\theta}}^l \right) \right)$$

where

$$\mathbf{R}_{I_{-\theta}}^l = \mathbf{F}(-\theta)\mathbf{R}_I^l\mathbf{F}^T(-\theta), \quad \mathbf{R}_I^l = 2 \cdot \left(\mathbf{C}_I^l . * \bar{D}\left(\mathbf{C}_I^l \right) \right).$$

For any square matrix $\mathbf{M}$, the operators $D(\cdot)$ and $\bar{D}(\cdot)$ are defined as

$$D(\mathbf{M}) = \begin{bmatrix} \{diag(\mathbf{M})\} . \wedge \left(\frac{\alpha-2}{2}\right) \\ \vdots \\ \{diag(\mathbf{M})\} . \wedge \left(\frac{\alpha-2}{2}\right) \end{bmatrix} \tag{24}$$

[10]For small-size images, the construction of a pyramid using subsampling results in a decreased performance of the covariation estimator (6) due to the very small number of samples of the subbands at the coarsest levels. For large-size images, we can employ a pyramid decomposition using subsampling in order to reduce the computational complexity, while maintaining the accuracy of the covariation estimator by decomposing the images in, at most, three or four levels. Then, the element $[\mathbf{C}_I^{l\rightarrow(l+1)}]_{ij}$ could be computed, for instance, by averaging each subband coefficient at the $(l+1)$th level with its associated four coefficients at the finer lth level.

$$\bar{D}(\mathbf{M}) = \begin{bmatrix} \{diag(\mathbf{M})\} .\wedge \left(\frac{2-\alpha}{\alpha}\right) \\ \vdots \\ \{diag(\mathbf{M})\} .\wedge \left(\frac{2-\alpha}{\alpha}\right) \end{bmatrix} \tag{25}$$

where $\mathbf{R}_I^l$ is the lth-level underlying covariance matrix corresponding to the multivariate sub-Gaussian model of image I and $\{diag(\mathbf{M})\}$ denotes a row vector formed from the main diagonal of the square matrix $\mathbf{M}$.

Proof: The proof of the above proposition follows by inverting (12) and expressing $\tilde{\mathbf{C}}^l$ as a function of $\mathbf{C}_\theta^l$. Then, the covariation matrix $\mathbf{C}_{I_{-\theta}}^l$, associated with the clockwise rotated version of I, corresponds to the matrix $\tilde{\mathbf{C}}^l$ and, similarly, $\mathbf{C}_\theta^l$ corresponds to the lth-level covariation matrix of the original (nonrotated) signature of image I, which is *assumed* to be a rotated version of the query Q. ■

The above proposition follows also similarly for the matrices $\mathbf{C}_{I_{-\theta}}^{l\to(l+1)}$, replacing $\mathbf{R}_I^l$ by $\mathbf{R}_I^{l\to(l+1)}$ and $\mathbf{C}_I^l$ by $\mathbf{C}_I^{l\to(l+1)}$, that is

$$\mathbf{C}_{I_{-\theta}}^{l\to(l+1)} = 2^{-\alpha/2} \cdot \left(\mathbf{R}_{I_{-\theta}}^{l\to(l+1)} .* D\left(\mathbf{R}_{I_{-\theta}}^{l\to(l+1)}\right)\right)$$

where

$$\mathbf{R}_{I_{-\theta}}^{l\to(l+1)} = \mathbf{F}(-\theta)\mathbf{R}_I^{l\to(l+1)}\mathbf{F}^T(-\theta),$$
$$\mathbf{R}_I^{l\to(l+1)} = 2 \cdot \left(\mathbf{C}_I^{l\to(l+1)} .* \bar{D}\left(\mathbf{C}_I^{l\to(l+1)}\right)\right).$$

The dimension of all the above matrices equals $J \times J$, where J is the number of orientations at each decomposition level. Moreover, notice that during the actual numerical implementation of the operators $D(\cdot)$ and $\bar{D}(\cdot)$, each element of the vector $\{diag(\mathbf{M})\}$ is rised to a different estimated α, that is, the values of α estimated from each of the various different orientation subbands are close to each other but not exactly equal in practice. On the other hand, our estimation of covariation assumes that we have two joint sub-Gaussian random variables, that is, with equal characteristic exponent values. We use the following strategy: first, we observe that from (6), the free p parameter, which depends on the characteristic exponent α, affects the second variable (as an exponent), that is, we should use the estimated characteristic exponent α, corresponding to the second variable (orientation 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. This is consistent with the definition of covariation.

The intuition behind this similarity function is as follows: consider Q to be the query image and I to be an image in a database. In the SM step, we measure the closeness of Q and I by computing the distance between their signatures. As we may have a database of images along with rotated versions of them, we assume that the given query Q may be in general a potentially rotated version of I. Thus, the signature of image I, $\mathcal{S}_E(I)$, corresponds to our steerable model

$$\mathcal{S}_{\mathcal{E}}(I) = \left\{\mathbf{C}_{I_\theta}^1, \ldots, \mathbf{C}_{I_\theta}^L, \mathbf{C}_{I_\theta}^{1\to2}, \ldots, \mathbf{C}_{I_\theta}^{(L-1)\to L}\right\}$$

for an arbitrary angle θ. Before measuring the similarity between Q and I we have to align their corresponding signatures, that is, to rotate clockwise the signature of I to align it with the signature of Q. This clockwise rotation is expressed by the terms $\mathbf{C}_{I_{-\theta}}^l$ and $\mathbf{C}_{I_{-\theta}}^{l\to(l+1)}$. Since we do not know *a priori* the angle θ^* that gives the best alignment, we have to search over a set of possible rotations by minimizing $d(Q, I_{-\theta})$ over θ.

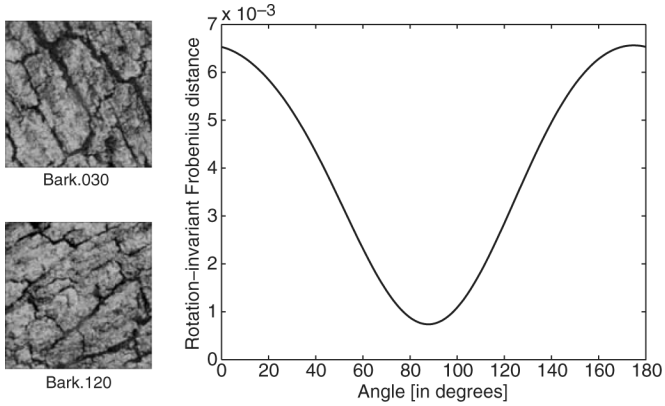

Fig. 1. (a) *Bark* physically rotated at 30° and 120°; (b) $\bar{d}(Q, I_{-\theta})$ for $J = 4$. Notice that the minimum is achieved for approximately $\theta^* = 90°$, which is the exact relative angle between the two texture samples.

Notice that when Q and I are in fact two rotated versions of the same image, the angle θ^*, for which the minimum is achieved in (23), should be close to their exact relative angle, that is, the angle one needs to rotate clockwise the image Q in order to get I. Thus, one way to evaluate the performance of the above rotation-invariant similarity function (23), 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, this functionality may be also useful on its own for many practical applications where there is a need to find out approximately this relative angle. Fig. 1 illustrates this by showing the function $\bar{d}(Q, I_{-\theta})$ given by (23), for the *Bark* texture sample obtained from the Brodatz database.

The rotation-invariance property of the similarity function (23) is also illustrated in Fig. 2, where the query image Q (*Bark*) and the database image I (*Water* and *Leather*) belong to different texture classes. Each curve in Fig. 2 corresponds to the distance $\bar{d}(Q_\phi, I_{-\theta})$ (as a function of θ), that is, the distance between the four versions of the *Bark* image [original ($\phi = 0°$) and three rotated versions (30°, 60°, 120°)] and the original *Water* image rotated clockwise at several angles $\theta \in \Theta$. Similarly, each curve in Fig. 2 corresponds to the distance between the same four versions of the *Bark* image and the clockwise rotations, at several angles $\theta \in \Theta$, of the rotated version of *Leather* image at 90°. It is clear that these curves are approximately shifted versions of each other, which shows (experimentally) the rotation-invariance property of the proposed similarity function (23). However, the expression of this similarity function is quite complicated and norm-dependent, and, thus, it is difficult to provide an analytical expression for the optimal angle θ^*, that is, the angle that achieves the minimum in (23). As we can also see in both cases of Fig. 2, the curve corresponding to the version of *Bark* rotated at $\phi = 120°$ has a higher amplitude than the other three curves. This is due to the nonhomogeneous nature of the *Bark* texture image. Besides, both of the higher amplitude curves seem to have a similar peak-to-peak range. How-

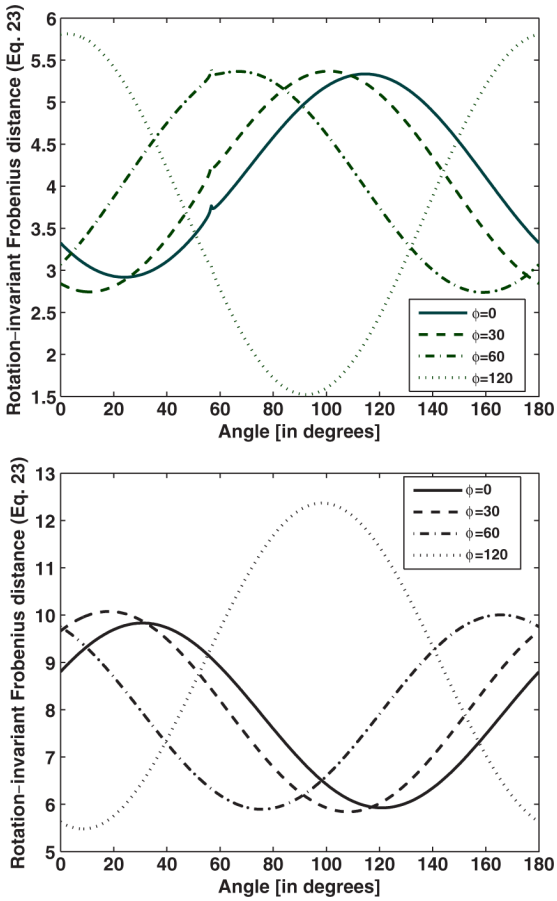

Fig. 2. Distance $\bar{d}(Q_\phi, I_{-\theta})$, as a function of θ, between four versions of *Bark* (original ($\phi = 0°$) and three physically rotated at $\phi = 30°$, 60°, 120°) and (a) the original image *Water*; (b) the image *Leather* rotated at 90°.

ever, it is important to note that this higher amplitude does not affect the rotation invariance property since the same phenomenon will appear when comparing the query texture image to the various texture images of the different classes. Notice also how in Fig. 2, all the curves corresponding to the *Water* texture are located vertically at lower positions than the same curves corresponding to the *Leather* texture (Fig. 2).

If we represent the texture information using the signature $\mathcal{S}$ instead of $\mathcal{S}_E$, the distance between two images Q and I is measured using (23), by taking only the first L summands. Moreover, under a Gaussian assumption (Gaussian signature), (23) can be simplified and takes the following form [13]:

$$\bar{d}_g(Q,I) = \min_{\theta} \sum_{l=1}^{L} \left\| \mathbf{C}_Q^l - \mathbf{F}(-\theta)\mathbf{C}_I^l \mathbf{F}^T(-\theta) \right\| \qquad (26)$$

where, in this case, $\mathbf{C}^l$ denotes the matrix with elements being the covariances between pairs of subbands at the lth pyramid level.

It is also important to notice that, in the proposed retrieval method, we achieve rotation invariance through alignment in angle between signatures extracted using a steerable sub-Gaussian model and not by calculating a fixed set of rotation-invariant features valid for all images of the same class, namely, the set of images that are rotated versions of each other. The latter approach would involve diagonalizing matrices and extracting eigenvalues, as in [14] for instance, which has an important complexity.

F. Computational Complexity

As mentioned before, one of the goals of this paper is to design a new texture retrieval system that provides *the best compromise between complexity and average retrieval performance*. In this section, we compare the computational complexity of the retrieval technique proposed in this work, with the complexity of three other rotation-invariant texture retrieval methods that have been proposed before and which, to the best of our knowledge, represent the current state-of-the-art.

1) M1 [13]: This method makes a Gaussian assumption for the statistics of the pyramid coefficients and uses a rotation-invariant similarity function based on the Frobenius norm of differences between covariance matrices.
2) M2 [10]: In contrast with M1, this retrieval scheme assumes that the marginal and joint statistics of the pyramid coefficients are best described using members of the family of $S\alpha S$ and multivariate sub-Gaussian distributions. Then, it applies a Gaussianization procedure that consists in extracting the signature of each texture image based on second-order statistics, followed by a rotation-invariant version of the KLD to measure the similarity.
3) M3 [14]: This method is the best current representative within the category of rotation-invariant texture retrieval methods that are based on the statistical fitting of the steerable pyramid coefficients using an HMM.

The retrieval method proposed in this paper, as we similarity measurement (SM). On the other hand, method M2 has one more additional component, which consists of a Gaussianization procedure applied before the FE step, adding, therefore, more complexity to the whole retrieval process. In the subsequent analysis, the computational cost is expressed as a function of the number of operations required, where an "operation" is defined as the pair consisting of a multiplication between two numbers and a summation $(a \cdot b + c)$.

Let $M \times M$ denote the dimension of the original image, which is decomposed via an L-level steerable pyramid with J basic orientations and P to be equal to the neighborhood size that is used during the Gaussianization process in method M2. In addition, let K denote the number of matrices contained in the selected signature, which varies depending on whether we exploit only intralevel, or both intra- and interlevel dependencies, that is, $K = L$ or $K = 2L - 1$.

First, we compare the complexity of the FE step for the four methods, including our proposed retrieval system. Using the above notation, the computational cost of the FE step of our method is approximately equal to $K \cdot J^2 \cdot \mathcal{O}(M^2)$, where $\mathcal{O}(M^2)$ is the approximate cost for the computation of the covariation between a pair of subbands using the FLOM covariation estimator of (6). Assuming that the steerable pyramid is constructed without subsampling between consecutive levels, the constant in the $\mathcal{O}(\cdot)$ function is small, at the order of 2.

Regarding methods M1 and M2, the cost of extracting the signature of an image is of order $K \cdot [J(J+1)/2] \cdot \mathcal{O}(M^2)$, since both methods estimate covariances between pairs of subbands,

which requires approximately the same order of numerical operations as the estimation of covariations (the constant in the $\mathcal{O}(\cdot)$ function is of order 1). In addition, the covariance matrices are symmetric, in contrast to the covariation matrices, and, thus, in methods M1 and M2, we have to estimate $[J(J+1)/2]$ elements per matrix, instead of J^2 covariations for each matrix of the signature of the texture retrieval scheme proposed in this work. Taking into consideration that in practice the value of J is usually 3 or 4 and does not depend on the image size, we can conclude that the computational cost of the FE step of our proposed method $(K \cdot J^2 \cdot \mathcal{O}(M^2))$ is very close to the cost of methods M1 and M2 $(K \cdot [J(J+1)/2] \cdot \mathcal{O}(M^2))$.

Finally, the cost of the FE step for the HMM-based method M3 is difficult to be estimated with precision, since it employs the expectation maximization (EM) algorithm using the ML criterion [12] to estimate the model parameters, and, thus, it may not even converge in some cases. In order to give a rough estimation of the computational cost of method M3 we argue as follows. The FE step consists of estimating $(2L-1)$ hidden states and $(2 \cdot L \cdot J)$ eigenvalues corresponding to the $(2L)$ covariance matrices (each one of dimension $J \times J$) of the Gaussian densities used in the model. The Expectation step of the EM algorithm is particularly difficult in this case because of the increased interplay between the states. Besides, for an HMM, the complexity of *each iteration* of EM is linear in the number of observations, that is, the number of subband coefficients in this case. Since method M3 constructs a steerable pyramid by subsampling between adjacent levels, the total number of wavelet coefficients for an L-level pyramid with J orientations per level is equal to $JM^2(1-4^{-L})/3$. The total number of model parameters to be estimated is equal to $(2L-1)+2L(J(J+1)/2)$, where $2L(J(J+1)/2)$ is the total number of covariances due to the symmetry of the $2L$ covariance matrices. Notice that an additional computational cost is required to perform the eigenvalue decomposition of the covariance matrices in order to obtain their eigenvalues. However, we consider this specific cost to be negligible (as compared to the rest of costs), since in practice the value of J is at most 3 or 4, and, thus, the dimension of the covariance matrices is very low. Therefore, the computational cost of the FE step of method M3 is of order $[(2L-1)+2LJ] \cdot \mathcal{O}(JM^2(1-4^{-L})/3)$, where the constant in the $\mathcal{O}(\cdot)$ notation depends on the number of iterations required for the EM algorithm to converge.

As a general conclusion, with respect to the computational cost of the FE step of the four methods, the cost of our proposed retrieval technique is approximately equal to the costs of methods M1 and M2, whereas taking into account the usual practical values of L and J, which are usually set to at most 3 or 4, we can conclude that the corresponding cost of method M3 is comparable with the costs of the other three methods only if the EM algorithm converges in very few iterations. For instance, for $L=3$ and $J=3$ the EM algorithm should converge in no more than 5 iterations. Unfortunately, such a fast convergence is not guaranteed *a priori*, making in practice the FE step of method M3 more computationally expensive, in general, than the FE step of the other three methods.

Regarding the SM step, the approach proposed in this work as well as the techniques M1 and M2, all apply Newton's method to minimize a rotation-invariant similarity function. The computational cost for performing the minimization in these three methods is basically the same, due to the high smoothness of the corresponding expressions inside the min operator in (23), as functions of θ. It is also important to note that, when the steering functions have only odd harmonics (as in our case), they oscillate at some finite speed, which implies an upper bound on the slope of these curves. Thus, it can be readily verified that the number of local minima of the rotation-invariant similarity functions in the texture retrieval method proposed here, as well as in M1 and M2, 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 nonoverlapping angular intervals. In particular, the application of Newton's method in these three retrieval schemes converges very rapidly in practice, namely, in, at most, five iterations in each interval (using the middle of each interval as the initial point).

The similarity function proposed in this work, requires for each iteration (during the minimization) about $K \cdot (3J^3+4J^2+J)$ operations. The corresponding cost of each iteration for method M1 is equal to $K \cdot (3J^3+2J)$, while method M2 requires about $K \cdot (5J^3+J+D)$, where D denotes the number of operations for the computation of the determinant of a $J \times J$ matrix (see [10] and [13] for more details). Using an LU decomposition of a $J \times J$ matrix, the determinant is computed in $J^3/3$ operations; thus, in practical applications, the value of D is small, since the value of J is small, usually $J=3$ or 4. Finally, the HMM-based method M3, uses Monte-Carlo simulations to evaluate the integral of the KLD, which, for the experimental setup described in the following section, consists of about 64 iterations [14]. In each iteration, data are randomly generated from the query model as trees of wavelet coefficients, and then its likelihood is computed against a candidate model. Thus, the computational cost of each iteration is proportional to the number of subband coefficients, making the SM step of method M3 much more computationally expensive than the corresponding costs of the other three methods, which depend only on the parameters of the steerable pyramid (L, J), which are much smaller than the dimension of the images $(M \times M)$.

Note that in the above analysis, the computational cost of each method during the SM step, except for method M3, is expressed only as a function of K and J. This is because of the fact that the computational cost for the estimation of the matrices constituting the signature of each method, which depends on M, was taken into account during the analysis for the computational complexity of the FE step, and, thus, the signature size that determines the computational cost depends only on K and J, since these matrices are already available during the SM.

Moreover, it is very important to take also into account that method M2 applies a Gaussianization process, in addition to the FE and SM steps, which adds a significant computational cost that is approximately equal to $L \cdot J \cdot [\mathcal{O}(P^3)+((5/4)M^2+2)P^2+(1/4)M^2P+(1/4)M^2]$, where $\mathcal{O}(P^3)$ is the number of operations for the inversion of a $P \times P$ matrix.

TABLE I
COMPUTATIONAL COMPLEXITY OF THE RETRIEVAL SCHEMES

Method	**Gaussianization step**	**FE step**	**SM step**
PROPOSED	-	$KJ^2\mathcal{O}(M^2)$	$C_{pr}K(3J^3+4J^2+J)$
M1.	-	$K[J(J+1)/2]\mathcal{O}(M^2)$	$C_{M1}K(3J^3+2J)$
M2.	$LJ[\mathcal{O}(P^3)+(\frac{5}{4}M^2+2)P^2$ $+\frac{1}{4}M^2P+\frac{1}{4}M^2]$	$K[J(J+1)/2]\mathcal{O}(M^2)$	$C_{M2}K(5J^3+J+D)$
M3.	-	$[(2L-1)+2LJ]\mathcal{O}(JM^2(1-4^{-L})/3)$	$C_{M3}\mathcal{O}(JM^2(1-4^{-L})/3)$

- $\mathbf{M1}$: Gaussian assumption for the statistics of the pyramid coefficients; Frobenius norm-based similarity function of differences between covariance matrices.
  $\mathbf{M2}$: multivariate sub-Gaussian assumption for the statistics of the pyramid coefficients + Gaussianization; rotation-invariant version of the KLD between multivariate Gaussians as a similarity function.
  $\mathbf{M3}$: HMM-based statistical fitting of the pyramid coefficients.
- $M \times M$: the dimension of the original image.
- L: number of steerable pyramid levels.
- J: number of basic orientations per level.
- K: number of matrices contained in the selected signature ($K = L$ or $K = 2L-1$).
- P: neighborhood size that is used during the Gaussianization process.
- D: number of operations for the computation of the determinant of a $J \times J$ matrix.
- $\{C_{pr}, C_{M1}, C_{M2}\} \approx 5$, $C_{M3} \approx 64$: number of iterations for the minimization of the similarity function.
- $g(n) = \mathcal{O}(f(n)) \Leftrightarrow \exists\ 0 < c_1 < c_2$ and n_0 s.t. $c_1 f(n) \leq g(n) \leq c_2 f(n)\ \forall n \geq n_0$

After all this analysis, we conclude the following.

- The texture retrieval method proposed in this work has a slightly increased complexity than method M1; however, as we show in the next section of the experimental results, our method results in a better retrieval performance.
- The similarity function of our method is less complex than the similarity function designed in method M2 after the Gaussianization process; on the other hand, the experimental results presented in the following section, show an increased retrieval performance of method M2 compared with the performance of our proposed method.
- Finally, as we analyzed above, the computational costs of the FE and SM steps of our method are in general lower than the corresponding costs of method M3, while achieving a comparable retrieval performance, as our experimental results show in the next section.

Thus, the above observations indicate that the texture retrieval method proposed in this work provides the best compromise among all other methods, in the sense of keeping a reduced complexity and offering a competitive retrieval performance. Table I summarizes the computational costs of our proposed method and those of methods M1, M2, and M3.

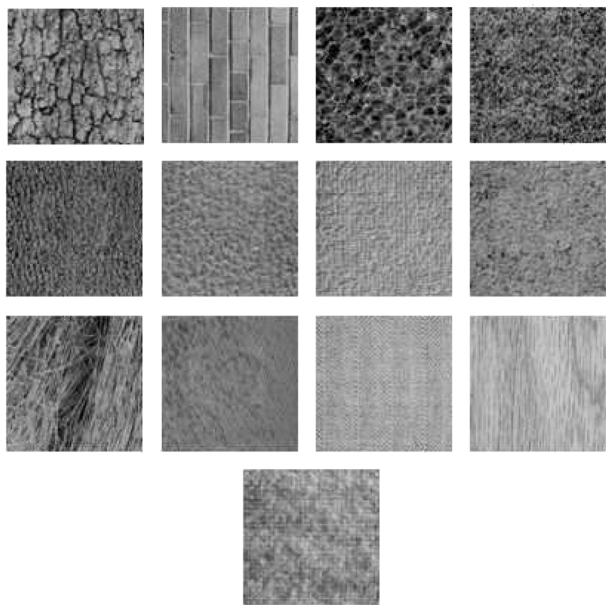

Fig. 3. Texture images from the VisTex database, from left to right and top to bottom: 1) bark, 2) brick, 3) bubbles, 4) grass, 5) leather, 6) pigskin, 7) raffia, 8) sand, 9) straw, 10) water, 11) weave, 12) wood, 13) wool.

III. EXPERIMENTAL RESULTS

In this section, we evaluate the efficiency of our overall texture retrieval system and compare it with the performance of the three methods M1, M2, and M3, presented in Section II-F.

In order to evaluate the retrieval effectiveness of our method and perform the comparison with the other methods, we apply the same experimental setup as in [10] and [14]. In particular, the image database consists of 13, 512×512 Brodatz texture images obtained from the USC SIPI database[11] (cf. Fig. 3). Each of them was physically rotated at 30°, 60°, and 120°, before being digitized. Then, the texture image dataset was formed by taking four nonoverlapping 128×128 subimages each from the original images at 0°, 30°, 60°, and 120°. Thus, the dataset used in the retrieval experiments contains 208 images that come from 13 texture classes.

[11]http://www.sipi.usc.edu/services/database.

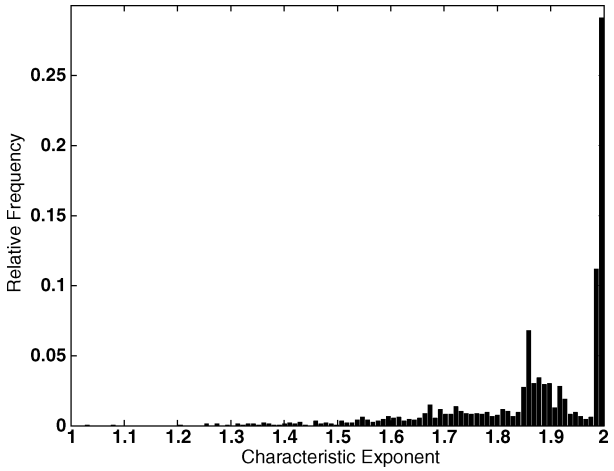

Fig. 4. Histogram of the estimated values for the characteristic exponent, α, for the set of 208 texture images of size 128 $\times$ 128.

The selection of the steerable-pyramid parameters, namely, the number of scales and orientations, affects the performance of the retrieval system. In particular, the set of scales captures a specific band of frequency components from an image, while the set of orientations is used to extract directional features. By construction, each subband at a given scale and orientation corresponds to a distinct filter. However, the more scales and orientations, or equivalently, the more filters we use, does not necessarily result in an increased retrieval performance. An explanation for this behavior is that the more filters we use, the more detailed and redundant the representation of a given image is. Thus, this may not result in a better retrieval performance, since similar features may be captured by the different filters when they are applied on distinct images. On the other hand, as it was analyzed before, the computational complexity of our proposed method increases, when the number of scales and orientations increases. For these reasons, and trying to compromise a high retrieval performance with a reduced computational cost, we implemented a three-level steerable pyramid decomposition with $J = 2$ basic orientations, $\phi_1 = 0$, $\phi_2 = \pi/2$, which means that the steering functions are [15]

$$f_1(\theta) = \cos(\theta), \quad f_2(\theta) = f_1\left(\frac{\pi}{2} - \theta\right) = \sin(\theta).$$

The histogram of the estimated characteristic exponent values for the 208 textures is shown in Fig. 4. We observe that only 28% of the textures exhibit very strong Gaussian statistics, corresponding to α values equal to 2, which belong mainly to pyramid subband coefficients of images 4, 5, and 9. Table II shows the average value of α, over all the pyramid subbands, for each texture class. We observe that images 4, 5, and 9 are exactly those images whose average value of α is closest to 2, compared to the remaining images.

In the following, the query is any of the nonoverlapping 128$\times$ 128 subimages in the image dataset. The relevant images for each query are defined as the other 15 subimages from the same original 512 $\times$ 512 image. The retrieval performance is evaluated in terms of the percentage of relevant images among the top 15 retrieved images.

First, we evaluate the performance of the steerable multivariate sub-Gaussian model combined with the rotation-invariant version of the Frobenius distance, which is proposed in this work. As mentioned before, the signature $\mathcal{S}$, containing the estimated covariation matrices between pairs of subbands at the same decomposition level, is a subset of the enhanced signature $\mathcal{S}_E$ given by (16). Thus, we extract the signature $\mathcal{S}_E$ from each texture sample and then, the similarity measurement for the simpler case of using signature $\mathcal{S}$ is given from (23) by simply using only the first L matrices of $\mathcal{S}_E$. For the 3-level pyramid decomposition that we apply, the signatures $\mathcal{S}$ and $\mathcal{S}_E$ contain 3 and 5 matrices, respectively.

TABLE II
AVERAGE VALUE OF α, OVER ALL THE PYRAMID SUBBANDS, FOR EACH TEXTURE CLASS

Texture Class	Aver. α	Texture Class	Aver. α
1	1.8056	8	1.8452
2	1.6512	9	1.9539
3	1.6992	10	1.7563
4	1.9637	11	1.9257
5	1.9695	12	1.7939
6	1.8692	13	1.9266
7	1.8741		

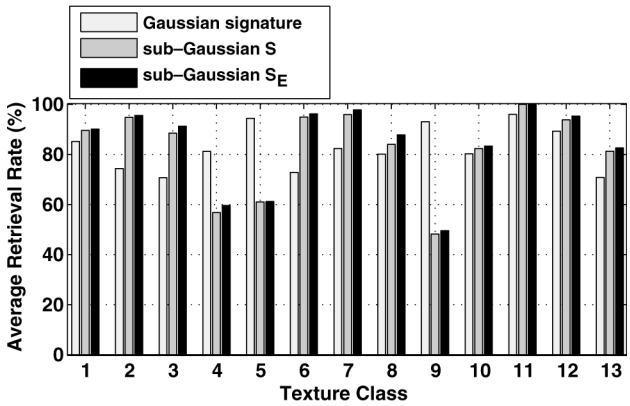

Fig. 5. Average percentages (%) of correct retrieval rate for each individual texture class.

Fig. 5 shows the average percentages of correct retrieval rates for each one of the 13 selected texture classes, for the Gaussian signature of size equal to 3, which contains the sample covariance matrices between pairs of subbands at the same level, combined with the corresponding similarity function (26) (method M1) and the sub-Gaussian signatures $\mathcal{S}$ and $\mathcal{S}_E$, combined with the similarity function (23). We observe that the Gaussian model is better only for 3 classes (4, 5, 9), which are exactly those with the greatest portion of characteristic exponent values (basically) equal to 2. Even for values of α that are slightly smaller than 2, our method provides already a superior performance. The explanation for the decreased performance of our proposed method in the Gaussian case is that the steerable sub-Gaussian model, described in Proposition 3, loses some of its intrinsic information, since the exponents $((\alpha - 2)/\alpha, (\alpha - 2)/2)$ in (12) tend to zero as α tends to 2 (Gaussian case), which tends to cause numerical instability.

Regarding our proposed method, we observe that the enhanced signature $\mathcal{S}_E$ results in only a slightly better retrieval performance than the signature $\mathcal{S}$. Although one could have expected to have a higher increase on the average retrieval rate

TABLE III
AVERAGE RETRIEVAL RATE (%) IN THE TOP 15 MATCHES, FOR THE SUBSET OF 160 (128 × 128) TEXTURES WITH HEAVY-TAILED MARGINAL DISTRIBUTIONS OF THE PYRAMID SUBBANDS

Methods		
Gaussian signature & Frobenius (26)	$\mathcal{S}$ & Frobenius (23)	$\mathcal{S}_{\mathcal{E}}$ & Frobenius (23)
80.18	90.52	92.01

when using $\mathcal{S}_E$ instead of $\mathcal{S}$, this does not happen because there is not a strong correlation between subbands at adjacent levels for the particular set of textures. This can be verified by examining the off-diagonal elements of the $\mathbf{C}_I^{l\rightarrow(l+1)}$ matrices, which have relatively small amplitudes compared with the corresponding elements of the $\mathbf{C}_I^l$ matrices. Due to the increased computational complexity, the enhanced signature should be employed only when a stronger interscale correlation is present.[12]

In order to focus on the non-Gaussian case (the usual case found also in practice in natural images), we omit the three classes with the strong Gaussian behavior and repeat the above experiment for the remaining set of 10 original, nonrotated textures (thus, giving 160 nonoverlapping subimages). Table III shows the corresponding average retrieval rates for the subset of 10 texture classes, for the Gaussian signature of size equal to 3, which contains the sample covariance matrices between pairs of subbands at the same level, combined with the corresponding similarity function (26) (method M1) and the sub-Gaussian signatures $\mathcal{S}$ and $\mathcal{S}_E$, combined with the similarity function (23). As expected, there is a significant improvement of the average retrieval performance when the pyramid coefficients follow a distinct heavy-tailed non-Gaussian distribution.

Fig. 6 shows the average retrieval rates for each individual texture class, of the HMM-based method M3, the method M2 that applies a Gaussianization process on the pyramid coefficients and the method proposed in this work using the enhanced signature $\mathcal{S}_E$. It is clear that, for the texture classes whose marginal and joint distributions between the subband coefficients present a heavy-tailed non-Gaussian behavior, our proposed texture retrieval method outperforms the method based on HMMs (which, in addition, has a larger computational cost in practice), while at the same time maintaining a high retrieval performance, which is comparable to the performance of the method M2, which applies a Gaussianization step before feature extraction. Again, we observe that only for the classes 4, 5, and 9, whose marginal distribution of each subband is close to a Gaussian, the rotation-invariant method proposed in this work results in the lowest retrieval performance among the three methods compared in this figure.

Even though, for the set of texture images that we use in our experiments, the retrieval performance of the method M2 is slightly better than the performance of the method proposed in this work, as mentioned in Section II-F, notice also that it has a substantially higher computational complexity due to the Gaussianization procedure. Thus, in the case of a retrieval system with limited computational resources, the implementation of the rotation-invariant scheme proposed here preserves a high retrieval efficiency, while keeping a reasonably low complexity.

[12]To the best of our knowledge, there does not exist electronic databases containing rotated versions of texture images, other than the one we use here.

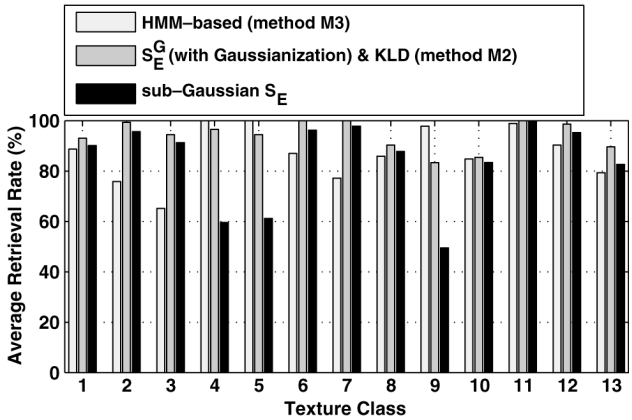

Fig. 6. Average percentages (%) of correct retrieval rate for each individual texture class.

In conclusion, our method provides an average retrieval performance that is: a) superior to the method M1, while keeping a similar complexity, b) superior to the method M3, while keeping at the same time even a lower complexity, c) competitive (slightly inferior) with the method M2, which has a substantially higher computational complexity.

IV. CONCLUSION

In this paper, we described the design of a rotation-invariant texture retrieval system that exploits the non-Gaussian heavy-tailed behavior of the distributions of the subband coefficients, representing the texture information via a steerable pyramid. In particular, we constructed a steerable multivariate sub-Gaussian model, which relates the fractional lower-order statistics of a rotated image with those of its original version. We performed the similarity measurement between two images by inverting this model, which is equivalent to performing a signature alignment in angle, and minimizing a rotation-invariant similarity function based on the Frobenius matrix norm.

The motivation for the development of the proposed rotation-invariant texture retrieval system was the increased computational complexity of our recently introduced method [10], which includes a Gaussianization step before extracting the image signatures, combined with the implementation of a rotation-invariant version of the KLD between multivariate Gaussian densities. We illustrated that, when the marginal and joint distributions of the subband coefficients are heavy-tailed (the usual case encountered in practice with natural images), the proposed steerable sub-Gaussian model achieves a high retrieval performance, while maintaining a sufficiently lower computational complexity.

A future research direction for the improvement of the retrieval performance of the retrieval method proposed in this work, is to study the construction of a rotation-invariant statistical similarity measure, such as the KLD, between multivariate sub-Gaussian distributions, to be used instead of a deterministic

norm-based measure. Additional future work could involve extension to nonhomogeneous textures by using segmentation.

REFERENCES

[1] J. Ashley, M. Flickner, J. Hafner, D. Lee, W. Niblack, and D. Petkovic, "The query by image content (QBIC) system," in *Proc. ACM SIGMOD Int. Conf. Management of Data*, 1995, p. 475.

[2] S. Belongie, J. Malik, and J. Puzicha, "Matching shapes," presented at the Int. Conf. Computer Vision, Vancouver, BC, Canada, 2001.

[3] B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[4] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.

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

[6] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Process.*, vol. 4, no. 11, pp. 1549–1560, Nov. 1995.

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

[8] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–692, Jul. 1989.

[9] E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *Proc. 31st Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, 1997, pp. 673–678.

[10] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides, "Rotation-invariant texture retrieval with gaussianized steerable pyramids," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2702–2718, Sep. 2006.

[11] G. Cross and A. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 1, pp. 25–39, Jan. 1983.

[12] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[13] B. Beferull-Lozano, H. Xie, and A. Ortega, "Rotation-invariant features based on steerable transforms with an application to distributed image classification," presented at the IEEE Int. Conf. Image Processing, Barcelona, Spain, 2003.

[14] M. N. Do and M. Vetterli, "Rotation invariant texture characterization and retrieval using steerable wavelet domain hidden Markov models," *IEEE Trans. Multimedia*, vol. 4, no. 4, pp. 517–527, Dec. 2002.

[15] 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, Sep. 1991.

[16] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

[17] J. Huang, "Study on the correlation properties of wavelet transform coefficients and the applications in a neural network-based hybrid image coding system," presented at the CISST, Las Vegas, NV, Jun. 22–26, 2003.

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

[19] J. P. Nolan, "Parameterizations and modes of stable distributions," *Statist. Probab. Lett.*, no. 38, pp. 187–195, 1998.

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

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

[22] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE* , vol. 81, no. 7, pp. 986–1010, Jul. 1993.

George Tzagkarakis received the B.S. degree (first in class, honors) in mathematics from the University of Crete (UOC), Greece, in 2002, and the M.Sc. degree (first in class, honors) from the Computer Science Department (CSD), UOC, in 2004, with scholarships from the CSD and the Institute of Computer Science (ICS) of the Foundation for Research and Technology-Hellas (FORTH). He is currently pursuing the Ph.D. degree at the UOC in the area of signal processing with applications in sensor networks.

Since 2000, he has been also collaborating with the Wave Propagation Group of the Institute of Applied and Computational Mathematics (FORTH), and he has been a Research Assistant in the ICS. His research interests lie in the fields of statistical signal and image processing with an emphasis on non-Gaussian heavy-tailed modeling, distributed signal processing for sensor networks, information theory, and applications in image classification/retrieval and inverse problems in underwater acoustics.

Baltasar Beferull-Lozano (S'01–M'02–SM'08) was born in Valencia, Spain, in 1972. He received the M.Sc. degree in physics from the Universidad de Valencia (UV), Valencia, Spain, in 1995 (first in class honors) and the M.Sc. and Ph.D. degrees in electrical engineering from University of Southern California (USC), Los Angeles, in 1999 and 2002, respectively. His Ph.D. work was supported by a National Graduate Doctoral Fellowship from the Ministry of Education of Spain.

From January 1996 to August 1997, he was a Research Fellow Assistant with the Department of Electronics and Computer Science, UV, and from September 1997 to September 2002, he was a Research Fellow Assistant in the Department of Electrical Engineering, the NSF Research Integrated Media Systems Center and the Signal and Image Processing Institute (SIPI), USC. He has also been with AT&T Shannon Laboratories. From October 2002 to June 2005, he was a Research Associate in the Department of Communication Systems at the Swiss Federal Institute of Technology, Lausanne, Switzerland, and a Senior Researcher within the Swiss National Competence Center in Research on Mobile Information and Communication Systems. From July 2005 to November 2005, he was a Visiting Professor at the Universidad Politécnica de Valencia. Since December 2005, he has been a Research Professor in Instituto de Robótica-Escuela Técnica Superior de Ingeniería at Universidad de Valencia. He has served as a member of the Technical Program Committees for several ACM & IEEE International Conferences. His research interests are in the general areas of signal and image processing, distributed signal processing and communications for wireless networks, information theory, and communication theory.

Dr. Beferull-Lozano received several awards while at USC, including the Best Ph.D. Thesis paper Award in April 2002 and the Outstanding Academic Achievement Awards in April 1999 and April 2002.

Panagiotis Tsakalides (M'95) received the Diploma in electrical engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1990, and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 1995.

He is an Associate Professor of computer science at the University of Crete, Greece, and a Researcher with the Institute of Computer Science, Foundation for Research and Technology-Hellas (FORTH-ICS), Greece. From 2004 to 2006, he served as the Department Chairman. From 1999 to 2002, he was with the Department of Electrical Engineering, University of Patras, Greece. From 1996 to 1998, he was a Research Assistant Professor with the Signal and Image Processing Institute, USC, and he consulted for the U.S. Navy and Air Force. His research interests lie in the field of statistical signal processing with emphasis in non-Gaussian estimation and detection theory, and applications in wireless communications, imaging, and multimedia systems. He has coauthored over 60 technical publications in these areas, including 20 journal papers.

Dr. Tsakalides was awarded the IEE's A. H. Reeve Premium in 2002 (with coauthors P. Reveliotis and C. L. Nikias) for the paper "Scalar quantization of heavy tailed signals" published in the October 2000 issue of the *IEE Proceedings-Vision, Image and Signal Processing*.