# ROTATION-INVARIANT TEXTURE RETRIEVAL WITH GAUSSIANIZED STEERABLE PYRAMIDS

*G. Tzagkarakis*[1], *B. Beferull-Lozano*[2], *and P. Tsakalides*[1]

[1]Department of Computer Science
University of Crete and
Institute of Computer Science (ICS-FORTH)
711 10 Heraklion, Crete, Greece
{gtzag,tsakalid}@ics.forth.gr

[2] School of Computer and Comm. Sciences
Audiovisual Communications Laboratory
Swiss Federal Institute of Technology (EPFL)
CH1015, Lausanne, Switzerland
Baltasar.Beferull@epfl.ch

**ABSTRACT**

This paper presents a novel rotation-invariant image retrieval scheme based on steerable pyramid transforms. First, we model the subband coefficients as *sub-Gaussian* random vectors 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 of two distinct images is measured by minimizing the Kullback-Leibler Divergence (KLD) between their corresponding multivariate Gaussian distributions, where the minimization is performed over a set of rotation angles. We provide analytical expressions for the minimum KLD and we demonstrate the effectiveness of our proposed method using a set of real texture images.

## 1. INTRODUCTION

The search of large digital multimedia libraries, unlike the search of conventional text-based digital databases, cannot be realized by simply searching text annotations. But the design of completely automatic mechanisms that extract meaning from multimedia data and characterize the information content in a compact and meaningful way is a challenging task. Accordingly, efficient content-based information retrieval systems must be developed on the basis of automatically-derived features that accurately specify the information content of the items in the database.

In a typical content-based image retrieval (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 to accurately represent the content of a given image, creating a better correlation of the pixel representation with image semantics. 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 to a query image each image in the database is, by comparing their signatures.

This work was supported by the Greek General Secretariat for Research and Technology under Program ΕΠΑΝ, Codes ΗΠΑ-011 and 1308/B1/3.3.1/317/12.04.2002.

In recent work, we have shown that by employing a $2D$ wavelet decomposition on a given image, successful modeling of the wavelet subband coefficients is achieved by taking into consideration the actual heavy-tailed behavior of their marginal densities [1]. Specifically, we have shown that the subband decompositions of many texture images have non-Gaussian statistics that are best described by families of distributions with algebraic tails, such as the Symmetric Alpha-Stable ($S\alpha S$) [2]. After extracting the $S\alpha S$ model parameters, we analytically derived and employed the KLD as the distance measure between two $S\alpha S$ distributions. Our formulation improved the retrieving performance of the system, resulting in a decreased probability error rate for images with distinct non-Gaussian statistics. However, this approach does not take into account the possible interdependencies between different subbands of a given image, which can be exploited in order to provide a more accurate representation of the texture image profile. More accurate modeling can be achieved by taking into account the correlation structure between the raw coefficients, their magnitudes, and other statistics [3].

An important property of a CBIR system is rotation invariance. Recently, a rotation-invariant image retrieval system based on steerable pyramids was proposed by Beferull-Lozano *et al.* [4]. In this system, the correlation matrices between the basic orientations at each level of the pyramid are chosen as the energy-based texture features. Then, the similarity measure between two images is defined as the minimum Frobenius norm, over all possible rotations $\theta$, of the difference between the correlation matrix of the original (query) image and that of each image in the database.

In the present work, we proceed first by defining a "neighborhood" for each transform coefficient and then by considering its components as a sample of a sub-Gaussian random process. Within the framework of sub-Gaussian processes we use the notion of covariation, pertaining to lower-than-two order correlation, in order to extract possible interdependencies between coefficients at different image orientations and scales. As a second step, we apply a Gaussianization procedure to normalize the distribution of the raw subband coefficients [5]. The normalization procedure is justified by the fact that the KLD between Gaussian densities has a closed-form expression. The similarity measurement between a given query and a rotated image in a database, is performed by minimizing the KLD between multivariate Gaussian distributions over a set of rotation angles. We derive an analytical expression for the rotation invariant KLD between a given query image $Q$ and a database image $I$ and we describe a numerical method for its computation.

## 2. JOINT STATISTICAL MODELING IN THE TRANSFORM DOMAIN

### 2.1. Variance-adaptive local modeling using multivariate sub-Gaussian distributions

The local dependencies of the coefficient magnitudes at a given subband and the associated marginal distributions, can be modeled using a homogeneous random field with a spatially changing variance. This requirement can be realized by modulating each coefficient (node of the field) with hidden scaling random variables.

A representative example of such a field is the *Gaussian scale mixture* (GSM) [6], which is the product of a Gaussian random vector and a hidden scalar random variable (multiplier). By definition, a random vector $\vec{X}$ follows a GSM distribution if and only if it can be written as $\vec{X} \stackrel{d}{=} \sqrt{A}\,\vec{G}$, where $\vec{G}$ is a zero-mean Gaussian vector and $A$ a positive scalar variable independent of $\vec{G}$ ($\stackrel{d}{=}$ denotes equality in distribution). 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 a set 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*. In this paper, we extract the interdependencies between coefficients at different subbands and levels by utilizing their joint statistics, expressed by a GSM model in which the multiplier $A$ is drawn from a $S\alpha S$ distribution. This results in a specific model, namely the *$\alpha$-sub-Gaussian* ($\alpha$-SG) model.

As an example of a neighborhood, we can define the neighbors of a coefficient to be the $m \times m$ adjacent coefficients at the same subband, the J-1 coefficients of the other subbands (orientations) at the same spatial location plus the coefficient at the same subband of the next level placed at the corresponding location. Then, the vector representing this neighborhood is supposed to be a sample of an $\alpha$-sub-Gaussian process, which is a variance mixture of a Gaussian process, defined as follows [2]:

***Definition*:** Let $\{G(t),\, t \in T\}$ be a Gaussian process with covariance function R(u,v) and $A \sim S_{\alpha/2}((\cos\frac{\pi\alpha}{4})^{2/\alpha},\, 1,\, 0)$ be a positive $\frac{\alpha}{2}$-stable random variable where $\alpha < 2$. Assume that the random variable $A$ is independent of $\{G(t),\, t \in T\}$. The symmetric alpha-stable ($S\alpha S$) process $\{X(t) = A^{1/2}G(t),\, t \in T\}$ is a sub-Gaussian process with an underlying Gaussian process $\{G(t),\, t \in T\}$.

The finite dimensional projections, $(X(t_1), ..., X(t_d))$, $d \geq 1$, are $\alpha$-SG($\mathbf{R}$) random vectors with underlying covariance matrix $\mathbf{R}$. So, the vector $\vec{X}_k^l = [x_1, x_2, ..., x_N]$, representing the $N$-dimensional neighborhood of coefficient $c_k^l$ at level $l$, is supposed to be drawn of an $\alpha$-SG($\mathbf{R}^l$) process, where $\mathbf{R}^l$ denotes the underlying covariance matrix of the $l$th-level. The notion of covariance between two random variables plays an important role in the second-order moment theory. However, covariances do not exist for the family of $S\alpha S$ random variables, due to the lack of finite variance. Instead, a quantity called *covariation*, which under certain constraints plays an analogous role for $S\alpha S$ random variables to the one played by covariance for Gaussian random variables, has been proposed [2]. Specifically, let $X$ and $Y$ be jointly $S\alpha S$ random variables with $\alpha > 1$, zero location parameters and dispersions $\gamma_X$ and $\gamma_Y$ respectively. The covariation of $X$ with $Y$ is defined by

$$[X, Y]_\alpha = \frac{E\{XY^{<p-1>}\}}{E\{|Y|^p\}}\gamma_y \qquad (1)$$

where for any complex number $z$ and $a \geq 0$ we use the notation $z^{<a>} = |z|^{a-1}\bar{z}$, with $\bar{z}$ denoting complex conjugation.

Let the set of vectors $\{\vec{X}_k = [x_1^k, x_2^k, ..., x_N^k]^T\}_{k=1,...,K}$ represent independent realizations of an $\alpha$-SG($\mathbf{R}$) process. Then, by discretizing (1), we can estimate the covariation between two components $x_m,\ x_n$ of an $\alpha$-SG($\mathbf{R}$) vector, denoted as $\mathbf{c}_{mn} = [x_m,\ x_n]_\alpha$, as follows:

$$\hat{\mathbf{c}}_{mn} = (c(p,\alpha))^{-\frac{\alpha}{p}}\left[\frac{1}{K}\sum_{k=1}^{K} x_m^k (x_n^k)^{<p-1>}\right]\left[\frac{1}{K}\sum_{k=1}^{K}|x_n^k|^p\right]^{\frac{\alpha}{p}-1} \qquad (2)$$

where $c(p, \alpha)$ is a constant with $0 < p < \alpha$ (cf. [2]). We define the estimated covariation matrix $\hat{\mathbf{C}}$ as the matrix with elements $[\hat{\mathbf{C}}]_{mn} = \hat{\mathbf{c}}_{mn}$. The relation between the covariation matrix $\mathbf{C}$ of the sub-Gaussian vector and the underlying covariance matrix $\mathbf{R}$ of the Gaussian vector is [2]

$$[\mathbf{C}]_{mn} = 2^{-\frac{\alpha}{2}}[\mathbf{R}]_{mn}[\mathbf{R}]_{nn}^{(\alpha-2)/2} \qquad (3)$$

resulting in the following estimators of the covariances

$$[\hat{\mathbf{R}}]_{nn} = \left(2^{\frac{\alpha}{2}}[\hat{\mathbf{C}}]_{nn}\right)^{\frac{2}{\alpha}}, \quad [\hat{\mathbf{R}}]_{mn} = 2^{\frac{\alpha}{2}}\frac{[\hat{\mathbf{C}}]_{mn}}{[\hat{\mathbf{R}}]_{nn}^{(\alpha-2)/2}}, \qquad (4)$$

which are consistent and asymptotically normal.

### 2.2. Normalization of the sub-Gaussian model

An important property of a GSM model is that the density of $\vec{X}$ is Gaussian when conditioned on $A$, i.e., the normalized vector $\vec{X}/\sqrt{A}$ follows a Gaussian distribution:

$$p(\vec{X}|A) = \frac{\exp(-\vec{X}^T(A\mathbf{R})^{-1}\vec{X}/2)}{(2\pi)^{N/2}\ |A\mathbf{R}|^{1/2}}\ . \qquad (5)$$

From (5), it can be seen that the maximum likelihood estimator for the multiplier $A$ is

$$\hat{A}(\vec{X}) = \frac{\vec{X}^T\mathbf{R}^{-1}\vec{X}}{N}, \qquad (6)$$

where the estimator is viewed as a function of a neighborhood $\vec{X}$ to emphasize the assumption of locality. This simplifies the computational procedure, as we assume that the multipliers associated with different neighborhoods are estimated independently, even though the neighborhoods are overlapping.

Summarizing, the "Gaussianization" of a given decomposed image proceeds as follows. For each subband (except the low-pass residual):

i) Estimate the underlying covariance matrix $\mathbf{R}$ using (4).

ii) For each coefficient $c_k$:

- Construct the corresponding neighborhood $\vec{X}_k$.
- Estimate the associated multiplier $\hat{A}_k(\vec{X}_k)$ using (6).
- Compute the normalized coefficient $\tilde{c}_k = c_k/\sqrt{\hat{A}_k}$.

From (6), it is obvious that the estimation accuracy depends on the underlying covariance matrix and the neighborhood structure. We evaluated the performance of the above estimator in terms of the neighborhood size $N$ and the value of characteristic exponent $\alpha$, by running 500 Monte-Carlo simulations on a set of $K = 1000$ generated $\alpha$-SG($\mathbf{R}$) vectors with known covariance matrix ($\mathbf{R} = \sigma^2\mathbf{I}$, where $\mathbf{I}$ denotes the $N \times N$ identity matrix). We concluded that a neighborhood size $N$ between 11 and 15 is adequate to achieve an appropriately small mean squared estimation error.

## 3. IMAGE RETRIEVAL

### 3.1. Feature Extraction

Following the normalization procedure, the marginal and joint statistics of the coefficients at adjacent positions, orientations, and levels are close to the Gaussian distribution. In the FE step, we compute the $J \times J$ covariance matrix at each decomposition level. Thus, for a given image $I$ decomposed in $L$ levels, the corresponding signature $\mathcal{S}$ is given by the set of the $L$ covariance matrices:

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

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 $\frac{J(J+1)L}{2}$.

### 3.2. Similarity Measurement

After the Gaussianization procedure, we model the distribution of each decomposition level using a multivariate Gaussian density (MvGD). The similarity between two images is measured by employing the KLD. In particular, let $Q^l, I^l$ be the $l$-th decomposition level of two images $Q$ and $I$, following MvGDs with zero-mean vectors and covariance matrices $\mathbf{\Sigma}_{Q^l}$ and $\mathbf{\Sigma}_{I^l}$, respectively. The KLD between the two images in the $l$-th level is given by:

$$D(Q^l \| I^l) = \frac{1}{2} tr(\mathbf{\Sigma}_{Q^l}\mathbf{\Sigma}_{I^l}^{-1} - I) - \frac{1}{2}\ln|\mathbf{\Sigma}_{Q^l}\mathbf{\Sigma}_{I^l}^{-1}| \tag{7}$$

Making an assumption of independence between scales, the overall KLD between images $Q$, $I$ is given by the sum:

$$D(Q\|I) = \sum_{l=1}^{L} D(Q^l\|I^l) \tag{8}$$

Considering databases that may contain rotated versions of a given image, the following relation of equivalence exists [4]:

$$\mathbf{\Sigma}_{I_\theta^l} = \mathbf{F}(\theta)\mathbf{\Sigma}_{I^l}\mathbf{F}^T(\theta), \tag{9}$$

where $\mathbf{\Sigma}_{I^l}$, $\mathbf{\Sigma}_{I_\theta^l}$ are the $l$-level covariance matrices of image $I$ and its rotated version at an angle $\theta$, $I_\theta$, respectively, and $\mathbf{F}(\theta)$ is an appropriate steering matrix. Consider $Q$ to be the query image and $\tilde{I} = I_\phi$ to be a counter-clockwise rotation, by an angle $\phi$, of the original image $I$ in the database. Obviously, in actual applications the value of $\phi$ is unknown. Thus, the distance between the $l$th-levels of $Q$ and $\tilde{I}$ ($Q^l$ and $\tilde{I}^l$, respectively) is defined as the minimum KLD between $Q^l$ and $\tilde{I}_{-\theta}^l$, where the minimization is over a set of rotations $\Theta$. By noticing that $\mathbf{\Sigma}_{\tilde{I}_{-\theta}^l} = \mathbf{F}(-\theta)\mathbf{\Sigma}_{\tilde{I}^l}\mathbf{F}^T(-\theta)$ and substituting (9) into (7), we get that the KLD between $Q^l$ and $\tilde{I}_{-\theta}$ is

$$\begin{aligned} D(Q^l\|\tilde{I}_{-\theta}^l) &= \frac{1}{2} tr\big(\mathbf{\Sigma}_{Q^l}\mathbf{F}^T(\theta)\mathbf{\Sigma}_{\tilde{I}^l}^{-1}\mathbf{F}(\theta) - I\big) - \\ &- \frac{1}{2}\ln(|\mathbf{\Sigma}_{Q^l}||\mathbf{\Sigma}_{\tilde{I}^l}^{-1}|). \end{aligned} \tag{10}$$

Finally, the overall KLD between $Q$ and $\tilde{I}$ is defined as:

$$\begin{aligned} D(Q\|\tilde{I}) &= \min_{\theta\in\Theta} \sum_{l=1}^{L} D(Q^l\|\tilde{I}_{-\theta}^l) \\ &= \min_{\theta\in\Theta}\Big[\frac{1}{2}\sum_{l=1}^{L} tr\big(\mathbf{\Sigma}_{Q^l}\mathbf{F}^T(\theta)\mathbf{\Sigma}_{\tilde{I}^l}^{-1}\mathbf{F}(\theta)\big)\Big] - \frac{JL}{2} - \\ &- \frac{1}{2}\sum_{l=1}^{L}\ln(|\mathbf{\Sigma}_{Q^l}||\mathbf{\Sigma}_{\tilde{I}^l}^{-1}|). \end{aligned} \tag{11}$$

Note that in the above analysis, we have assumed equispaced basic orientations, which results in orthogonal matrices $\mathbf{F}(\theta)$, i.e., $\mathbf{F}^T(\theta) = \mathbf{F}^{-1}(\theta) = \mathbf{F}(-\theta)$.

## 4. EXPERIMENTS AND CONCLUSIONS

In order to evaluate its efficiency, the proposed retrieval scheme was applied on a set of 14, $512 \times 512$ texture images, obtained from the USC SIPI database[1]. We divided each image into 4 ($256 \times 256$) non-overlapping subimages. Then, each subimage was associated with 4 physically rotated versions at $30^o$, $60^o$, $90^o$, and $120^o$ degrees, resulting in a database containing 280 images. We implemented a 3-level steerable pyramid decomposition, by employing the following oriented basis (steering) functions:

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

$$f_3(\theta) = f_1(\frac{\pi}{2} - \theta) \qquad f_4(\theta) = f_1(\frac{3\pi}{4} - \theta)$$

with basic angles $\phi_1 = 0$, $\phi_2 = \pi/4$, $\phi_3 = \pi/2$, $\phi_4 = 3\pi/4$, resulting in 4 oriented subbands at each level. Since the steering functions have only odd harmonics which oscillate at some finite speed, the number of local extrema of (11) (as a function of theta) 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 extrema is lower bounded making it possible to search for them in a few non-overlapping angular intervals [7].

The Gaussianization process for a coefficient $c_k$ is cur-

[1] http://sipi.usc.edu/services/database

| Computational Complexity | | | |
|---|---|---|---|
| Method | Gaussianization step | FE step | SM step |
| **1.** | - | $M^2J(J+1)(1-4^{-L})$ | $KL(6J^3-2J^2+J+1)$ |
| **2.** | $LJ\cdot\mathcal{O}(N^3)+JM^2[8N^2+N+1](1-4^{-L})/3$ | $M^2J(J+1)(1-4^{-L})$ | $KL(6J^3-2J^2+J+1)$ |
| **3.** | $LJ\cdot\mathcal{O}(N^3)+JM^2[8N^2+N+1](1-4^{-L})/3$ | $M^2J(J+1)(1-4^{-L})$ | $KL[\mathcal{O}(J^3)+6J^3+3J^2+J+1]+$ $+L[\mathcal{O}(J^4)+3]$ |

**Table 1**. Computational complexity of the retrieval schemes.

| Methods | | |
|---|---|---|
| **1.** Non-Gaussianized + Frobenius | **2.** Gaussianized + Frobenius | **3.** Gaussianized + KLD |
| 88.62 | 89.26 | 94.01 |

**Table 2**. Average retrieval rate (%) in the top 16 matches.

ried out by forming its neighborhood consisting of the $3\times 3$ adjacent coefficients at the same subband, the 3 coefficients of the rest of the subbands at the same level, placed at the same position and the 1 coefficient of the same subband at the next level, placed at the corresponding location according to a quad-tree structure.

In the following, we compare the performance of our proposed rotation invariant KLD and Gaussianization approach (denoted as "method 3"), with the performances obtained by minimizing the Frobenius norm of the differences between the corresponding covariance matrices [4] with and without Gaussianization (methods 2 and 1, respectively). The relevant images for each query are defined as the other 15 subimages obtained from the same original image.

Table 1 presents a rough estimation of the computational complexity of the above three methods, where $M\times M$ is the dimension of the original image, $N$ is equal to the neighborhood size and $K$ is the dimension of the discrete grid of angles, $\Theta$, for the minimization of the overall Frobenius and KLD similarity measures. In the above implementation, we discretized the $[0,\pi]$ interval using a step equal to $1^o$. From this table, it is clear that the main computational cost is due to the Gaussianization step, while the complexity of the two similarity measures is similar, since in real applications the values of $L$ and $J$ are small (usually 3 or 4). The computational complexity can be reduced by employing an efficient implementation of the Gaussianization process, as well as a fast algorithm for the minimization of the similarity function. Table 2 shows the comparison in performance in average percentages of retrieving relevant images in the top 16 matches. Figure 1 depicts the average percentages of retrieving relevant subimages as a function of the number of top matches. The performance improvement achieved by our proposed method is evident.

As a conclusion, we observe that the implementation of a Gaussianization procedure on the original coefficients of a multi-orientation, multi-scale steerable pyramid decomposition of the images in a given database, combined with the application of a rotation invariant version of the Kullback-Leibler divergence as a measure of similarity, results in a decreased probability of retrieval error. The key feature of the proposed scheme is that the Gaussianization process considers a heavy-tails modeling of the original subband coefficients, resulting in an improved performance.

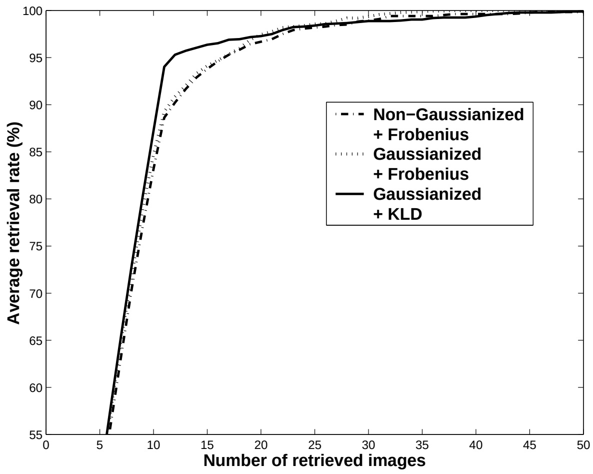


**Fig. 1**. Retrieval performance according to the number of top matches considered.

## 5. REFERENCES


[1] G. Tzagkarakis and P. Tsakalides, "A statistical approach to texture image retrieval via alpha-stable modeling of wavelet decompositions," *Proc. 5th WIAMIS*, April 21-23 2004.

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

[3] 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.

[4] B. Beferull-Lozano, H. Xie, and A. Ortega, "Rotation-invariant features based on steerable transforms with an application to distributed image classification," in *Proc. of IEEE Int. Conf. on Image Proc.*, (Barcelona, Spain), Sept. 2003.

[5] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Adv. Neural Information Processing Systems (NIPS*99)*, vol. 12, (Cambridge, MA), pp. 855–861, MIT Press, 2000.

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

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