Wavelet-based Texture Retrieval using Generalized Gaussian Density and Kullback-Leibler Distance

Minh N. Do, *Student Member, IEEE*, and Martin Vetterli, *Fellow, IEEE*

***Abstract*—We present a statistical view of the texture retrieval problem by combining the two related tasks, namely feature extraction (FE) and similarity measurement (SM), into a joint modeling and classification scheme. We show that using a consistent estimator of texture model parameters for the FE step followed by computing the Kullback-Leibler distance (KLD) between estimated models for the SM step is asymptotically optimal in term of retrieval error probability. The statistical scheme leads to a new wavelet-based texture retrieval method that is based on the accurate modeling of the marginal distribution of wavelet coefficients using generalized Gaussian density (GGD) and on the existence a closed form for the KLD between GGD's. The proposed method provides greater accuracy and flexibility in capturing texture information, while its simplified form has a close resemblance with the existing methods which uses energy distribution in the frequency domain to identify textures. Experimental results on a database of 640 texture images indicate that the new method significantly improves retrieval rates, e.g. from 65% to 77%, compared with traditional approaches, while it retains comparable levels of computational complexity.**

***Index Terms*—Content-based image retrieval, similarity measurement, texture characterization, texture retrieval, wavelets, statistical modeling, generalized Gaussian density, Kullback-Leibler distance.**

I. Introduction

A. Motivation

DIGITAL image libraries are becoming more widely used as more visual information is put in digital form as well as on-line. To improve human access, however, there must be an effective and precise method for users to search, browse, and interact with these collections and to do so in a timely manner. As a result, content-based image retrieval (CBIR) from unannotated image databases has been a fast growing research area recently: see [1] for a recent extensive review on the subject.

We consider a simple architecture of a typical CBIR system (Figure 1), where there are two major tasks. The first one is *Feature Extraction* (FE), where a set of features, called image signatures, is generated to accurately represent the content of each image in the database. A signature is much smaller in size than the original image, typically on the order of hundreds of elements (rather than millions). The second task is *Similarity Measurement* (SM), where a distance between the query image and each image in the database using their signatures is computed so that the top N "closest" images can be retrieved. Typically, the features used in CBIR systems are low-level image features such as color, texture, shape and layout. In this work, we focus on the use of texture information for image retrieval.

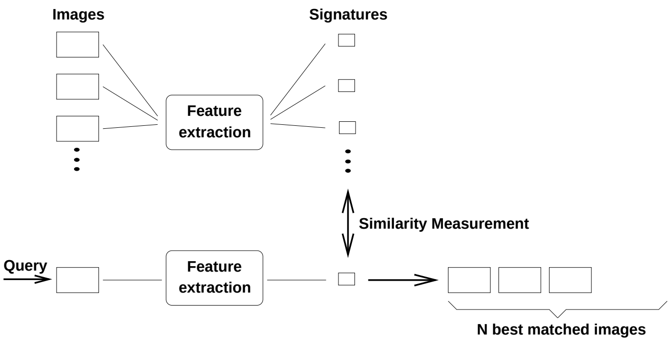


Fig. 1. Image retrieval system architecture

Some of the most popular texture extraction methods for retrieval are based on filtering or wavelet-like approaches [2], [3], [4], [5], [6], [7], [8]. Essentially, these methods measure energy (possibly weighted) at the output of filter banks as extracted features for texture discrimination. The basic assumption of these approaches is that the energy distribution in the frequency domain identifies a texture. Besides providing acceptable retrieval performance from large texture databases, those approaches are partly supported by physiological studies of the visual cortex [9], [10]. Furthermore, as wavelets are a core technology in the next generation of still image coding format, JPEG-2000 [11], the choice of wavelet features enables the implementation of retrieval systems that can work directly in the compressed domain. Other possible transforms are wavelet packets, wavelet frames and Gabor wavelet transforms.

Finding good similarity measures between images based on some feature set is a challenging task. On the one hand, the ultimate goal is to define similarity functions that match with human perception, but how humans judge the similarity between images is a topic of ongoing research. Perceptual studies [12], [13] identified texture dimensions by conducting experiments that asked observers to group textures according to perceived similarity. The detected perceptual criteria and rules for similarity judgment from this type of subjective experiments can be used in building image retrieval system [14]. On the other hand, many current retrieval systems take a simple approach by using typically norm-based distances (e.g. Euclidean distance) on the extracted feature set as a similarity function [1]. The main premise behind these CBIR systems is that

M. N. Do is with the Audio-Visual Communications Laboratory, Department of Communication Systems, Swiss Federal Institute of Technology, Lausanne, Switzerland (e-mail: Minh.Do@epfl.ch).

M. Vetterli is with the Audio-Visual Communications Laboratory, Department of Communication Systems, Swiss Federal Institute of Technology, Lausanne, Switzerland, and with the Department of Electrical Engineering and Computer Science, University of California, Berkeley CA 94720 (e-mail: Martin.Vetterli@epfl.ch).

This work was supported by a PhD Fellowship from the Department of Communication Systems, EPFL, and the Swiss National Science Foundation under Grant 21-52439.97.

given a "good set" of features extracted from the images in the database (the ones that significantly capture the content of images) then for two images to be "similar" their extracted features have to be "close" to each other. Therefore, any reasonable similarity functions defined on the features space should perform well. Sometimes, weighting factors are necessary to normalize extracted features over the entire database to comparable ranges so that they have approximately the same influence on the overall distance.

Note that this "global" normalization process is *different* with the one often used in classification problems where the normalized factors are computed using a training set of feature vectors from each class. Furthermore, the commonly used inverse variance weighted Euclidean distance as in CBIR [15] is questionable in the case of a feature component that has small global variance, thus leading to a large weight in the overall distance. By contrast, it can be argued that a small variation component should have little discrimination power and should thus carry a small weight in the overall distance.

B. Our Approach and Related Works

In this work we consider jointly the problems of FE and SM in texture retrieval using a statistical approach. Our point is that, given only a low-level representation, statistical modeling provides a natural mean to formulate the retrieval problem, as is typically done in pattern recognition. Considering the two related retrieval tasks FE and SM as estimation and detection problems, respectively, provides us with a justified way of defining similarity functions on the feature space. The implication of this approach is twofold. First, it provides a confidence on the optimality of the defined similarity function under some explicit assumptions. Secondly, as we will see, this approach provides a common ground for many existing similarity functions by simply modifying the underlying assumptions.

Statistical modeling has been used in CBIR systems before. Perhaps the most well-known examples are the use of histograms to capture the distribution of image features such as color [16]. Wouwer et al. [7] employed generalized Gaussian density functions to represent texture images in the wavelet domain. The model parameters are estimated using a method of moment matching, and the similarity function is again defined as weighted Euclidean distances on extracted model parameters. Independently of our work, Vasconcelos and Lippman [17] recently took a similar approach where they introduced a probabilistic formulation of the CBIR problem as a common ground for several currently used similarity functions.

As an important case of CBIR, we demonstrate in this work the application of the statistical framework in the wavelet-based texture retrieval problem. The statistical approach fits nicely into this case, since a texture image is often regarded as a realization of an underlying stochastic process. In the end, we will briefly discuss how such approach can be applied to other features and integrated into more general image retrieval systems.

The outline of this paper is as follows. In the next section, we set up the CBIR problem in a general statistical framework. In Section III, we apply this to wavelet-based texture retrieval where the statistical approach provides a justified way of defining new similarity function which has certain optimality properties and can be simplified to explain other common used metrics. In Section IV, experimental results on a large texture image database indicate the significant improvement in retrieval rate using the new approach. Section V concludes with some discussions.

II. Content-based Image Retrieval in a Statistical Framework

A. General Setting

The problem of searching for the top N images similar to a given query image from a database of total M images ($N \ll M$) can be formulated as a multiple hypotheses problem.[1] The query image $\mathcal{I}_q$ is represented by its data set $\boldsymbol{x} = (x_1, x_2, \ldots, x_L)$, which is typically obtained after a pre-processing stage. Each candidate image in the database $\mathcal{I}_i : i = 1, 2, \ldots, M$ is assigned with a hypothesis $\mathcal{H}_i$. The goal is to select among the M possible hypotheses the N best ones (with a ranking order) that describe the data $\boldsymbol{x}$ from the query image.

To select the N top matches from those M hypotheses we can use the multiple hypotheses testing argument recursively. That is, we first choose the best one among the M possible hypotheses $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_M\}$, and then we choose the next best one among the remain $(M-1)$ hypotheses, and keep doing so for N times. Under the common assumption that all prior probabilities of the hypotheses are equal, it can be shown [18] that, for each recursive step the optimum rule (with the minimum probability of error criterion) is to choose the hypothesis with the highest likelihood among the possible ones. Thus for CBIR, it is optimal to select N hypotheses with highest likelihood, i.e. $\mathcal{H}_{k_1}, \mathcal{H}_{k_2}, \ldots, \mathcal{H}_{k_N}$ where

$$p(\boldsymbol{x}|\mathcal{H}_{k_1}) \geq p(\boldsymbol{x}|\mathcal{H}_{k_2}) \geq \ldots \geq p(\boldsymbol{x}|\mathcal{H}_{k_N}) \geq p(\boldsymbol{x}|\mathcal{H}_i) \qquad i \neq k_j \ (j = 1, 2, \ldots, N). \quad (1)$$

This is referred to as the *maximum likelihood* (ML) selection rule. The problem with (1) is that it requires M computational steps with a typically large data set $\boldsymbol{x}$. This turns out to be impractical in CBIR applications since this operation has to be done *on-line* in the interactive mode. Therefore, we need to find an approximation with much less computational cost.

In the *parametric* approach, the conditional probability density $p(X|\mathcal{H}_i)$ is modeled by a member of a family of probability density functions (PDF's), denoted by $p(X; \boldsymbol{\theta}_i)$ where $\boldsymbol{\theta}_i$ is a set of model parameters. With this setting, the extracted features for the image $\mathcal{I}_i$ is the estimated model parameter $\hat{\boldsymbol{\theta}}_i$, which is computed in the FE step. We denote the space of model parameters as $\boldsymbol{\Theta}$.

[1] However the term "hypotheses" is used here in a loose sense. Since in CBIR applications, the search is not for the *exact match* but rather for *most similar* ones, hence we can allow for more than one hypothesis to be valid.

Consider the query data $\boldsymbol{x} = (x_1, x_2, \ldots, x_L)$ as an independent and identically distributed (i.i.d.) sequence from the model $p(X; \boldsymbol{\theta}_q)$ of the query image. Then for large L, using the weak law of large number, the ML selection rule (1) is equivalent to maximizing

$$\begin{aligned}\frac{1}{L}\log p(\boldsymbol{x};\boldsymbol{\theta}_i) &= \frac{1}{L}\sum_{j=1}^{L}\log p(x_j;\boldsymbol{\theta}_i) \\ &\xrightarrow{L\to\infty} E_{\boldsymbol{\theta}_q}\left[\log p(X;\boldsymbol{\theta}_i)\right] \\ &= \int p(x;\boldsymbol{\theta}_q)\log p(x;\boldsymbol{\theta}_i)dx.\end{aligned}$$

This can be seen as equivalent to minimizing the *Kullback-Leibler distance* (KLD) or the *relative entropy* [19] between the two PDF's $p(x;\boldsymbol{\theta}_q)$ and $p(x;\boldsymbol{\theta}_i)$

$$D(p(X;\boldsymbol{\theta}_q)||p(X;\boldsymbol{\theta}_i)) = \int p(x;\boldsymbol{\theta}_q)\log\frac{p(x;\boldsymbol{\theta}_q)}{p(x;\boldsymbol{\theta}_i)}dx \qquad (2)$$

Under the same asymptotic condition (L is large), if the FE step uses a *consistent* estimator, which ensures the estimated parameter $\hat{\boldsymbol{\theta}}$ converges to the true parameter $\boldsymbol{\theta}$, then the distance (2) can be computed using the *estimated* model parameters $\hat{\boldsymbol{\theta}}_q$ and $\hat{\boldsymbol{\theta}}_i$. For such consistent estimator, we could employ the ML estimator [20], which means that for the query image, it computes

$$\hat{\boldsymbol{\theta}}_q = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \log p(\boldsymbol{x};\boldsymbol{\theta}). \qquad (3)$$

In summary, by combining FE and SM into a joint modeling and classification scheme, the optimum ML selection rule can be *asymptotically* realized (as the data sets for each image become large) by:

<u>*Feature Extraction:*</u> Given the data from each image, extracting features as estimated model parameters using a consistent estimator such as the ML estimator.

<u>*Similarity Measurement:*</u> To select the top N matches to a query, the images in the database are ranked based on the KLD's between the estimated model for the query and estimated models for each image.

The advantage of this scheme is that the SM step can be computed entirely on the estimated model parameters, which are typically small size, so that it can meet the timing constraint of the CBIR application. The method is generic as it allows the use of any feature data and statistical models for indexed images. Such image models can incorporate the knowledge from perceptual studies to closely match human judgment.

We point out that the Kullback-Leibler distance has been used in comparing images (e.g. in [21], [22]) but, to our knowledge, its use has not yet been sufficiently justified in the context of the image retrieval problem by jointly considering the two related tasks FE and SM.

To combine the KLD's from multiple data sets, such as from different channels or feature sets, we can use the *chain rule* [19] which states that the KLD between two joint PDF's $p(X,Y)$ and $q(X,Y)$ is

$$\begin{aligned}D(p(X,Y)||q(X,Y)) = D(p(X)||q(X)) + \\ D(p(Y|X)||q(Y|X)). \quad (4)\end{aligned}$$

Especially, when data is considered to be independent between combined sets then the joint KLD is simply the sum of KLD's from each set. Finally, the convexity property of KLD permits it to be used in an efficient search scheme using multiscale representations [21].

B. Relation to Histogram Methods

Histograms have been used since the early days of image retrieval, especially for representing color features [16], as well as for texture or local geometric properties [1]. In this section, we demonstrate that the histogram method can be interpreted through our statistical approach by using an appropriate model setup.

Let us partition the range of image data into R disjoint intervals of equal length, $\{S_1, S_2, \ldots, S_R\}$. Now consider the family of piecewise constant densities, defined as

$$p(x;\theta) = p_i \quad \text{for } x \in S_i, \quad i = 1, \ldots, R. \qquad (5)$$

Here the set of model parameters is $\boldsymbol{\theta} := (p_1, \ldots, p_R)$, where $p_i \geq 0$ ($i = 1, \ldots, R$) and $\sum_{i=1}^{R} p_i = 1$.

Given a sequence of i.i.d. data samples from an image, $\boldsymbol{x} = (x_1, \ldots, x_L)$, using the ML estimator in our statistical retrieval the FE step amounts to computing the feature vector $\hat{\boldsymbol{\theta}}$ where,

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i=1}^{L}\log p(x_i;\boldsymbol{\theta}) \\ &= \arg\max_{\sum_{k=1}^{R} p_k = 1,\, p_k\geq 0} \sum_{k=1}^{R} n_k \log p_k. \qquad (6)\end{aligned}$$

Here we denote n_k the number of data samples in $(x_1, \ldots, x_L)$ that belong to the region S_k. Solving (6) using the Lagrange multiplier gives,

$$\hat{\boldsymbol{\theta}} := \boldsymbol{p} \quad \text{with } p_k = n_k/L, \quad k = 1, \ldots, R. \qquad (7)$$

So the extracted feature vector $\boldsymbol{p}$ is in fact the normalized histogram of the image data $\boldsymbol{x}$. When searching for similar images given the query data $\boldsymbol{x}^{(q)}$ and its features $\boldsymbol{p}^{(q)}$, the following (discrete) KLD between the query image and each candidate image $\mathcal{I}_i$ can be used to rank the images in the database

$$\begin{aligned}D(\boldsymbol{p}^{(q)}||\boldsymbol{p}^{(i)}) &= \sum_{k=1}^{R} p_k^{(q)} \log\frac{p_k^{(q)}}{p_k^{(i)}} \\ &= \sum_{k=1}^{R} p_k^{(q)}\log p_k^{(q)} - \sum_{k=1}^{R} p_k^{(q)}\log p_k^{(i)}. \quad (8)\end{aligned}$$

Whereas the ML selection rule is based on the following log-likelihoods

$$\begin{aligned} \log p(\boldsymbol{x}^{(q)}|\mathcal{H}_i) &= \sum_{j=1}^{L} \log[p(x_j^{(q)};\boldsymbol{p}^{(i)})] \\ &= \sum_{k=1}^{R} n_k^{(q)} \log p_k^{(i)} \\ &= L^{(q)} \sum_{k=1}^{R} p_k^{(q)} \log p_k^{(i)}. \qquad (9) \end{aligned}$$

Comparing (8) with (9) indicates that in this case, the ranking based on the KLD is *exactly* (rather than asymptotically) the same with the optimum ML selection rule.

A drawback of the histogram method is that it requires a large number of extracted features, typical several hundreds histogram bins, to capture accurately image information. Thus it leads to impractical complexity in both storage of image indices and retrieval timing. In the next section, we employ the wavelet transform and the generalized Gaussian density to efficiently solve the texture retrieval problem within our statistical framework.

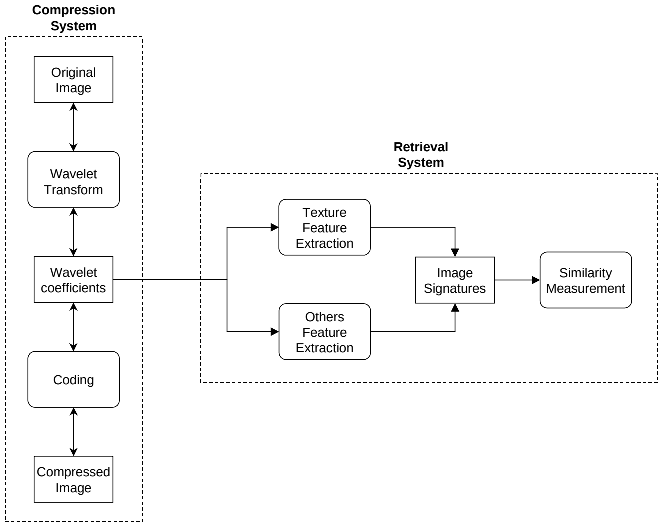


Fig. 2. Block diagram of the wavelet-based texture retrieval system in conjunction with a compression system. Feature extraction tasks could be done concurrently and efficiently either when an image is compressed and inserted into the database or when an input compressed image is decoded.

III. Wavelet Coefficients Modeled using Generalized Gaussian Density

A. Wavelet Representation

Statistical modeling is much easier if some preprocessing is carried out on the input images. Typical preprocessing is done via transformation of image pixel values into a suitable space where simple models with a small number of parameters can describe the data. Wavelets have recently emerged as an effective tool to analyze texture information as they provide a natural partition of the image spectrum into multiscale and oriented subbands via efficient transforms [2], [3], [4], [5], [6], [7], [8]. Furthermore, since wavelets are used in major future image compression standards [11] and are also shown to be prominent in searching for images based on color and shape [23], [24], a wavelet-based texture retrieval system can be used effectively in conjunction with a compression system and retrieval systems using other image features (see Figure 2).

Using the assumption that the energy distribution in frequency domain identifies texture, traditional approaches computed energies of wavelet subband as texture features. Commonly, $\mathbf{L}^1$ and $\mathbf{L}^2$ norms are used as measures.[2] More specifically, given the wavelet coefficients $x_{i,1}, x_{i,2}, \ldots, x_{i,L}$ at the i-th subband, typically the following two values are used as features:

$$f_i^{(1)} = \frac{1}{L}\sum_{j=1}^{L}|x_{i,j}|, \text{ and} \qquad (10)$$

$$f_i^{(2)} = \left(\frac{1}{L}\sum_{j=1}^{L}x_{i,j}^2\right)^{1/2}. \qquad (11)$$

[2] This is an abuse of terminology since strictly speaking $\mathbf{L}^1$ norm is not an energy function. Sometimes it is chosen due to its simplicity. Results from several studies indicate no general conclusion in favor of a particular measure.

On the other hand, statistical approaches treat texture analysis as a probability inference problem (e.g. see [25]). A natural extension of the energy method is to model a texture by the marginal densities of wavelet subband coefficients. This is justified by recent psychological research on human texture perception which suggests that two homogeneous textures are often difficult to discriminate if they produce similar marginal distributions of responses from a bank of filters [26]. In fact, Heeger and Bergen [27] successfully synthesized many natural looking texture images by matching the histograms of filter responses from a wavelet-liked transform. More accurate texture models can be obtained via a *fusion* of marginal distributions using minimax entropy principles [25] or by taking into account the joint statistics of wavelet coefficients across subbands [28]. However, considering complexity as a major constraint in the image retrieval application, in this work we simply characterize texture images via marginal distributions of their wavelet subband coefficients. Still, this representation of texture is more precise than the ones that use wavelet subband energies alone. But more importantly, the statistical modeling leads to a more justifiable way of defining similarity functions between images.

B. Generalized Gaussian Density Modeling of Wavelet Coefficients

Experiments show that a good PDF approximation for the marginal density of coefficients at a particular subband produced by various type of wavelet transforms may be achieved by adaptively varying two parameters of the *generalized Gaussian* density (GGD) [29], [30], [7], [31], which is defined as:

$$p(x;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)}e^{-(|x|/\alpha)^\beta}, \qquad (12)$$

where $\Gamma(.)$ is the Gamma function, i.e. $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$, $z > 0$.

Here α models the width of the PDF peak (standard deviation), while β is inversely proportional to the decreasing rate of the peak. Sometimes, α is referred to as the *scale* parameter while β is called the *shape* parameter. The GGD model contains the Gaussian and Laplacian PDF's as special cases, using $\beta = 2$ and $\beta = 1$, respectively.

Within a CBIR statistical framework, the desired estimator in our case is the maximum-likelihood (ML) estimator. Furthermore, in [32] evaluation of accuracy of estimates for both large and small samples for GGD models among classic statistical methods shows that the ML estimator is significantly superior for heavy-tailed distribution (which is often the case for subband coefficients). We now describe the ML estimator for GGD.

Let us define the likelihood function of the sample $\boldsymbol{x} = (x_1, \ldots, x_L)$ having independent component as

$$L(\boldsymbol{x}; \alpha, \beta) = \log \prod_{i=1}^{L} p(x_i; \alpha, \beta),$$

where α and β are parameters to be estimated. It was shown in [32] that in this case the following likelihood equations have a unique root in probability, which is indeed the maximum-likelihood estimator

$$\frac{\partial L(\boldsymbol{x}; \alpha, \beta)}{\partial \alpha} = -\frac{L}{\alpha} + \sum_{i=1}^{L} \frac{\beta \, |x_i|^\beta \alpha^{-\beta}}{\alpha} = 0 \qquad (13)$$

$$\frac{\partial L(\boldsymbol{x}; \alpha, \beta)}{\partial \beta} = \frac{L}{\beta} + \frac{L\Psi(1/\beta)}{\beta^2} - \sum_{i=1}^{L} \left(\frac{|x_i|}{\alpha} \right)^\beta \log \left(\frac{|x_i|}{\alpha} \right) = 0, \quad (14)$$

here $\Psi(.)$ is the digamma function [33], i.e. $\Psi(z) = \Gamma'(z)/\Gamma(z)$.

Fix $\beta > 0$ then (13) has a unique, real, and positive solution as:

$$\hat{\alpha} = \left(\frac{\beta}{L} \sum_{i=1}^{L} |x_i|^\beta \right)^{1/\beta}. \qquad (15)$$

Substitute this into (14), the shape parameter β is the solution of the following *transcendental* equation:

$$1 + \frac{\Psi(1/\hat{\beta})}{\hat{\beta}} - \frac{\sum_{i=1}^{L} |x_i|^{\hat{\beta}} \log |x_i|}{\sum |x_i|^{\hat{\beta}}} + \frac{\log \left(\frac{\hat{\beta}}{L} \sum_{i=1}^{L} |x_i|^{\hat{\beta}} \right)}{\hat{\beta}} = 0, \quad (16)$$

which can be solved numerically. We propose an effective determination of $\hat{\beta}$ using the Newton-Raphson iterative procedure [20] with the initial guess from the moment method described in [30]. This algorithm is detailed in Appendix A. Experiments show that typically only around three iteration steps are required to obtain solutions with an accuracy of the order of 10^{-6}.

Figure 3 shows a typical example of a histogram of wavelet subband coefficients together with a plot of the fitted GGD using the ML estimator. The fits are generally quite good. As a result, with only two parameters for the GGD, we can accurately capture the marginal distribution of wavelet coefficients in a subband that otherwise would require hundreds of parameters by using histogram. This significantly reduces the storage of the image features, as well as the computational complexity in similarity measurement.

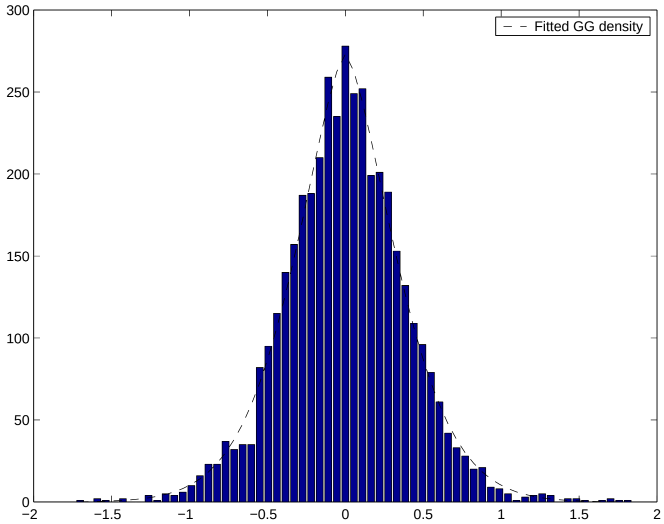


Fig. 3. Wavelet subband coefficient histogram fitted with a generalized Gaussian density. Example for the Fabric15 subimage of size 128 × 128 at the highest horizontal frequency scale. The estimated parameters are: $\alpha = 0.46$ and $\beta = 1.52$.

C. Similarity Measurement between GGD's

Given the GGD model, the PDF of wavelet coefficients in each subband can be completely defined via two parameters α and β. Substitute (12) into (2) and after some manipulations we obtain the following closed form for the Kullback-Leibler distance (KLD) between two GGD's as:

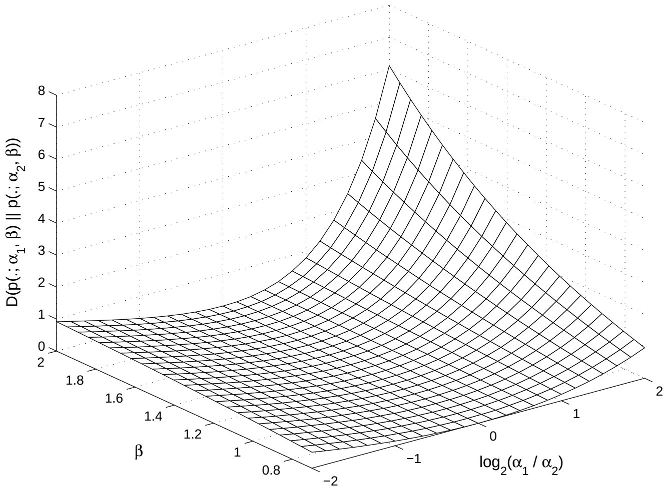
$$D(p(.; \alpha_1, \beta_1) || p(.; \alpha_2, \beta_2)) = \log \left(\frac{\beta_1 \alpha_2 \Gamma(1/\beta_2)}{\beta_2 \alpha_1 \Gamma(1/\beta_1)} \right) + \left(\frac{\alpha_1}{\alpha_2} \right)^{\beta_2} \frac{\Gamma((\beta_2 + 1)/\beta_1)}{\Gamma(1/\beta_1)} - \frac{1}{\beta_1}. \quad (17)$$

Therefore, the similarity measurement between two wavelet subbands can be computed very effectively using the model parameters. Furthermore, applying (4) with the reasonable assumption that wavelet coefficients in different subbands are independent, the overall similarity distance between two images is precisely the sum of KLD's given in (17) between corresponding pairs of subbands. That is, if we denote $\alpha_i^{(j)}$ and $\beta_i^{(j)}$ as the extracted texture features from the wavelet subband j of the image $\mathcal{I}_i$ then the overall distance between two images $\mathcal{I}_1$ and $\mathcal{I}_2$ (where $\mathcal{I}_1$ is the query image) is the sum of all the distances across all wavelet subbands:

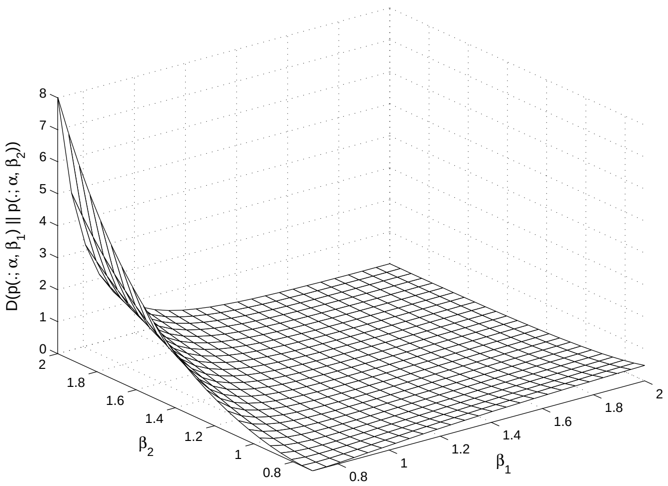
$$D(\mathcal{I}_1,\mathcal{I}_2)=\sum_{j=1}^{B} D(p(.;\alpha_1^{(j)},\beta_1^{(j)})||p(.;\alpha_2^{(j)},\beta_2^{(j)})), \quad (18)$$

where B is the number of analyzed subbands. Thus the KLD theory provides us with a justified way of combining distances into an overall similarity measurement, and no normalization on the extracted features is needed.

The distance function defined in (17) is a function of three variables: the ratio of two scales α_1/α_2 and two shape parameters β_1 and β_2. Figure 4 plots the distance function when the two distributions have the same shape parameter and when they have the same scale parameter. The chosen ranges for the ratio α_1/α_2 as $[0.25, 4]$ comes from the fact that we are only interested in the two relatively close distributions when searching for most similar images. The selected range for β as $[0.7, 2.0]$ is based on the experimental results (refer to Section IV-C)



(a) When $\beta_1=\beta_2=\beta$.



(b) When $\alpha_1=\alpha_2=\alpha$.

Fig. 4. Kullback-Leibler distance between two generalized Gaussian densities given in (17).

D. A Special Case of GGD

To gain more insight on the similarity function (17), let us consider the special case of a GGD where the shape parameter β is fixed. That means we model the wavelet coefficients using the following single parameter α distribution family.

$$\left\{p_\beta(.;\alpha),\alpha\in\mathbb{R}_+ \ : \ p_\beta(x;\alpha)=\frac{\beta}{2\alpha\Gamma(1/\beta)}e^{-(|x|/\alpha)^\beta}\right\}. \quad (19)$$

This simplified model is often used in practical image processing problems like denoising [31]. From the sample data sequence $\boldsymbol{x}=\{x_1,x_2,\dots,x_L\}$, the extracted feature is just the estimated parameter α as given in (15)

$$\hat{\alpha}_{\boldsymbol{x}}=\left(\frac{\beta}{L}\sum_{i=1}^{L}|x_i|^\beta\right)^{1/\beta}. \quad (20)$$

The KLD between two PDF's from the family given in (19) is:

$$D(p_\beta(.;\alpha_1)||p_\beta(.;\alpha_2))=\log\left(\frac{\alpha_2}{\alpha_1}\right)+\left(\frac{\alpha_1}{\alpha_2}\right)^\beta\frac{1}{\beta}-\frac{1}{\beta}. \quad (21)$$

Next, consider again the optimum ML selection rule. At a particular wavelet subband, denote $\boldsymbol{x}=\{x_1,x_2,\dots,x_L\}$ the wavelet coefficients of the query image and $p_\beta(x;\alpha_k)$, $k=1,2,\dots,M$ as the estimated marginal distributions for each candidate images in the database. The ranking procedure should be based on the following normalized log-likelihood values, which can be simplified as

$$\begin{aligned} l_k(\boldsymbol{x}) &= \frac{1}{L}\sum_{i=1}^{L}\log p_\beta(x_i;\alpha_k) \\ &= \log\beta-\log 2-\log\alpha_k-\log\Gamma(1/\beta)-\frac{\sum_{i=1}^{L}|x_i|^\beta}{L\alpha_k^\beta} \\ &= -\log\alpha_k-\left(\frac{\hat{\alpha}_{\boldsymbol{x}}}{\alpha_k}\right)^\beta\frac{1}{\beta}+C_\beta, \end{aligned}$$

where $\hat{\alpha}_{\boldsymbol{x}}$ is the extracted feature from the query data given in (20) and C_β is a constant only dependent on β.

The KLD's between the query model parameter and a candidate image model parameter is

$$\begin{aligned} d_k(\boldsymbol{x}) &= D\left(p_\beta(.|\hat{\alpha}_{\boldsymbol{x}})||p_\beta(.|\alpha_k)\right) \\ &= \log \alpha_k + \left(\frac{\hat{\alpha}_{\boldsymbol{x}}}{\alpha_k}\right)^\beta \frac{1}{\beta} - \log \hat{\alpha}_{\boldsymbol{x}} - \frac{1}{\beta}. \end{aligned}$$

So we can see that, with a given query data $\boldsymbol{x}$, in searching for the best matches, maximizing the log-likelihood $l_k(\boldsymbol{x})$ for $k = 1, 2, \ldots, M$ is *exactly* the same as minimizing the KLD's $d_k(\boldsymbol{x})$ for $k = 1, 2, \ldots, M$. Thus in this simplified case, the retrieval process using the KLD's provides the same result as the optimum ML selection rule that uses direct data from the query image. Note that in the general case, this is only true asymptotically when the size of data $L \to \infty$. Here, however, it is true *for every* L.

E. Relation to Energy-based Methods in the Laplacian Case

Furthermore, consider the case when the parameter β is fixed and equal 1. That is we are modeling the wavelet coefficients using the Laplacian distribution. The extracted feature from wavelet coefficients $\boldsymbol{x}$ of a particular subband is:

$$\hat{\alpha}_{\boldsymbol{x}} = \frac{\sum_{i=1}^{L} |x_i|}{L}. \tag{22}$$

This is precisely the $\mathbf{L}^1$-norm feature of wavelet coefficients as in (10).

From (21), the KLD between two Laplacian distribution is

$$D(p_1(.;\alpha_1)||p_1(.;\alpha_2)) = \log\left(\frac{\alpha_2}{\alpha_1}\right) + \frac{\alpha_1}{\alpha_2} - 1.$$

This is a convex function of α_2/α_1 and is minimum when $\alpha_2/\alpha_1 = 1$. Therefore in term of selecting the most similar images, we are only interested in the situation when the ratio α_2/α_1 is in the vicinity of 1. Using first-order Taylor approximation of $\log x$ around 1, $\log x \approx x - 1$ when $x \approx 1$, we have

$$\begin{aligned} D(p_1(.;\alpha_1)||p_1(.;\alpha_2)) &\approx \frac{\alpha_2}{\alpha_1} - 1 + \frac{\alpha_1}{\alpha_2} - 1 \\ &= \frac{(\alpha_2 - \alpha_1)^2}{\alpha_1 \alpha_2}. \end{aligned}$$

Substitute this into (18), the overall similarity measurement between two images $\mathcal{I}_1$ and $\mathcal{I}_2$ becomes

$$D(\mathcal{I}_1, \mathcal{I}_2) \approx \sum_{j=1}^{B} \frac{(\alpha_2^{(j)} - \alpha_1^{(j)})^2}{\alpha_2^{(j)} \alpha_1^{(j)}}. \tag{23}$$

This distance is essentially the same as the popular weighted Euclidean distance between extracted features $\alpha_i^{(j)}$ where "global" normalization factors $w^{(j)} = \mathrm{var}\{\alpha_i^{(j)} : i = 1, 2, \ldots, M\}$ are replaced by "local" normalization factors $w_{1,2}^{(j)} = \alpha_2^{(j)} \alpha_1^{(j)}$.

Therefore, we demonstrated that our statistical method with a GGD model on the wavelet coefficients can be particularized to closely resemble and thus provide a justification for the weighted Euclidean distance between $\mathbf{L}^1$-norms of wavelet subbands. This is an interesting fact since the two approaches are based on totally different assumptions. The former relies on an underlying stochastic process of the texture image while the later is based on the energy distribution in the frequency domain.

F. Summary of Different Forms of KLD

Let us summarize the different forms of KLD that we have seen so far. In Section III-C we introduced the general formula (17) for the KLD between GGD's for two wavelet subbands together with the overall similarity measurement between two images (18) as the sum of all the distances across wavelet subbands. Section III-D and III-E considered special cases to gain more insight of the technique as well as providing explanation for existing methods. For practical applications, as well as for the following experiments, the general form of KLD's in Section III-C are used.

IV. Experimental Results

We used 40 textures obtained from the MIT Vision Texture (VisTex) database [34] and displayed them in Figure 5. These are real world 512 × 512 images from different natural scenes. Only gray-scale levels of the images (computed from the luminance component) were used in the experiments. Since we define similar textures as subimages from a single original one, we selected texture images whose visual properties do not change too much over the image.

Each of the 512 × 512 images was divided into sixteen 128 × 128 non-overlapping subimages, thus creating a test database of 640 texture images. Furthermore, to eliminate the effect of common range in the gray level of subimages from a same original image and to make the retrieval task less biased, each subimage was individually normalized to zero mean and unit variance before the processing.

A. Testing the Texture Discrimination Power of the GGD Model

The accuracy of GGD's in modeling wavelet coefficients from texture images has been shown in [7] by fitting the estimated PDF curve with the actual histogram of the coefficients. In this section we explore another way of testing the GGD models in terms of comparing synthesized textures visually. This allow us to *visualize* the capability of the GGD models not only in capturing texture information but also in discriminating textures.

In this experiment, we employed the conventional pyramid wavelet decomposition with three levels using the Daubechies' maximally flat orthogonal filters of length 8 (D_4 filters) [35]. From a single image in the database (of size 128 × 128), two GGD parameters were estimated from each of 9 wavelet subbands (except for the lowest band or scale coefficients which corresponds to the approximation of the image) using the ML estimator described in the previous section. Our hypothesis is that those 18 (2×9) model

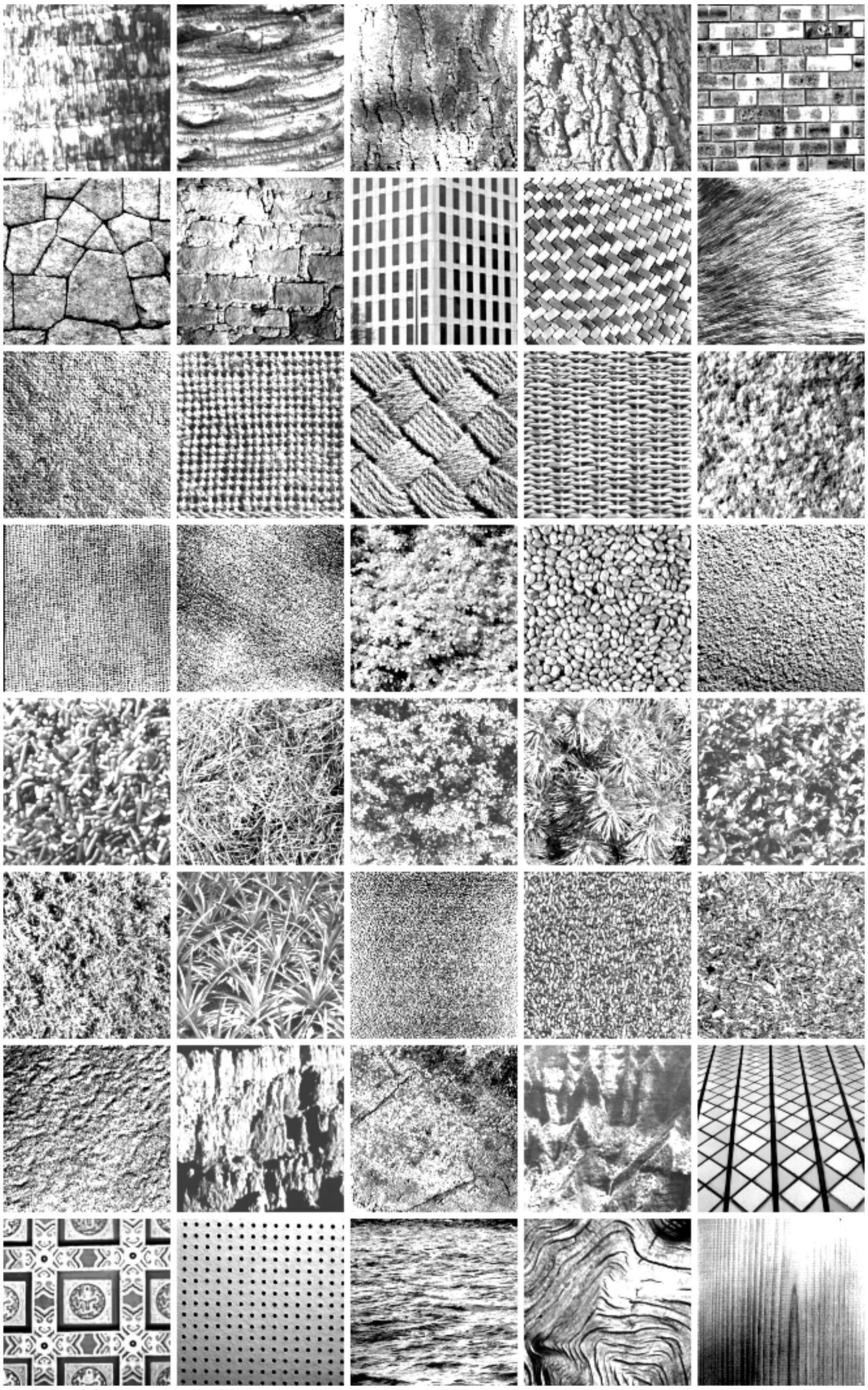
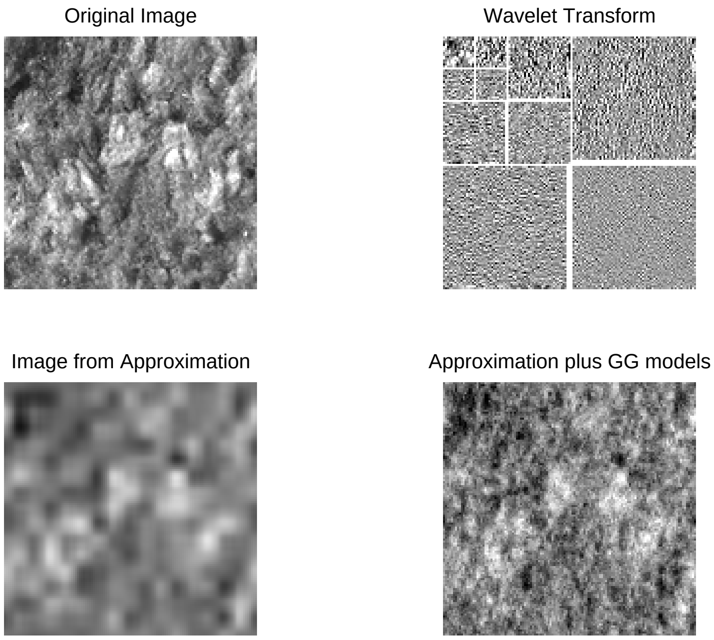


Fig. 5. Texture images from the VisTex collection that are used in the experiments; from left to right and top to bottom: Bark0, Bark6, Bark8, Bark9, Brick1, Brick4, Brick5, Buildings9, Fabric0, Fabric4, Fabric7, Fabric9, Fabric11, Fabric14, Fabric15, Fabric17, Fabric18, Flowers5, Food0, Food5, Food8, Grass1, Leaves8, Leaves10, Leaves11, Leaves12, Leaves16, Metal0, Metal2, Misc2, Sand0, Stone1, Stone4, Terrain10, Tile1, Tile4, Tile7, Water5, Wood1, Wood2.

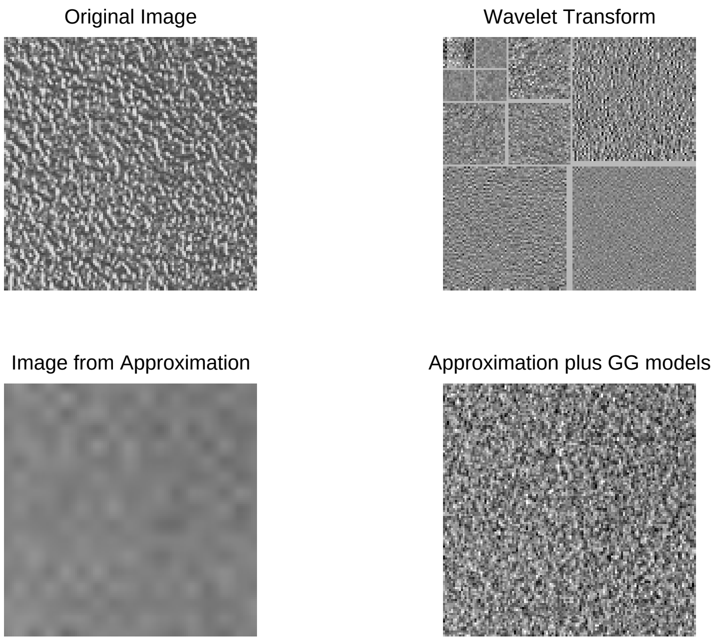
parameters capture important texture-specific features and have discrimination power among texture classes.

To show that, wavelet coefficients at each wavelet subband are generated as i.i.d. sample sequence from the GGD with parameters estimated from the true coefficients. Using those generated wavelet coefficients and the scale coefficients from the original image, the synthesis texture image is obtained by the inverse wavelet transform. Figure 6 shows the example results for two images of size 128×128 from different texture classes in our database. The comparison is also made with the reconstructed images using the scale coefficients only (referred to as coarse approximation images).

We emphasize that the goal here is not texture synthesis



(a) Fabric15 subimage.



(b) Metal0 subimage.

Fig. 6. Texture synthesis examples using generalized Gaussian density for wavelet coefficients on images of size 128×128.

but rather texture discrimination. In this regard, it can be seen from Figure 6 that the extracted 18 GGD model parameters in fact capture some of the main texture-specific information of the indexed images. The synthetic images from two different images using additive GGD models are clearly more distinguishable than the ones using scaling coefficients only.

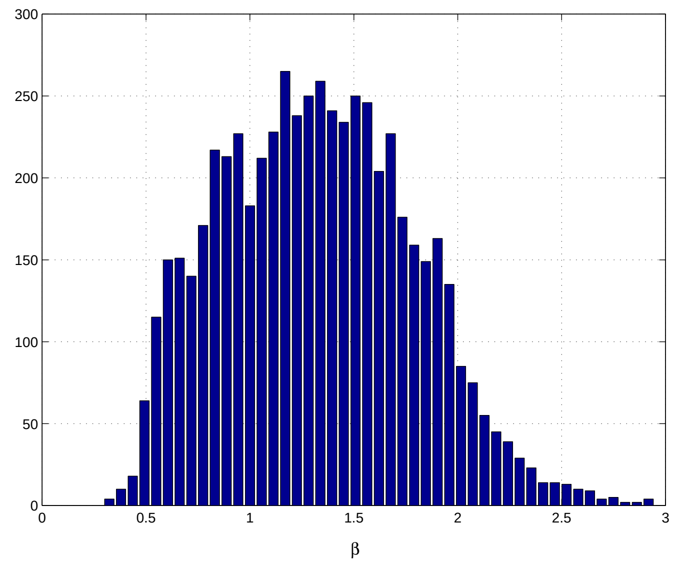


Fig. 7. Histogram of estimated values for β from 640 texture images of size 128 $\times$ 128.

B. Computational Complexity

The proposed texture retrieval system has been implemented in a Matlab environment. The Feature Extraction (FE) step involves talking a wavelet transform of the input image and estimating the GGD model parameters at each subband using ML estimator. It was found that roughly the same amount of time is spent on wavelet transformation and parameter estimation, giving a total of less than 1 second of CPU time on a Sun's Ultra 5 workstation for extracting features from one image.

We applied three levels of wavelet decomposition (which generates 9 wavelet subbands). Therefore, to represent each texture image using the GGD model we need only 18 numbers as an extracted feature set. Thanks to the closed form of distance in (17), the Similarity Measurement (SM) between two images involves simple computation using a small number of model parameters. Optimized implementation using lookup tables yield comparable computation time as normalized Euclidean distance.

C. Range of β in GGD Models

It is of interest to know the common range for the values of β in GGD's for texture images. For typical natural images which are dominated by smooth regions, the values for β are found to be between 0.5 and 1 [29]. Figure 7 shows the histogram of the estimated values of β from our database of 640 texture images using the method described in Section III-B. The discrete wavelet transform of three levels using D_4 filters was used. As can be seen from the figure, the 5-th and 95-th percentile values of the estimated β values for our texture database are around 0.7 and 2.0, respectively.

D. Retrieval Effectiveness

In retrieval experiments, a simulated query image is any one of 640 images in our database. The relevant images for each query are defined as the other 15 subimages from the same original VisTex image. Following [6] we evaluated the performance in terms of the average rate of retrieving relevant images as a function of the number of top retrieved images. The new approach is compared with the traditional methods using the energy-based features in the wavelet domain given in (10-11) together with normalized Euclidean distance as the similarity measurement.

Evaluation of all possible wavelet transform schemes is beyond the scope of the experiments. Thus we restricted our attention to the Daubechies family of orthogonal wavelets [35]. This family is indexed by the number of zeros at $\omega = \pi$ of the lowpass filter. Wavelet D_n uses an orthogonal filter bank with length $2n$ lowpass and highpass filters.

For the number of decomposition levels, our experiments agree with [3] that the size of the smallest subimages should not be less than 16 $\times$ 16 so the estimated energy values or model parameters would be robust. Hence for the input image size 128 $\times$ 128, a maximum of three levels of decomposition is chosen.

In a first series of experiments, the wavelet pyramid transform (DWT) with different filters and decomposition levels were employed. Table I shows the comparison in performance in average percentages of retrieving relevant images in the top 15 matches. Here $\mathbf{L}^1$ and $\mathbf{L}^2$ denote the methods which use texture features computed as in (10) and (11), respectively; while $\mathbf{L}^1 + \mathbf{L}^2$ uses both set of features. Note that the $\mathbf{L}^1 + \mathbf{L}^2$ feature set is essentially equivalent to the one composed of means and standard deviations of the magnitude of the wavelet coefficients that was used in [6]. It also yields the same number of features per images as the statistical method using GGD's (2 features per wavelet subband). The proposed method use the combination of GGD & KLD. We also report the results when the normalized Euclidean distance (ED) was used on GGD's model parameters, which is similar to the method used in [7]. Therefore, except GGD & KLD, all other listed methods use normalized Euclidean distance as the similarity measurement. Following are the main points that we observed:

1. First, the statistical approach (GGD & KLD) always outperforms the traditional methods. This is consistent with our expectation since the GGD parameters are more expressive in characterizing textures than the energy-based ones. Furthermore, the inferior results of the GGD & ED method (where the same features with the statistical method were used but with the normalized Euclidean distance) shows that good performance in retrieval comes not just from a good set of extracted features but also together with a suitable similarity measurement. Hence this supports our approach of considering the two problems FE and SM jointly.
2. Secondly, the length of the filter has little effect in performance in all methods.
3. Finally, in our database, most of the texture discrimination information live in the first two scales of

Filters	Methods				
	$\mathbf{L}^1$	$\mathbf{L}^2$	$\mathbf{L}^1 + \mathbf{L}^2$	GGD & KLD	GGD & ED
D_2	55.50	53.37	56.89	**68.36**	58.73
D_4	55.02	51.37	55.47	**66.36**	57.47
D_6	55.39	53.49	56.15	**65.60**	56.41
D_8	55.80	54.37	56.54	**65.37**	56.45
D_{10}	55.04	52.71	54.97	**64.22**	56.10

(a) One-level decomposition.

Filters	Methods				
	$\mathbf{L}^1$	$\mathbf{L}^2$	$\mathbf{L}^1 + \mathbf{L}^2$	GGD & KLD	GGD & ED
D_2	61.29	57.54	61.43	**74.10**	64.63
D_4	62.97	59.68	62.82	**73.35**	65.87
D_6	61.88	59.57	61.65	**72.72**	64.89
D_8	62.83	61.25	62.99	**72.65**	65.15
D_{10}	62.21	60.62	62.11	**71.61**	63.45

(b) Two-level decomposition.

Filters	Methods				
	$\mathbf{L}^1$	$\mathbf{L}^2$	$\mathbf{L}^1 + \mathbf{L}^2$	GGD & KLD	GGD & ED
D_2	62.72	61.76	64.48	**76.93**	62.22
D_4	63.89	62.54	64.83	**76.57**	61.18
D_6	65.07	63.03	65.11	**75.51**	60.35
D_8	65.60	63.64	65.48	**75.63**	61.73
D_{10}	65.90	64.25	65.81	**74.66**	59.73

(c) Three-level decomposition.

TABLE I

Average retrieval rate (%) in the top 15 matches using pyramid wavelet transform with different filters and decomposition levels.

wavelet decomposition since there is little improvement in retrieval rates when we increased from 2 to three levels of decomposition.

Figure 8 details the comparison between the $\mathbf{L}^1 + \mathbf{L}^2$ and GGD & KLD methods on each texture class using there levels of wavelet decomposition with the D_4 filters. Again we can see that the new method consistently gives superior performance for almost all texture classes, especially for the ones that have structural patterns. Note that those two methods extract the same number of features, 18 ($= 2 \times 9$), from each indexed image.

Figure 9 shows a graph illustrating this comparison in retrieval performances as functions of number of top matches considered. As can be seen, almost the double number of retrieved images is required in the traditional method compared to the new method in order to retrieve the same number of relevant images.

A second series of experiments was conducted for the non-subsampled discrete wavelet frames (DWF). The results of retrieval rates are summarized in Table II. We also listed the results for the wavelet pyramids (DWT) for

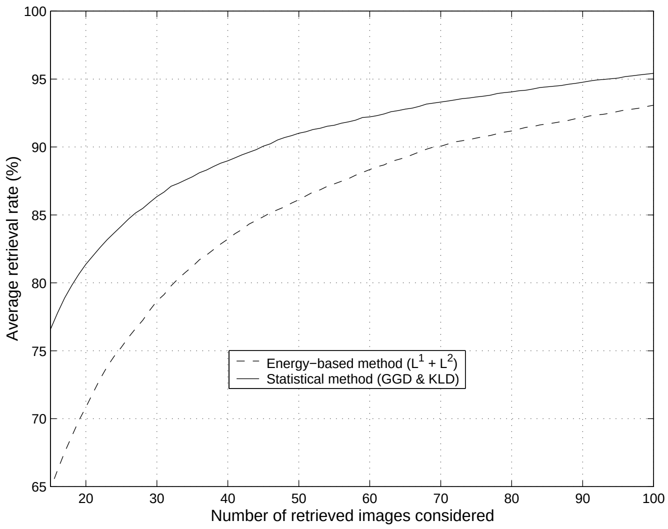


Fig. 9. Retrieval performance according to the number of top matches considered.

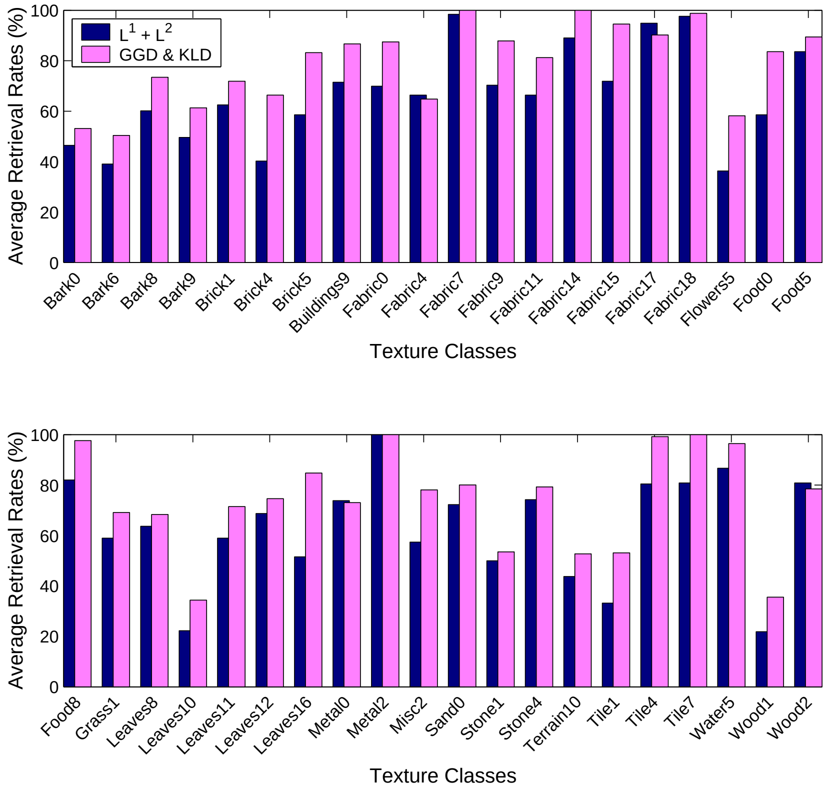


Fig. 8. Average retrieval rates for individual texture class using wavelet pyramid transform with Daubechies' D_4 filters and three decomposition levels.

comparison. As expected, the full rate filterbanks improve the performance over the critical-sampled filterbanks on both the old and new methods. However this improvement is marginal compared to the one achieved when replacing the traditional approach by our proposed method.

E. Image Retrieval Examples

Qualitative evaluation of our method was carried out by visually examining the images of retrieval results. However, this can only be based on a subjective perceptual similarity since there exists no "correct" ordering that is agreed upon by all people [36].

Figure 10 shows some examples of retrieval results to demonstrate the capability of our method. In Figure 10(a), the query image is "leaves". The system almost perfectly retrieves all images of the same leaves and also images of other types of leaves. In Figure 10(b), the query is a fabric patch. In this case, all relevant images are correctly ranked as the top matches following by images of similar textures.

V. Conclusion and Discussions

We have introduced a statistical framework for texture retrieval in CBIR applications by jointly considering the two problems of FE and SM while keeping in mind the complexity constraint of such applications. In our approach, the FE step becomes an ML estimator for the model parameters of image data and the SM step amounts to computing the Kullback-Leibler distances between model parameters. For large data sets, this achieves the same optimality as the ML selection rule.

The statistical framework has been applied successfully in a wavelet-based texture retrieval application, where wavelet coefficients in each subband are independently modeled by a generalized Gaussian density (GGD). This results in a new texture similarity measurement in wavelet domain which has a sound theoretical justification with no need for normalization steps. Furthermore, by restricting to simpler models, the new similarity distance becomes closely related to the popular variance-normalized Euclidean distance. Hence, the statistical approach can be used as a common framework for other existing methods. Experimental results on 640 texture images of 40 classes from the VisTex collection indicated that the new method significantly improves retrieval rates, e.g. from 65% to 77%, over the traditional approaches, using both the pyramid wavelet transform and wavelet frames, while requiring comparable computational time.

We want to emphasize that our method is specially designed for the retrieval problem where the classes are not defined *a priori*. If one looks at a pure classification problem, better results might be obtained by taking into account the distribution (e.g. covariances) of the feature vector itself from each predefined class and then employing the optimal Bayesian classifier [37]. Of course, this requires an additional training step which one cannot usually afford in the general retrieval problem.

Type of decomposition	Methods		
	$\mathbf{L}^1 + \mathbf{L}^2$	GGD & KLD	GGD & ED
1 scale (6 features)			
DWT	55.47	**66.36**	57.47
DWF	56.92	**67.09**	61.34
2 scales (12 features)			
DWT	62.82	**73.35**	65.87
DWF	63.32	**74.01**	69.78
3 scales (18 features)			
DWT	64.83	**76.57**	61.18
DWF	68.48	**78.12**	71.13

TABLE II

AVERAGE RETRIEVAL RATE (%) IN THE TOP 15 MATCHES USING PYRAMID WAVELET TRANSFORM (DWT) AND WAVELET FRAMES (DWF) WITH D_4 FILTERS.

The proposed statistical framework can be applied to other and more general retrieval methods. The GGD was used effectively here for modeling the coefficients from the wavelet transforms and wavelet frames and can applied to other similar filtering schemes such as wavelet packets and Gabor transforms. In [38], [39], [40], we employed the statistical framework to more complex texture models that aggregate wavelet descriptors across scales and orientations using hidden Markov models. Furthermore, we can extend the statistical model for texture using the Wold theory which was shown to closely match human texture perception [36]. As shown in Section II-B, the popular histogram method fits into our scheme. Thus beyond texture, color and local shape features can also be captured. Finally, assuming that different feature sets (color, texture, shape) are independent, the chain rule of the KLD suggests that the overall similarity measurement is simply the sum of KLD's from each feature.

Acknowledgments

The authors would like to thank Professor Michael Unser and Michael Gastpar for many simulating discussions, Zoran Pečenović for developing the image retrieval interface used in our experiments, and one referee for some very helpful comments on the original version of the manuscript.

Appendix

I. Maximum Likelihood Estimator for Generalized Gaussian Density

The MLE for GGD amounts to solve the highly nonlinear equation (16). We thus have to resort to iterative root finding procedures like the Newton-Raphson method.

Define the left hand side of (16) as a function of $\hat{\beta}$, $g(\hat{\beta})$. The Newton-Raphson iteration finds the new guess for the root of $g(\beta)$, β_{k+1}, based on the previous one, β_k, using

$$\beta_{k+1} = \beta_k - \frac{g(\beta_k)}{g'(\beta_k)} \tag{24}$$

We have,

$$\begin{aligned} g'(\beta) = & -\frac{\Psi(1/\beta)}{\beta^2} - \frac{\Psi'(1/\beta)}{\beta^3} + \frac{1}{\beta^2} \\ & - \frac{\sum_{i=1}^{L} |x_i|^\beta (\log|x_i|)^2}{\sum_{i=1}^{L} |x_i|^\beta} + \frac{\left(\sum_{i=1}^{L} |x_i|^\beta \log|x_i|\right)^2}{\left(\sum_{i=1}^{L} |x_i|^\beta\right)^2} \\ & + \frac{\sum_{i=1}^{L} |x_i|^\beta \log|x_i|}{\beta \sum_{i=1}^{n} |x_i|^\beta} - \frac{\log\left(\frac{\beta}{L}\sum_{i=1}^{L} |x_i|^\beta\right)}{\beta^2} \end{aligned} \tag{25}$$

where $\Psi'(z)$ is known as the first polygamma or trigamma function [33]. Note the fact that $g(\beta)$ and $g'(\beta)$ share many common terms which can be used for saving computation at each iteration step in (24).

A good initial guess for the root of $g(\beta)$ can be found based on the matching moments of the data set with those of the assumed distribution [30]. For a GGD, it can be shown that the ratio of mean absolute value to stand deviation is a steadily increasing function of the β:

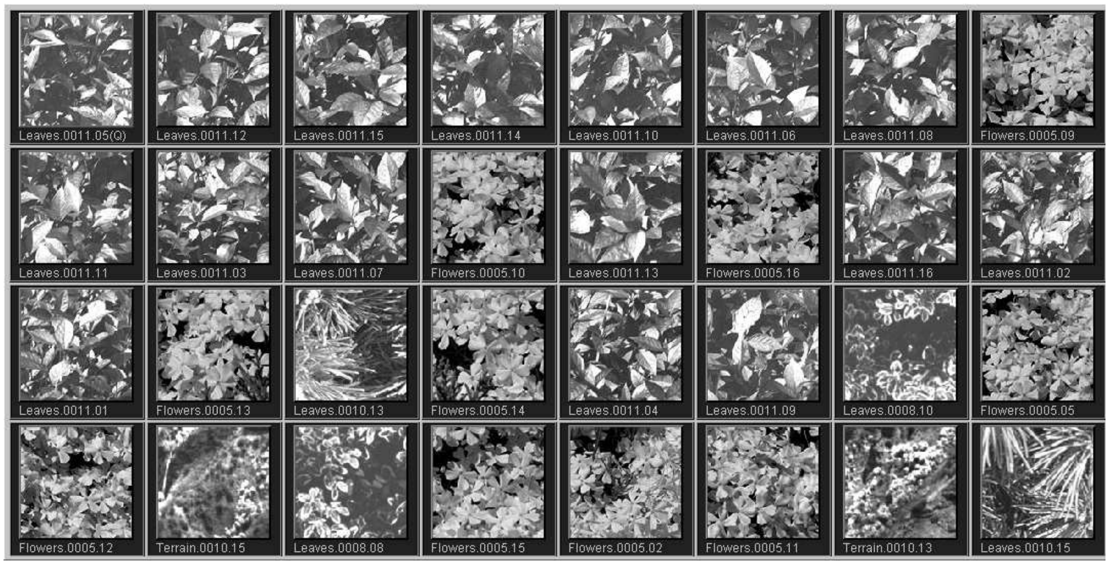
$$\mathcal{F}_M(\beta) = \frac{\Gamma(2/\beta)}{\sqrt{\Gamma(1/\beta)\Gamma(3/\beta)}} \tag{26}$$

Hence if let $m_1 = (1/L)\sum_{i=1}^{L} |x_i|$ and $m_2 = (1/L)\sum_{i=1}^{L} x_i^2$ be the estimate of the mean absolute value and the estimate of the variance of the sample data set, respectively, then β is estimated by solving

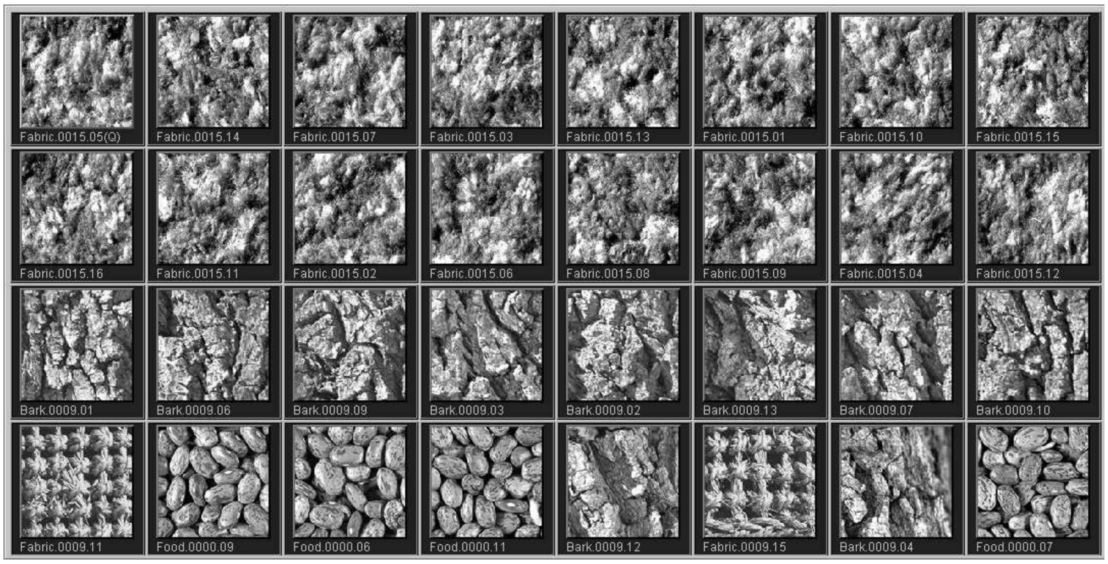
$$\bar{\beta} = \mathcal{F}_M^{-1}\left(\frac{m_1}{\sqrt{m_2}}\right) \tag{27}$$

In a practical implementation, the solution of (27) can be found quickly using interpolation and a look-up table whose entries are the corresponding values of $m_1/\sqrt{m_2}$ and $\bar{\beta}$.

Finally, the initial guess $\beta_0 = \bar{\beta}$ of the ML estimator $\hat{\beta}$ can be "polished up" with a few number steps of Newton-Raphson (experiments showed that only around three steps are adequate).



(a)



(b)

Fig. 10. Examples of retrieval results from 640 texture images based on the VisTex collection. In each case, the query image is on the top left corner; all other images are ranked in the order of similarity with the query image from left to right, top to bottom.

References

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Patt. Recog. and Mach. Intell.*, vol. 22, pp. 1349–1380, Dec. 2000.

[2] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Patt. Recog. and Mach. Intell.*, vol. 15, pp. 1186–1191, 1993.

[3] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structure wavelet transform," *IEEE Trans. Image Proc.*, vol. 2, no. 4, pp. 429–441, 1993.

[4] J. R. Smith and S.-F. Chang, "Transform features for texture classification and discrimination in large image databases," in *Proc. of IEEE Int. Conf. on Image Processing*, 1994.

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

[6] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Patt. Recog. and Mach. Intell.*, vol. 18, no. 8, pp. 837–842, August 1996.

[7] G. V. Wouwer, P. Scheunders, and D. V. Dyck, "Statistical texture characterization from discrete wavelet representations," *IEEE Trans. Image Proc.*, vol. 8, no. 4, pp. 592–598, April 1999.

[8] T. Randen and J. H. Husoy, "Filtering for texture classification: a comparative study," *IEEE Trans. Patt. Recog. and Mach. Intell.*, vol. 21, pp. 291–310, 1999.

[9] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology*, , no. 160, pp. 106–154, 1962.

[10] J. Daugman, "Two-dimensional spectral analysis of cortical receptive field profile," *Vision Research*, vol. 20, pp. 847–856,

1980.
[11] JPEG committee, "JPEG home page," http://www.jpeg.org.
[12] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," *IEEE Trans. Systems, Man and Cybernetics*, vol. 8, pp. 460–473, 1982.
[13] A. R. Rao and G. L. Lohse, "Towards a texture naming system: Identifying relevant dimensions of texture," *Visual Res.*, vol. 36, no. 11, pp. 1649–1669, 1996.
[14] A. Mojsilovic, J. Kovacevic, J. Hu, J. Safranek, and S. K. Ganapathy, "Matching and retrieval based on the vocabulary and grammar of color patterns," *IEEE Trans. Image Proc.*, vol. 9, no. 1, pp. 38–54, Jan. 2000.
[15] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Stelle, and P. Yanker, "Query by image and video content: The QBIC system," *Computer*, pp. 23–32, Sept. 1995.
[16] M. Swain and D. Ballard, "Color indexing," *Int. Journal of Computer Vision*, vol. vol. 7, no. 1, 1991.
[17] N. Vasconcelos and A. Lippman, "A unifying view of image similarity," in *Proc. of IEEE Int. Conf. on Pattern Recognition (ICPR)*, Barcelona, Spain, 2000.
[18] H. V. Poor, *An Introduction to Signal Estimation and Detection*, Springer-Verlag, New York, 2nd edition, 1994.
[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Interscience, New York, NY, 1991.
[20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.
[21] J. Y. Chen, C. A. Bouman, and J. P. Allebach, "Multiscale branch and bound algorithm image database," in *Proc. of SPIE/IS&T Conf. on Storage and Retrieval for Image and Video Data bases V*, San Jose CA, 1997, vol. 3022, pp. 133–144.
[22] J. S. De Bonet and P. Viola, "Texture recognition using a non-parametric multi-scale statistical model," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1998.
[23] C.E. Jacobs, A. Finkelstein, and D.H. Salesin, "Fast multiresolution image querying," in *Computer graphics proceeding of SIGGRAPH*, Los Angeles, 1995, pp. 278–280.
[24] M. Do, S. Ayer, and M. Vetterli, "Invariant image retrieval using wavelet maxima moment," in *Proc. of 3rd Int. Conf. in Visual Information and Information Systems*, 1999, pp. 451–458.
[25] S. C. Zhu, Y. N. Wu, and D. Mumford, "FRAME: Filters, random field and maximum entropy: – towards a unified theory for texture modeling," *Int'l Journal of Computer Vision*, vol. 27, no. 2, pp. 1–20, 1998.
[26] J. R. Bergen and E. H. Adelson, "Theories of visual texture perception," in *Spatial Vision*, D. Regan, Ed. CRC press, 1991.
[27] D. Heeger and J. Bergen, "Pyramid-based texture analysis/synthesis," in *Proc. ACM SIGGRAPH*, 1995.
[28] E. P. Simoncelli and J. Portilla, "Texture characterization via joint statistics of wavelet coefficient magnitudes," in *Proc. of IEEE Int'l Conf. on Image Processing*, 1998.
[29] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Patt. Recog. and Mach. Intell.*, vol. 11, no. 7, pp. 674–693, July 1989.
[30] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 5, pp. 52–56, 1995.
[31] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. Inform. Th.*, vol. 45, pp. 909–919, 1999.
[32] M. K. Varanasi and B. Aazhang, "Parametric generalized Gaussian density estimation," *J. Acoust. Soc. Amer.*, vol. 86, pp. 1404–1415, 1989.
[33] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Tables*, Dover Publications, New York, 1970.
[34] MIT Vision and Modeling Group, "Vision texture," http://vismod.www.media.mit.edu.
[35] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.
[36] F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Trans. Patt. Recog. and Mach. Intell.*, vol. 18, no. 7, pp. 722–733, July 1996.
[37] M. Unser, "Local linear transforms for texture measurements," *Signal Processing*, vol. 11, pp. 61–79, 1986.
[38] M. N. Do, A. C. Lozano, and M. Vetterli, "Rotation invariant texture retrieval using steerable wavelet-domain hidden Markov models," in *Proc. of SPIE Conf. on Wavelet Applications in Signal and Image Processing VIII*, San Diego, USA, Aug. 2000.
[39] M. N. Do and M. Vetterli, "Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models," *IEEE Trans. Multimedia*, Apr. 2001, submitted.
[40] M. N. Do, "Fast approximation of Kullback-Leibler distance for dependence trees and hidden Markov models," *IEEE Signal Proc. Letters*, Oct. 2001, submitted.

Minh N. Do was born in Thanh Hoa, Vietnam in 1974. He received the B.Eng. degree in Computer Engineering (First Class Honours) from the University of Canberra, Australia in 1997. He completed the Graduate School in Communication Systems in 1998, and is doing a Ph.D. in the Audio-Visual Communications Laboratory at the Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland.
He received a Silver Medal from the 32nd International Mathematical Olympiad in Sweden (1991), and a University Medal from the University of Canberra, Australia (1997). His research interests include wavelets, multirate systems, multidimensional signal processing, and visual information representation and retrieval.

Martin Vetterli received the Dipl. El.-Ing. degree from ETH Zürich (ETHZ), Switzerland, in 1981, the MS degree from Stanford University in 1982, and the Doctorat ès Science degree from EPF Lausanne (EPFL), Switzerland, in 1986.
He was a Research Assistant at Stanford and EPFL, and has worked for Siemens and AT&T Bell Laboratories. In 1986, he joined Columbia University in New York where he was last an Associate Professor of Electrical Engineering and co-director of the Image and Advanced Television Laboratory. In 1993, he joined the University of California at Berkeley where he was a Professor in the Dept. of Electrical Engineering and Computer Sciences until 1997, and holds now Adjunct Professor position. Since 1995, he is a Professor of Communication Systems at EPF Lausanne, Switzerland, where he chaired the Communications Systems Division (1996/97), and heads of the Audio-Visual Communications Laboratory. He held visiting positions at ETHZ (1990) and Stanford (1998). He is a fellow of the IEEE, a member of SIAM, and was the Area Editor for Speech, Image, Video, and Signal Processing of the IEEE Transactions on Communications. He is also on the editorial boards of Annals of Telecommunications, Applied and Computational Harmonic Analysis and The Journal of Fourier Analysis and Applications.
He received the Best Paper Award of EURASIP in 1984 for his paper on multidimensional subband coding, the Research Prize of the Brown Bovery Corporation (Switzerland) in 1986 for his doctoral thesis, the IEEE Signal Processing Society's Senior Award in 1991 and in 1996 (for papers with D.LeGall and K.Ramchandran, respectively), and he is a IEEE Signal Processing Distinguished lecturer in 1999. He received the Swiss National Latsis Prize in 1996 and the SPIE Presidential award in 1999. He is a member of the Swiss Council on Science and Technology.
He was a plenary speaker at various conferences (e.g. 1992 IEEE ICASSP) and is the co-author, with J.Kovacevic, of the book **Wavelets and Subband Coding** (Prentice-Hall, 1995). He has published about 85 journal papers on a variety of topics in signal and image processing and holds 5 patents.
His research interests include wavelets, multirate signal processing, computational complexity, signal processing for telecommunications, digital video processing and compression and wireless video communications.