# Rotation Invariant Texture Characterization and Retrieval Using Steerable Wavelet-Domain Hidden Markov Models

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

*Abstract*—We present a new statistical model for characterizing texture images based on wavelet-domain hidden Markov models. With a small number of parameters, the new model captures both the subband marginal distributions and the dependencies across scales and orientations of the wavelet descriptors. Applying to the steerable pyramid, once it is trained for an input texture image, the model can be easily steered to characterize that texture at any other orientation. Furthermore, after a diagonalization operation, we obtain a rotation-invariant model of the texture image. We also propose a fast algorithm to approximate the Kullback–Leibler distance between two wavelet-domain hidden Markov models. We demonstrate the effectiveness of the new texture models in retrieval experiments with large image databases, where significant improvements are shown.

*Index Terms*—Hidden Markov models, image retrieval, Kullback–Leibler distance, rotation invariance, steerable pyramids, texture characterization, wavelets.

## I. INTRODUCTION

**W**ITH the explosive growth of multimedia databases and digital libraries, there is high demand for effective and efficient tools that allow users to search and browse through such collections. The focus of this paper is on the use of texture information for image retrieval applications. Some of the most popular texture extraction methods for retrieval are based on filtering or wavelet-like approaches [1]–[7]. 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. In an image retrieval system, once those features are extracted from each image (in the Feature Extraction step), distances between feature sets of the query image and of each candidate image in the database are computed (in the Similarity Measurement step).

M. N. Do was with the Audiovisual Communications Laboratory, Department of Communication Systems, Swiss Federal Institute of Technology, Lausanne, Switzerland. He is now with the Department of Electrical and Computer Engineering, the Coordinated Science Laboratory, and the Beckman Institute, University of Illinois, Urbana, IL 61801 USA (e-mail: minhdo@uiuc.edu).

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

Typically, the similarity measure is chosen using heuristic arguments, for example the Euclidean distance or its weighted versions between the feature vectors [8].

An alternative approach is to set up the image retrieval in a statistical framework by jointly considering the two problems of feature extraction (FE) and similarity measurement (SM) into a joint modeling and classification scheme, while taking into account the complexity constraint for such applications [9]. In this framework, the FE step becomes a maximum likelihood (ML) estimator for model parameters of image data, and the SM step amounts to computing the Kullback–Leibler distances between the models of the query and of each candidate image. The framework is asymptotically optimal in terms of retrieval error probability, and thus the similarity measurement has a sound theoretical justification.

Using the statistical framework, a natural extension of the wavelet subband energy method for texture retrieval is to model each texture by the marginal densities of its wavelet subband coefficients. In [10], [9], we applied this framework to a simple model where wavelet coefficients in each subband are independently modeled by a generalized Gaussian density (GGD). This method leads to a significant improvement in the retrieval rate over the traditional wavelet subband energy method using both the pyramid wavelet transform and wavelet frames, while requiring comparable computational time.

While having low complexity, the marginal distribution model ignores some important texture-specific information, notably the dependencies of wavelet descriptors across scales and orientations. Furthermore, like most other wavelet-based texture analysis methods, the extracted features are sensitive to the orientation of the analyzed image. This is a drawback in the retrieval applications since a same texture can appear at different orientations in the image database.

In this paper, we address these problems by using a coherent statistical model that captures both wavelet subband marginal distributions and inter-subband dependencies, while being rotation invariant. The proposed model uses a wavelet domain hidden Markov tree [11] and steerable pyramids [12]. Rotation invariance is achieved via a diagonalization of the covariance matrices in the model.

### A. Related Works

Several authors have developed rotation invariant texture features. Kashyap and Khotanzad [13] developed a circular simultaneous autoregressive model for the extraction of rotation invariant texture features. Chen and Kundu [14] modeled

the features of wavelet subbands as a hidden Markov model (HMM). These models are trained using texture samples with different orientations that are treated as being in the same class. Greenspan *et al.* [15], Haley and Manjunath [16] used the magnitude of a discrete Fourier transform in the rotation dimension of features obtained from a multiresolution filtering. Yet another rotation invariant method proposed by Wu and Wei [17] that first converts two-dimensional (2-D) texture images into a (one-dimensional) 1-D signal via spiral sampling, and then applies a HMM on the subband features of the 1-D signal. A comparative study of several rotation invariant texture analysis methods was performed by Fountain *et al.* [18].

Most rotation invariant texture analysis methods were designed for the classification problem, where the classes are defined *a priori*. Therefore, these methods are not suitable for the retrieval application, where each database image forms a separate class and must be individually trained.

### B. Main Contributions

First, this work enhances the recent technique on wavelet-domain hidden Markov models (WD-HMM) [11] for better dealing with images by incorporating the dependency of wavelet coefficients across orientations. Second, by replacing the standard wavelet transform with an overcomplete representation via steerable pyramids [19], we obtain a steerable statistical model that can facilitate rotation invariant applications. Third, for the WD-HMMs to be used effectively in the image retrieval application, we derive a fast algorithm to compute the distance between two WD-HMMs. Finally, our experiments with WD-HMMs in the image retrieval application provide a large scale evaluation of their capacity in discriminating among many different texture classes.

The outline of this paper is as follows. In the next section, we briefly review the statistical framework for image and texture retrieval. Section III discusses the original WD-HMM [11] for one-dimensional signals and presents our extension for 2-D images that takes into account the cross-orientation dependency of wavelet coefficients. By replacing the standard wavelet decomposition with the steerable pyramid [19], Section IV describes a WD-HMM that can be steered to characterize a given texture at any orientation and thus lead to a rotation-invariant model. Section V describes a fast algorithm to approximate the Kullback–Leibler distance between two WD-HMMs, which is crucial for the retrieval application. Experimental results on several texture databases are given in Section VI.

## II. IMAGE RETRIEVAL IN A STATISTICAL FRAMEWORK

### A. General Setting

We start by briefly reviewing the statistical framework for image retrieval [9]. 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. 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 (like wavelet transform). 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.

Suppose that each hypothesis $\mathcal{H}_i$ is modeled by a probability density function (PDF), 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 $\boldsymbol{\Theta}$.

Consider that the query data $\boldsymbol{x} = (x_1, x_2, \ldots, x_L)$ was drew from a model $p(X; \boldsymbol{\theta}_q)$ for the query image. Optimal retrieval (with minimum error probability) is obtained by searching for $\mathcal{I}_i$ that maximizes $\log p(\boldsymbol{x}; \boldsymbol{\theta}_i)$. For large $L$, this can be shown as equivalent to minimizing the *Kullback–Leibler distance* (KLD) or the *relative entropy* [20] between the two PDFs $p(x; \boldsymbol{\theta}_q)$ and $p(x; \boldsymbol{\theta}_i)$

$$D(p(X; \boldsymbol{\theta}_q)\|p(X; \boldsymbol{\theta}_i)) = \int p(x; \boldsymbol{\theta}_q) \log \frac{p(x; \boldsymbol{\theta}_q)}{p(x; \boldsymbol{\theta}_i)}\, dx. \quad (1)$$

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 (1) 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 [21], which means that for the query image, it computes

$$\hat{\boldsymbol{\theta}}_q = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \log p(\boldsymbol{x}; \boldsymbol{\theta}). \quad (2)$$

In summary, by combining FE and SM into a joint modeling and classification framework, the following retrieval scheme is *asymptotically* optimal:

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

*Similarity Measurement:* To select the top $N$ matches to a query, the images in the database are ranked based on the KLDs 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 in size, so that it can meet the timing constraint of the image retrieval 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.

Let us emphasis that, the joint consideration of the two steps FE and SM here is only conceptually, which proves the optimality of our scheme. Computationally, the two steps are performed separately, and thus they fit in the traditional setting of the image retrieval application.

### B. Texture Retrieval Using Generalized Gaussian Density

For wavelet-based texture retrieval, instead of simply describing each subband by its energy measurements, one could use an estimated marginal density. Experiments show that a good PDF approximation for marginal distribution of wavelet

coefficients in a subband is the generalized Gaussian density (GGD) [22], [6], which is defined as:

$$p(x; \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^\beta} \qquad (3)$$

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

Thus, under the GGD model, each wavelet subband is represented by two parameters $\alpha$ and $\beta$: $\alpha$ is the width of the PDF peak (variance), and $\beta$ is inversely proportional with the decreasing rate of the peak. There exists a closed form expression for the KLD between two GGDs [9]

$$D(p(.; \alpha_1, \beta_1)\|p(.; \alpha_2, \beta_2)) = \log\left(\frac{\beta_1\alpha_2\Gamma(1/\beta_2)}{\beta_2\alpha_1\Gamma(1/\beta_1)}\right)$$
$$+ \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} \frac{\Gamma((\beta_2+1)/\beta_1)}{\Gamma(1/\beta_1)} - \frac{1}{\beta_1}. \quad (4)$$

Assume that wavelet subbands are independent, the overall KLD between two images is simply the sum of KLDs across subbands. Experiments in [9] show that the GGD and KLD method lead to significant improvements in retrieval rates over the energy method. Furthermore, the GGD model can be simplified to closely resemble, and thus provide a justification, for the weighted Euclidean distance between $\mathbf{L_1}$-norms of wavelet subbands.

### III. WAVELET-DOMAIN HIDDEN MARKOV MODELS

#### A. Scalar Model

Recently, Crouse *et al.* [11] proposed a new framework for statistical signal processing based on wavelet-domain hidden Markov models (WD-HMMs). It provides an attractive approach to model both the non-Gaussian statistics and the persistence across scale property of wavelet coefficients that are often found in real-world signals. The concept of WD-HMM is briefly reviewed in this section together with the introduction of the notation.

In a WD-HMM, to each wavelet coefficient $w_{j,k}$ at level $j$, $1 \le j \le J$ (the index is such that $j = 1$ corresponds to the coarsest *wavelet* scale, while the scaling coefficients are disregarded in the WD-HMM) there is an associated discrete hidden state $S_{j,k}$ with the probability mass function $P(S_{j,k} = m) = p_{j,k}^{(m)}; m = 1, \dots, M$. Conditioning on its state $S_{j,k} = m$, the coefficient $w_{j,k}$ follows a Gaussian density. Since the wavelet coefficients are obtained from convolutions with filters that have zero sum (the wavelet high-pass filters), they can be assumed to have zero-mean. Furthermore, to reduce the number of parameters in the models, wavelet coefficients at the same subband are *tied* together to share the same statistics. If we take $M = 2$, the marginal distribution wavelets coefficients $w_{j,k}$ at the $j$th level can be written as a mixture of two Gaussian densities

$$w_{j,k} \sim f_j(w) = p_j^{(1)}g\left(w; \sigma_j^{(1)}\right) + p_j^{(2)}g\left(w; \sigma_j^{(2)}\right) \quad (5)$$

where $p_j^{(1)} + p_j^{(2)} = 1$, and $g(w; \sigma)$ denotes the zero-mean Gaussian density with variance $\sigma^2$, i.e., $g(w; \sigma) = (2\pi\sigma^2)^{-1/2}$

$\exp(-w^2/2\sigma^2)$. In this model, $p_j^{(1)}$ and $p_j^{(2)}$ have physical interpretation as the probabilities of the wavelet coefficient $w_{j,k}$ being in the states "small" and "large," respectively. Small coefficients can be considered as outcomes of a small variance probability density function, whereas large cofficients can be considered as outcomes of a large variance density.

There is an inter-scale dependency, most notably between a wavelet coefficient at a coarse level (parent) to the four coefficients at the next intermediate level that correspond to the same location (children) in the image [see Fig. 2(a)]. In order to capture this persistence across scales, there are state transition probability matrices for the *parent* $\to$ *child* link between the hidden states

$$\mathbf{A}_j = \begin{pmatrix} p_j^{1\to 1} & p_j^{1\to 2} \\ p_j^{2\to 1} & p_j^{2\to 2} \end{pmatrix}, \qquad j = 2, 3, \dots, J. \quad (6)$$

Here $p_j^{m\to m'}$ is the probability that a child coefficient at the level $j$ is in the state $m'$ given its parent coefficient is in the state $m$. In other words, across scale, the states of the wavelet coefficients follow a Markov chain. With this, we can relate the state probability at level $j$ with the state probability at the parent level $j - 1$ by

$$p_j^{(m)} = \sum_{m'} p_{j-1}^{(m')} p_j^{m'\to m}, \qquad j = 2, 3, \dots, J. \quad (7)$$

If we denote $\boldsymbol{p}_j = \left( p_j^{(1)} \quad p_j^{(2)} \right)$, then (7) can be written as $\boldsymbol{p}_j = \boldsymbol{p}_{j-1}\mathbf{A}_j$. Hence,

$$\boldsymbol{p}_j = \boldsymbol{p}_1\mathbf{A}_2\mathbf{A}_3\cdots\mathbf{A}_j, \quad \text{for all } j = 2, \dots, J. \quad (8)$$

Therefore, the WD-HMM for a tree of wavelet coefficients (also called hidden Markov tree model) is completely defined by a set of model parameters:

$$\Theta = \left\{\boldsymbol{p}_1, \mathbf{A}_2, \dots, \mathbf{A}_J; \sigma_j^{(m)}, (j=1, \dots, J, m=1, 2)\right\}. \quad (9)$$

where $J$ is the number of wavelet tree levels. The result is a statistical model that effectively captures both the marginal and the joint parent-child distributions of wavelet coefficients. Moreover, there exists an efficient Expectation-Maximization (EM) algorithm for fitting a WD-HMM to observed signal data using the ML criterion [11].

Originally developed for 1-D signals, the WD-HMM has been generalized for images in segmentation [23] and denoising [24] applications. For images, the wavelet transform leads to a decomposition with three orientations, often called horizontal (H), vertical (V) and diagonal (D). The authors in [23], [24] took a simple approach by considering these three orientations separately, thus requiring three independent WD-HMMs to characterize an image, one for each orientation. We refer to those models as *scalar* WD-HMMs.

Fig. 1 shows a typical example of the histogram of the wavelet coefficients from an image subband, together with the plot of the subband marginal density function obtained from the trained WD-HMM. By construction, the estimated marginal density is a mixture of two Gaussian densities as given in (5). For comparison we also show the fitted GGD using the ML estimator [9]. As can be seen from the figure, the WD-HMM provides a close
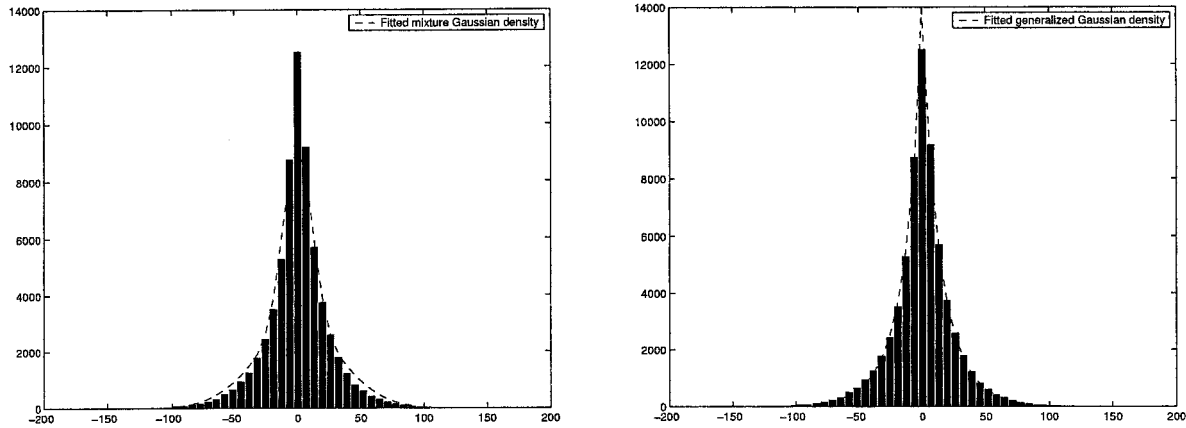
Fig. 1.    Example of wavelet subband coefficient histogram fitted with the marginal distribution curves by WD-HMM and GGD model.



(a)                                                           (b)
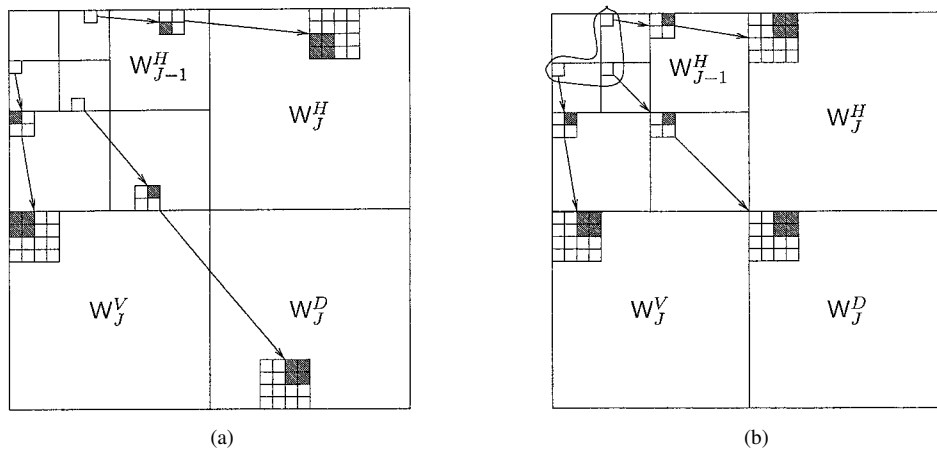
Fig. 2.    Tree structures on the WD-HMMs. In the scalar WD-HMM, there are three scalar models, whereas in the vector WD-HMM, there is one vector model. (a) Scalar model. (b) Vector model.

match to the GGD in terms of modeling the marginal distribution from a wavelet subband. However, the WD-HMM is more expressive than the GGD model by including the dependencies between parent-child coefficients across scales.

### B. Vector Model

The underlying assumption for the scalar WD-HMM approach is that wavelet coefficients at different orientations are independent. However, experiments in [25] show the importance of the cross-correlation of each subband with other orientations at the same scale in characterizing texture images.

To enhance the capacity of WD-HMM in capturing the cross-orientation dependency of wavelet coefficient, we propose to group coefficients at the same location and scale into a vector and then model these vectors by a single multidimensional WD-HMM [see Fig. 2(b)]. The result is one *vector* WD-HMM for the whole input image.

More specifically, denote the wavelet coefficients at the orientation $d$ ($d = 1, 2, 3$ for H, V, D, respectively), scale $j$ and location $\boldsymbol{k}$ as $w_{j,\boldsymbol{k}}^{(d)}$. The grouping operation will produce vectors of coefficients

$$\boldsymbol{w}_{j,\boldsymbol{k}} = \left( w_{j,\boldsymbol{k}}^{(1)} \quad w_{j,\boldsymbol{k}}^{(2)} \quad w_{j,\boldsymbol{k}}^{(3)} \right)^T .$$

Note that $\boldsymbol{w}_{j,\boldsymbol{k}}$ can be seen as the result of the inner products between the input image $\boldsymbol{x}$ with the three local directional wavelet functions $\psi_{j,\boldsymbol{k}}^{(d)}$ at scale $j$ and location $\boldsymbol{k}$ [26]

$$w_{j,\boldsymbol{k}}^{(d)} = \left\langle \boldsymbol{x}, \psi_{j,\boldsymbol{k}}^{(d)} \right\rangle, \qquad d = 1, 2, 3.$$

The marginal distribution function of the wavelet coefficient vectors $\boldsymbol{w}_{j,\boldsymbol{k}}$ at the level $j$ in the vector WD-HMM with tying is expressed as

$$\boldsymbol{w}_{j,\boldsymbol{k}} \sim f_j(\boldsymbol{w}) = p_j^{(1)} g\left( \boldsymbol{w}; \mathsf{C}_j^{(1)} \right) + p_j^{(2)} g\left( \boldsymbol{w}; \mathsf{C}_j^{(2)} \right). \quad (10)$$

Here, $g(\boldsymbol{w}; \mathsf{C})$ denotes the zero-mean multivariate Gaussian density with covariance matrix $\mathsf{C}$, i.e.,

$$g(\boldsymbol{w}; \mathsf{C}) = \frac{1}{\sqrt{(2\pi)^n |\det(\mathsf{C})|}} \exp(-\boldsymbol{w}^T \mathsf{C}^{-1} \boldsymbol{w}) \quad (11)$$

where $n$ is the number of orientations (in this case $n = 3$).

The wavelet coefficient vectors are then organized into a quad-tree structure that connects each vector to its four children at the next intermediate level of the same location [see Fig. 2(b)]. The *parent* $\rightarrow$ *child* link relationships of

these vectors are captured in the same way as in (6) for the scalar WD-HMM. Thus, an image is modeled by one vector WD-HMM with a set of parameters:

$$\Theta = \left\{ \boldsymbol{p}_1, \mathtt{A}_2, \ldots, \mathtt{A}_J; \mathtt{C}_j^{(m)}, (j=1, \ldots, J, m=1, 2) \right\}. \tag{12}$$

Therefore, in a vector WD-HMM, wavelet coefficients at the same scale and location but different orientations are "tied up" to have a same hidden state. The justification for this is that around the edges in an image, wavelet coefficients at all orientations have a high probability of being significant; whereas in smooth regions, all wavelet coefficients are small. In addition, in the vector WD-HMM, the across orientation dependencies are captured via the nondiagonal entries in the covariance matrices of the multivariate Gaussian densities (11).

Since any marginal density from a multivariate Gaussian density is also a Gaussian density, from (10), the marginal density for each wavelet subband in a vector WD-HMM is also a mixture of two zero-mean Gaussian densities. Thus, one can expect that the vector WD-HMM also captures the subband dependent marginal probability distributions of wavelet coefficients as the scalar WD-HMM.

In [27], Fan and Xia proposed a different way of grouping wavelet coefficients across orientation for the 2-D WD-HMM, in that each combination of the three hidden states for three wavelet coefficients at the same location and scale is represented by a single state. If the number of the hidden states for each wavelet coefficient is two, then there will be eight states for each wavelet coefficient vector, and thus each transition matrix is $8 \times 8$. Their method leads to a model with a large number of parameters, $62J - 7$, where $J$ is the number of wavelet decomposition levels. This increases the computational and storage costs significantly, which might not be suitable for the image retrieval application.

### C. Relations Between Models

In this section we draw the connections between the generalized Gaussian density (GGD) model and the scalar and vector WD-HMMs. As already discussed, all of these models capture the subband-dependent marginal probability density function. This is a crucial point since psychological research on human texture perception suggests that two homogeneous textures are often difficult to discriminate if they produce similar marginal distributions of responses from a bank of filters [28].

In [10], by simply modeling those PDFs by GGDs, we obtained good retrieval results, compared to the traditional subband energy approach. The scalar WD-HMM adds on extra texture-specific information by capturing the inter-scale dependencies (via the state transition matrices). The vector WD-HMM furthermore adds on the inter-orientation dependencies information (via the nondiagonal entries in the covariance matrices) in characterizing textures.

Table I shows the number of free parameters needed to describe each image using different models, when the wavelet transform is decomposed with $J$ levels. Note that due to the row sums property, each $\mathtt{A}_j$ has only two free parameters. The covariance matrices $\mathtt{C}$ are symmetric, thus they contain six free parameters each.

TABLE I
NUMBER OF FREE PARAMETERS NEEDED TO SPECIFY DIFFERENT MODELS FOR AN IMAGE WHEN THE WAVELET TRANSFORM IS TAKEN TO $J$ LEVELS

|  | GGD | scalar WD-HMM | vector WD-HMM |
|---|---|---|---|
| Hidden states | -- | $3 \times (2J - 1)$ | $2J - 1$ |
| Densities | $3 \times J \times 2$ | $3 \times J \times 2$ | $J \times 2 \times 6$ |
| Total | $6J$ | $12J - 3$ | $14J - 1$ |

## IV. ROTATION INVARIANCE USING STEERABLE WD-HMM

### A. Steerable WD-HMM

Both the scalar and the vector WD-HMMs described above have drawbacks in that they are sensitive to the orientation of the input image. This problem has roots in the standard wavelet transform. If the image is rotated, then in the wavelet domain the wavelet coefficients change completely. In fact, the wavelet coefficients of the rotated image are *not* just be simply rotated, but are also modified.

One way to remedy this situation is to replace the standard wavelet decomposition with the steerable pyramid [12], [19]. The steerable pyramid is a linear multiscale, multi-orientation image decomposition where the basis functions are directional derivative operators. This transform satisfies the *shiftability* in the orientation condition, which means that at a fixed location and scale the response at an arbitrary orientation is equals to a linear combination of coefficients corresponding to the oriented basis functions at that location and scale. More specifically, Denote $\boldsymbol{w}$ and $\boldsymbol{w}_\phi$ as the vectors of the steerable pyramid coefficients at fixed scale $j$ and location $\boldsymbol{k}$ for an input image and its rotated copy by $\phi$ respectively, then we have

$$\boldsymbol{w}_\phi = \mathtt{R}_\phi \boldsymbol{w}. \tag{13}$$

The columns of $\mathtt{R}_\phi$ are a set of interpolation functions that depend only on the rotation angle $\phi$ and the steerable basis functions. Furthermore, orientation shiftability ensures the orientation invariance of response power [12], i.e., $\|\boldsymbol{w}_\phi\|^2 = \|\boldsymbol{w}\|^2$ for any $\boldsymbol{w}$. This is equivalent [29] to saying that $\mathtt{R}_\phi$ is an orthogonal matrix, i.e., $\mathtt{R}_\phi^{-1} = \mathtt{R}_\phi^T$. As a bonus, the steerable pyramid representation is also translation-invariant.

*Proposition 1:* Suppose that $\Theta = \{\boldsymbol{p}_1, \mathtt{A}_2, \ldots, \mathtt{A}_J; \mathtt{C}_j^{(m)}$ $(j = 1, \ldots, J; m = 1, 2)\}$ is the vector WD-HMM on a steerable pyramid of an image. Then the corresponding model for the rotated version of that image by $\phi$ is $\Theta_\phi = \{\boldsymbol{p}_1, \mathtt{A}_2, \ldots, \mathtt{A}_J; \mathtt{C}_{j,\phi}^{(m)}(j = 1, \ldots, J; m = 1, 2)\}$. The only change is the covariance matrices

$$\mathtt{C}_{j,\phi}^{(m)} = \mathtt{R}_\phi \mathtt{C}_j^{(m)} \mathtt{R}_\phi^T, \qquad j = 1, \ldots, J; m = 1, 2. \tag{14}$$

*Proof:* Using (10) and (13), we can write the marginal distribution function of the coefficient vectors $\boldsymbol{w}_\phi$ at the level $j$ of the rotated texture as

$$f_{j,\phi}(\boldsymbol{w}_\phi) = \frac{1}{J(\boldsymbol{w}_\phi, \boldsymbol{w})} f_j(\mathtt{R}_\phi^{-1} \boldsymbol{w}_\phi)$$

$$= p_j^{(1)} g\left(\mathtt{R}_\phi^{-1} \boldsymbol{w}_\phi; \mathtt{C}_j^{(1)}\right) + p_j^{(2)} g\left(\mathtt{R}_\phi^{-1} \boldsymbol{w}_\phi; \mathtt{C}_j^{(2)}\right)$$

since the Jacobian $J(\boldsymbol{w}_\phi, \boldsymbol{w}) = |\det(\mathtt{R}_\phi)| = 1$. Using the fact that $\mathtt{R}_\phi$ is an orthogonal matrix again, by manipulating (11) we have

$$g\left(\mathtt{R}_\phi^{-1}\boldsymbol{w}_\phi; \mathtt{C}\right) = g\left(\boldsymbol{w}_\phi; \mathtt{R}_\phi \mathtt{C}\mathtt{R}_\phi^T\right).$$

Thus, $f_{j,\phi}(\boldsymbol{w}_\phi)$ is also a mixture of two zero-mean multivariate Gaussian densities which has the same probability mass function $p_j^{(m)}$ for the hidden state as in $f_j(\boldsymbol{w})$, whereas the covariance matrices are transformed by (14). Combining this across scales we obtain the desired result. □

As a result, the vector WD-HMM on a steerable pyramid is a *steerable* model. In other words, one can train a WD-HMM for a single orientation of a texture and then steer this model, with a simple transformation, to describe that texture at any other orientation.

### B. Rotation Invariance Using Steerable WD-HMM

Using the steerable WD-HMM above, we now develop a rotation-invariant model for texture retrieval. Recall that the only difference between the steerable WD-HMMs $\Theta$ and $\Theta_\phi$ of a given texture and its rotated version is among the covariance matrices. These covariance matrices are related by (14), or $\mathtt{C}_j^{(m)}$ and $\mathtt{C}_{j,\phi}^{(m)}$ are said to be orthogonally equivalent [29].

Using the Takagi's factorization [29], we factor each covariance matrix in the steerable WD-HMM into a product

$$\mathtt{C}_j^{(m)} = \mathtt{U}_j^{(m)} \Lambda_j^{(m)} \mathtt{U}_j^{(m)T}, \qquad j = 1, \dots, J; m = 1, 2 \quad (15)$$

where $\mathtt{U}_j^{(m)}$ is the orthogonal matrix whose columns are the normalized eigenvectors of $\mathtt{C}_j^{(m)}$ and $\Lambda_j^{(m)}$ is the diagonal matrix containing the real, nonnegative eigenvalues of $\mathtt{C}_j^{(m)}$ in descending order. This factorization is always possible since all covariance matrices are symmetric and positive semidefinite.

Let $\lambda$ be an eigenvalue of $\mathtt{C}_{j,\phi}^{(m)}$. That means there exists a vector $\boldsymbol{v}$ such that $\mathtt{C}_{j,\phi}^{(m)}\boldsymbol{v} = \lambda\boldsymbol{v}$. Using (14), we have

$$\mathtt{R}_\phi \mathtt{C}_j^{(m)} \mathtt{R}_\phi^T \boldsymbol{v} = \lambda\boldsymbol{v} \quad \text{or} \quad \mathtt{C}_j^{(m)} \mathtt{R}_\phi^T \boldsymbol{v} = \lambda\mathtt{R}_\phi^T\boldsymbol{v}.$$

If we denote $\boldsymbol{v}' = \mathtt{R}_\phi^T\boldsymbol{v}$, then $\mathtt{C}_j^{(m)}\boldsymbol{v}' = \lambda\boldsymbol{v}'$. Hence, $\lambda$ is also an eigenvalue of $\mathtt{C}_j^{(i)}$. Thus, the diagonalization operation on the rotated model $\Theta_\phi$ leads to

$$\Lambda_{j,\phi}^{(m)} = \Lambda_j^{(m)}, \quad \text{for all } j = 1, \dots, J; m = 1, 2.$$

In summary, given a steerable WD-HMM, we can factorize the covariance matrices into the form of (15), where the $\mathtt{U}_j^{(m)}$ matrices are responsible for the orientation of the input image while the $\Lambda_j^{(m)}$ matrices contain rotation-invariant texture information. Thus we have the following result.

*Proposition 2:* The *diagonalized* steerable WD-HMM

$$\overline{\Theta} = \left\{ \boldsymbol{p}_1, \mathtt{A}_2, \dots, \mathtt{A}_J; \Lambda_j^{(m)}(j=1, \dots, J; m=1, 2) \right\} \quad (16)$$

is a rotation-invariant model.

*Remark 1:* In practice one estimates a WD-HMM for an input image via the EM algorithm using the ML criterion. So the rotation invariant property of the estimated model relies on the assumption that the ML solution of the WD-HMM is unique and the EM training algorithm is able to find it.

## V. KULLBACK–LEIBLER DISTANCE BETWEEN WD-HMMS

The statistical framework in Section II suggests that the Kullback–Leibler distance (KLD) should be used to compute the dissimilarity between WD-HMMs. An additional advantage of using the KLD is that since it is defined directly on the extracted model's parameters, therefore with rotation-invariant models it leads to a rotation-invariant image retrieval system.

However, there is no closed form expression for the KLD between hidden Markov tree models. A simple solution is to resort to a Monte-Carlo method for computing the integral in the KLD [30], [31]. More specifically, from the query model we randomly generate a data set as wavelet coefficient trees (each tree consists of a coefficient or a vector coefficient at the coarsest level and all of its descendants), and then compute its likelihood against each candidate model. With this method, for an accurate approximation of the KLD, the generated data set has to contain a large number of trees. This can be prohibitively expensive in the retrieval application, where the distance has to be computed for a large number of images in the *interactive* mode. Furthermore, due to the "random" nature of the Monte-Carlo method, the approximations of the distance could vary in different computations. In [32], we propose a fast algorithm to approximate the KLD between two general dependent tree models. Apply this to the WD-HMMs, due to the tying of parameters, the algorithm is significantly simplified and is described next.

Consider the KLD between two vector WD-HMMs $\Theta$ and $\tilde{\Theta}$ that are defined in (12). Essentially, the proposed algorithm employs the "upward" procedure to compute an upper bound for the KLD between two WD-HMMs by successively using the following inequality.

*Lemma 1 [32]:* The KLD between two mixture densities $\sum_i w_i f_i$ and $\sum_i \tilde{w}_i \tilde{f}_i$ is upper bounded by

$$D\left(\sum_i w_i f_i \middle\| \sum_i \tilde{w}_i \tilde{f}_i\right) \le D(\boldsymbol{w}\|\tilde{\boldsymbol{w}}) + \sum_i w_i D\left(f_i \middle\| \tilde{f}_i\right), \quad (17)$$

with equality if and only if $w_i f_i(x) = c(x)\tilde{w}_i \tilde{f}_i(x)$, for all $i$.

Here $D(\boldsymbol{w}\|\tilde{\boldsymbol{w}})$ denotes the KLD between two probability mass functions $\boldsymbol{w} = (w_i)_i$ and $\tilde{\boldsymbol{w}} = (\tilde{w}_i)_i$

$$D(\boldsymbol{w}\|\tilde{\boldsymbol{w}}) = \sum_i w_i \log \frac{w_i}{\tilde{w}_i}. \quad (18)$$

Denote $\beta_j^{(m)}$ to be the conditional probability density of the wavelet coefficient subtrees that have root from a node at level $j$, given the state of that node is $m$. The key property of the WD-HMM is that given the state of a node in the wavelet coefficient tree, the wavelet coefficients attached to that node and its subtrees are independent. Thus, applying the chain rule of the KLD, we have

$$D\left(\beta_{j-1}^{(m)} \middle\| \tilde{\beta}_{j-1}^{(m)}\right) = D\left(g\left(.; \mathtt{C}_{j-1}^{(m)}\right) \middle\| g\left(.; \tilde{\mathtt{C}}_{j-1}^{(m)}\right)\right)$$

$$+ 4D\left(\sum_{m'=1}^2 p_j^{m \to m'} \beta_j^{(m')} \middle\| \sum_{m'=1}^2 \tilde{p}_j^{m \to m'} \tilde{\beta}_j^{(m')}\right)$$

since there are four subtrees at each node, and these subtrees share the same statistics.

Let $\boldsymbol{a}_j^{(m)}$ be the $m$th row vector of the transition probability matrix $\mathtt{A}_j$, then by applying (17) to the above equation, we obtain

$$D\left(\beta_{j-1}^{(m)} \left\| \tilde{\beta}_{j-1}^{(m)}\right.\right) \leq D\left(g\left(.; \mathtt{C}_{j-1}^{(m)}\right) \left\| g\left(.; \tilde{\mathtt{C}}_{j-1}^{(m)}\right)\right.\right)$$

$$+4D\left(\boldsymbol{a}_j^{(m)} \left\| \tilde{\boldsymbol{a}}_j^{(m)}\right.\right)+4\sum_{m'=1}^{2} p_j^{m\to m'} D\left(\beta_j^{(m')} \left\| \tilde{\beta}_j^{(m')}\right.\right). \quad (19)$$

Denote

$$d_j^{(m)} = D\left(g\left(.; \mathtt{C}_{j-1}^{(m)}\right) \left\| g\left(.; \tilde{\mathtt{C}}_{j-1}^{(m)}\right)\right.\right) + 4D\left(\boldsymbol{a}_j^{(m)} \left\| \tilde{\boldsymbol{a}}_j^{(m)}\right.\right),$$

$$D_j^{(m)} = D\left(\beta_j^{(m)} \left\| \tilde{\beta}_j^{(m)}\right.\right),$$

$$\boldsymbol{d}_j = \begin{pmatrix} d_j^{(1)} & d_j^{(2)} \end{pmatrix}^T,$$

and

$$\boldsymbol{D}_j = \begin{pmatrix} D_j^{(1)} & D_j^{(2)} \end{pmatrix}^T.$$

Then, (19) can be written in a more compact form as

$$\boldsymbol{D}_{j-1} \leq \boldsymbol{d}_j + 4\mathtt{A}_j \boldsymbol{D}_j. \quad (20)$$

For the KLD between two zero-mean $n$-dimensional Gaussians, we have the following closed form expression [33]

$$D\left(g(.; \mathtt{C}) \left\| g\left(.; \tilde{\mathtt{C}}\right)\right.\right)$$

$$= \frac{1}{2}\left(\log\frac{\left|\det\left(\tilde{\mathtt{C}}\right)\right|}{\left|\det(\mathtt{C})\right|} - n + \text{trace}\left(\tilde{\mathtt{C}}^{-1}\mathtt{C}\right)\right). \quad (21)$$

Initially, at the lowest level $j = J$, we simply have

$$D_J^{(m)} = D\left(g\left(.; \mathtt{C}_J^{(m)}\right) \left\| g\left(.; \tilde{\mathtt{C}}_J^{(m)}\right)\right.\right).$$

And finally, at the top level $j = 1$

$$D\left(\Theta \left\| \tilde{\Theta}\right.\right) = D\left(\sum_{m=1}^{2} p_1^m \beta_1^{(m)} \left\| \sum_{m=1}^{2} \tilde{p}_1^m \tilde{\beta}_1^{(m)}\right.\right)$$

$$\leq D\left(\boldsymbol{p}_1 \left\| \tilde{\boldsymbol{p}}_1\right.\right) + \sum_{m=1}^{2} p_1^m D\left(\beta_1^{(m)} \left\| \tilde{\beta}_1^{(m)}\right.\right)$$

$$= D\left(\boldsymbol{p}_1 \left\| \tilde{\boldsymbol{p}}_1\right.\right) + \boldsymbol{p}_1 \boldsymbol{D}_1.$$

Thus, we can use (20) recursively upward to compute an upper bound for $D(\Theta\|\tilde{\Theta})$. This bound is tighter, hence provides a more accurate approximation of the true KLD, when the two models' parameters are close. This property makes the proposed bound particularly fits to the retrieval application. Recall that the task of a retrieval system is to find a small set of images that are similar to the query image. For candidate images with model parameters that are significantly different with the query's model, the proposed upper bound will "overestimate" the true KLD, but this does not matter since we will discard these images anyway. For candidate images with model parameters that are close to the query's model, the proposed bound will be closer to the true KLD, thus leads to accurate selection and ranking of top similar images.
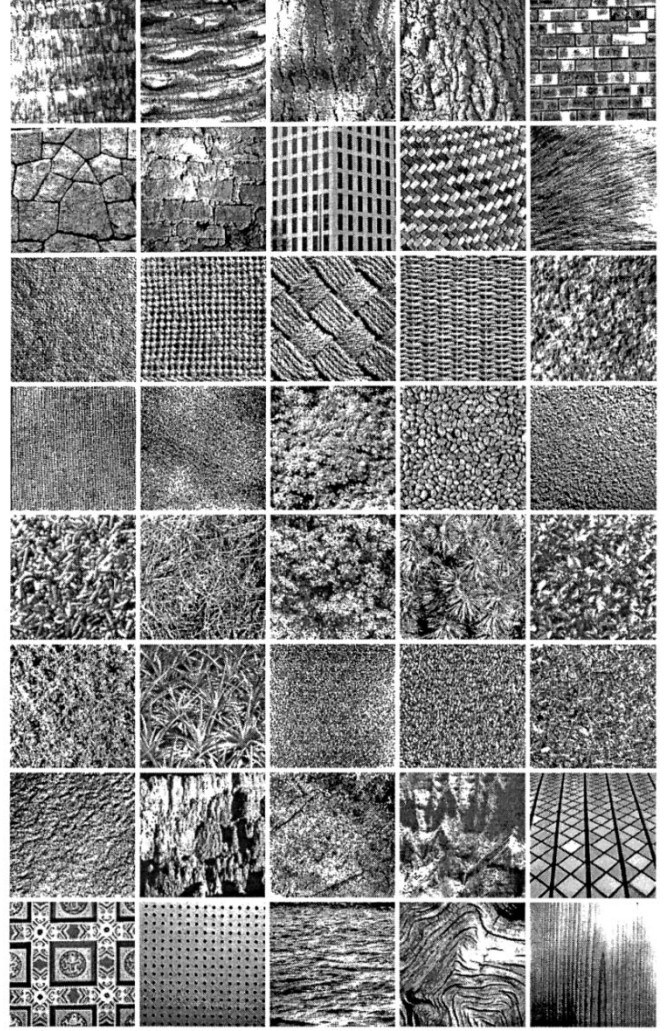


Fig. 3. Texture images from the VisTex collection that are used in the experiments; from left to right and top to bottom: Bark0, Bark6, Bark8, Bark9, Brick1, Brick4, Brick5, Buildings9, Fabric0, Fabric4, Fabric7, Fabric9, Fabric11, Fabric14, Fabric15, Fabric17, Fabric18, Flowers5, Food0, Food5, Food8, Grass1, Leaves8, Leaves10, Leaves11, Leaves12, Leaves16, Metal0, Metal2, Misc2, Sand0, Stone1, Stone4, Terrain10, Tile1, Tile4, Tile7, Water5, Wood1, and Wood2.

The algorithm has low computational complexity that is linear with the number of wavelet decomposition levels. For instance, with the rotation invariant WD-HMMs described in Section IV-B, approximating a KLD requires about $15J$ multiplications and $8J$ additions. In other words, the cost of approximating the KLD between WD-HMMs using the proposed algorithm is compatible to computing the Euclidean distance between feature vectors.

## VI. EXPERIMENTAL RESULTS

### A. Databases

We use two texture databases in our experiments. In a first series of experiments, we evaluate the retrieval effectiveness of both scalar and vector WD-HMMs against the GGD method in a large database. For this, we used the same 40 VisTex [34] textures that were tested in [9], and displayed them in Fig. 3. These are real world $512 \times 512$ images from different natural
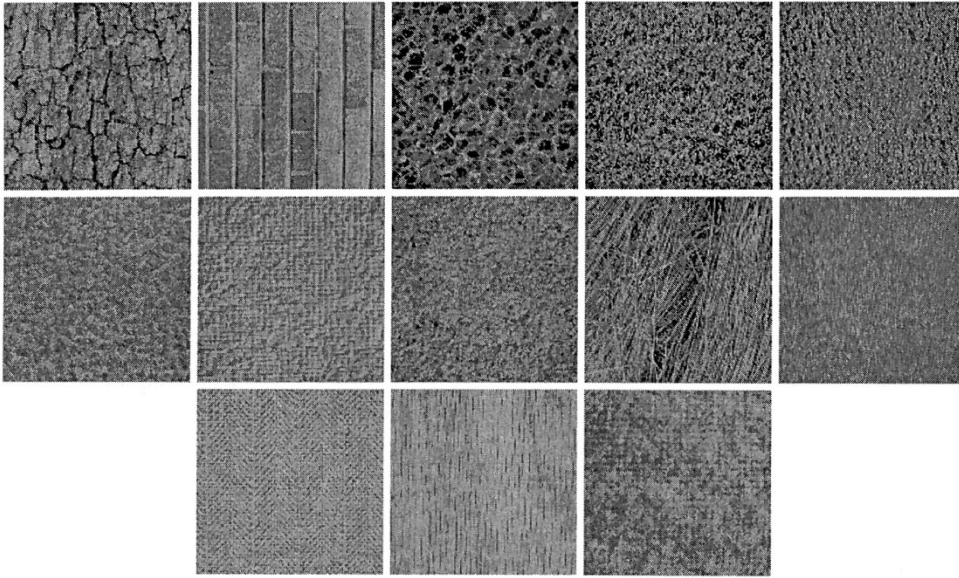
Fig. 4. Texture images from the rotate collection that are used in the experiments. The images are at $0°$. From left to right and top to bottom are Bark, Brick, Bubbles, Grass, Leather, Pigskin, Raffia, Sand, Straw, Water, Weave, Wood, and Wool.

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 $512 \times 512$ original images was divided into 16 $128 \times 128$ nonoverlapping 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.

The second image collection is used to test the rotation-invariant property of WD-HMMs. It consists of 13 $512 \times 512$ Brodatz texture images that were rotated to various degrees *before* being digitized [35]. Fig. 4 displays the original textures at the $0°$ or nonrotated position. From these images, we first construct the *nonrotated* image set by dividing each of the original $0°$ image into 16 $128 \times 128$ nonoverlapping subimages. Next, we construct the *rotated* image set by taking four nonoverlapping $128 \times 128$ subimages each from the original images at 0, 30, 60, and $120°$. Both databases contain 208 images that come from 13 texture classes. The *nonrotated* set serves as the ideal case, where all images in a same class have the same orientation, for the *rotated* set.

In retrieval experiments, a simulated query image is any one of $128 \times 128$ images in a database. The relevant images for each query are defined as the other 15 subimages from the same original image. Following [5] we evaluated the performance in terms of the percentage of relevant images among the top 15 retrieved images.

### B. Effectiveness of WD-HMMs

For this series of experiments, we used the standard discrete wavelet transform (DWT) with Haar filters and three decom-

position levels. We chose three levels of decomposition for our experiments since most of the texture information of our database is concentrated in those three levels.

Fig. 5 details the retrieval performance on the database of 640 textures images from the VisTex collection by using three statistical models: GGD, scalar WD-HMM and vector WD-HMM, to characterize wavelet coefficients. For comparison, we also show the performance of the common approach that uses the wavelet subband energies, i.e., $\mathbf{L_1}$ and $\mathbf{L_2}$ norms, from each subband as extracted features and uses the variance-weighted Euclidean distance between feature vectors for similarity measurement (for more details of this method, see [5], [9]).

We observe that all of the statistical methods outperform the traditional energy-based method. The scalar WD-HMM method gives compatible results to the GGD method, whereas the vector WD-HMM method significantly improves the retrieval rates in many texture classes, as well on average (Table II). Focusing on the WD-HMM methods, we see that the vector model outperforms the scalar model in almost all texture classes. This is consistent with the argument in Section III-B that the vector model is more precise in characterizing textures as it includes the inter-orientation dependency information.

### C. Effectiveness of Rotation Invariance

In the second series of experiments, we test the rotation invariant property of the steerable WD-HMM that is described in Section IV-B. We use a steerable pyramid having two directions and three levels of decomposition. Fig. 6 shows the comparison of the performances in average percentages of retrieving relevant images for the *nonrotated* set, the *rotated* set **without** using rotation invariant model, and the *rotated* set **with** rotation invariant model.

First, we compare the retrieval results obtained from the *nonrotated* set to the *rotated* set, **without** using rotation invariance. We see that textures which have similar results for both sets
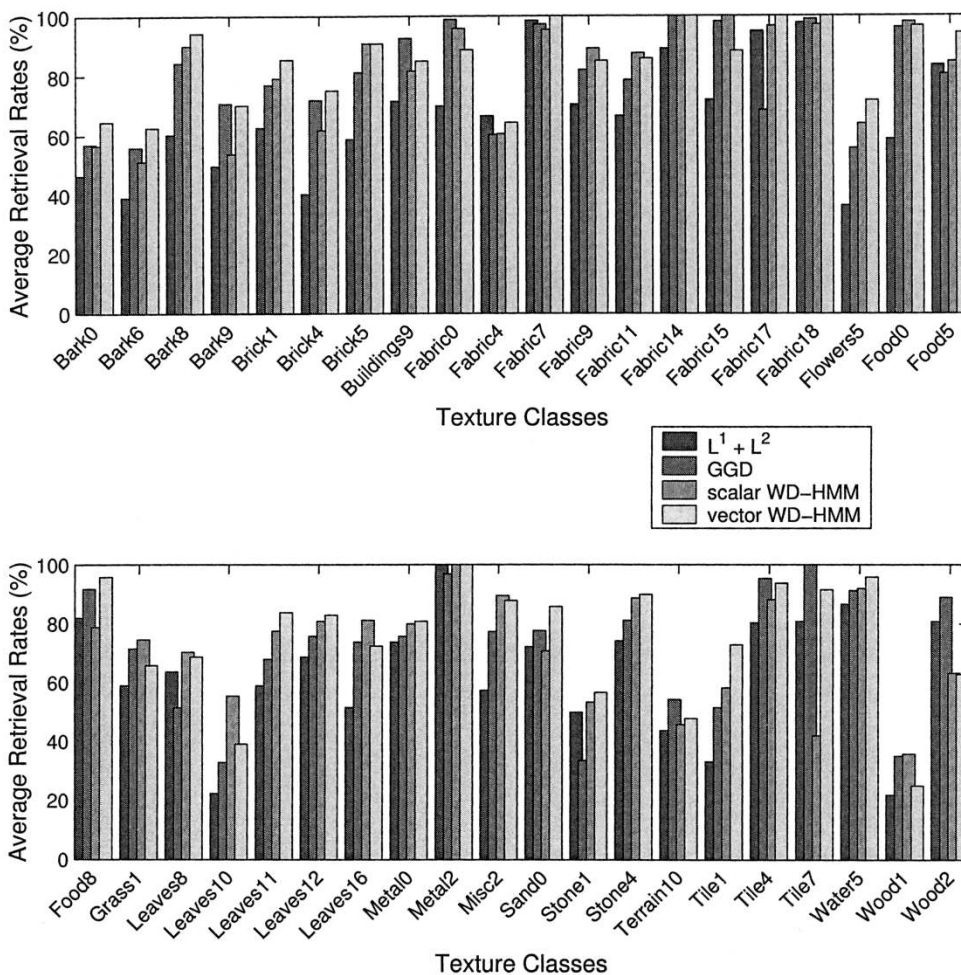
Fig. 5.   Average retrieval rates for individual texture class using standard wavelet transform with Haar filters and three decomposition levels.
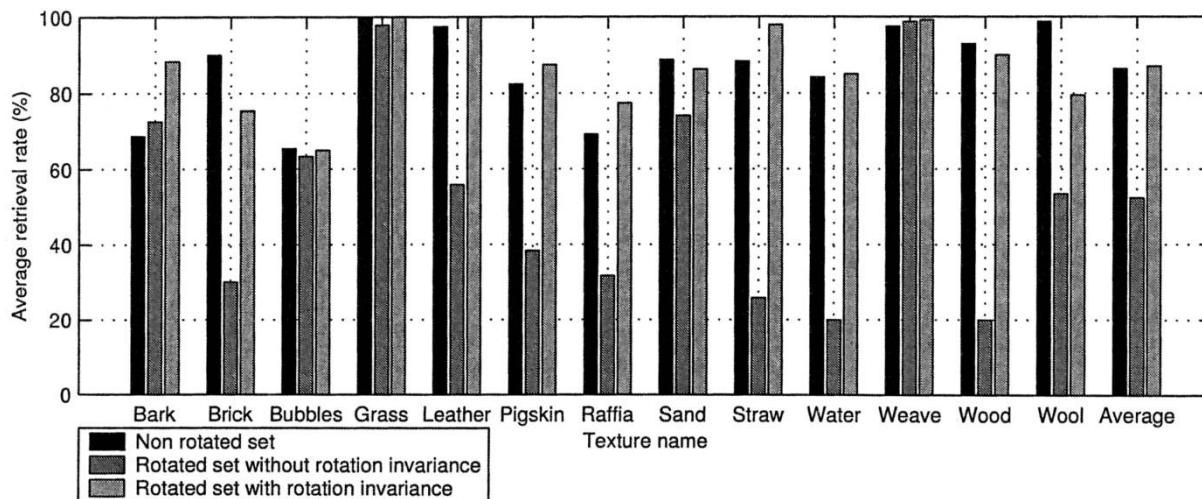


Fig. 6.   Average retrieval rates for individual texture class using order one steerable filters and three decomposition levels for *nonrotated* set, *rotated* set **without** rotation invariance and *rotated* set **with** rotation invariance.

(Bark, Bubbles, Grass, Weave) are the ones that have no strong direction, as those textures are not affected by rotation. Moreover they all have very distinct texture patterns. Textures which are most seriously affected by rotation (Brick, Leather, Pigskin,

Raffia, Straw, Water, Wood, Wool) are the ones that are strongly directional.

By contrast, the retrieval results obtained for the *rotated* set **with** rotation invariance are almost the same as those from the

TABLE II
AVERAGE RETRIEVAL RATES OVER THE WHOLE DATABASE FOR
DIFFERENT METHODS IN FIG. 5

| | |
|---|---|
| $L^1 + L^2$ | 64.83 |
| GGD | 75.73 |
| scalar WD-HMM | 76.51 |
| vector WD-HMM | 80.05 |

*nonrotated* set. Thus, the rotation invariant model is indeed insensitive to the orientation of the analyzed image. The results obtained by exploiting the rotation invariance are very conclusive. The average retrieval rate for the *rotated* set improves by 36.68% when the rotation invariance is effective. The improvement is more striking for the strong directional textures.

## VII. CONCLUSION AND DISCUSSION

We have introduced a new statistical model for images, the vector WD-HMM, as an extension of the WD-HMM from 1-D to 2-D. The model captures both the subband marginal distributions and the dependencies of wavelet coefficients across scales and orientations. By applying the vector WD-HMM to the steerable pyramid, we obtain a steerable model that can be diagonalized to become rotation invariant. To facilitate the use of WD-HMMs in the image retrieval application, we derive a fast algorithm to approximate the Kullback–Leibler distance between two WD-HMMs. Experimental results indicate that the new WD-HMM improves the texture retrieval performance compared to the independent subband model. The rotation invariant property was also tested, and results obtained were consistent with the theory.

## ACKNOWLEDGMENT

## REFERENCES

[1] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 1186–1191, Nov. 1993.

[2] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structure wavelet transform," *IEEE Trans. Image Processing*, vol. 2, pp. 429–441, Oct. 1993.

[3] J. R. Smith and S.-F. Chang, "Transform features for texture classification and discrimination in large image databases," in *Proc. IEEE ICIP*, 1994.

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

[5] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 837–842, Aug. 1996.

[6] G. V. Wouwer, P. Scheunders, and D. V. Dyck, "Statistical texture characterization from discrete wavelet representations," *IEEE Trans. Image Processing*, vol. 8, pp. 592–598, Apr. 1999.

[7] T. Randen and J. H. Husoy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 291–310, Apr. 1999.

[8] 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. Pattern Anal. Machine Intell.*, vol. 22, pp. 1349–1380, Dec. 2000.

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

[10] ——, "Texture similarity measurement using Kullback–Leibler distance on wavelet subbands," in *Proc. IEEE ICIP*, Vancouver, BC, Canada, Sep. 2000.

[11] M. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based signal processing using hidden Markov models," *IEEE Trans. Signal Processing, Special Issue on Wavelets and Filterbanks*, vol. 46, pp. 886–902, Apr. 1998.

[12] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory, Special Issue on Wavelet Transforms and Multiresolution Signal Analysis*, vol. 38, pp. 587–607, March 1992.

[13] R. L. Kashyap and A. Khotanzed, "A model-based method for rotation invariant texture classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 472–481, July 1986.

[14] J.-L. Chen and A. Kundu, "Rotational and gray-scale transform invariant texture identification using wavelet decomposition and hidden Markov model," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 208–214, Feb. 1994.

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

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

[17] W.-R. Wu and S.-C. Wei, "Rotation and gray-scale transform-invariant texture classification using spiral resampling, subband demposition, and hidden Markov model," *IEEE Trans. Image Processing*, vol. 5, pp. 1423–1434, Oct. 1996.

[18] S. R. Fountain, T. N. Tan, and K. D. Baker, "Comparative study of rotation invariant classification and retrieval of texture images," in *Proc. British Computer Vision Conf.*, 1998.

[19] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. IEEE ICIP*, Washington, DC, 1995.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[22] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[23] H. Choi and R. G. Baraniuk, "Image segmentation using wavelet-domain classification," in *Proc. SPIE, Math. Model., Bayesian Estim. Inv. Probl.*, vol. 3816, 1999, pp. 306–320.

[24] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," in *Proc. of SPIE, Math. Model., Bayesian Estim. Inv. Probl.*, vol. 3816, 1999, pp. 306–320.

[25] E. P. Simoncelli and J. Portilla, "Texture characterization via joint statistics of wavelet coefficient magnitudes," in *Proc. IEEE ICIP*, 1998.

[26] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. New York: Academic, 1999.

[27] G. Fan and X.-G. Xia, "Maximum likelihood texture analysis and classification using wavelet-domain hidden Markov models," in *Proc. 34th Asilomar Conf. Signals, Systems, and Computers*, 2000.

[28] J. R. Bergen and E. H. Adelson, "Theories of visual texture perception," in *Spatial Vision*, D. Regan, Ed. Boca Raton, FL: CRC Press, 1991.

[29] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[30] B. H. Juang and L. R. Rabiner, "A probabilistic distance measure for hidden Markov models," *AT&T Tech. J.*, vol. 64, pp. 391–408, Feb. 1985.

[31] B. Hendricks, H. Choi, and R. G. Baraniuk, "Analysis of texture segmentation using wavelet-domain hidden Markov trees," in *Proc. 33rd Asilomar Conference*, Pacific Grove, CA, Oct. 1999.

[32] M. N. Do, "Fast approximation of Kullback–Leibler distance for dependence trees and hidden Markov models," *IEEE Signal Processing Lett.*, to be published.

[33] Y. Singer and M. K. Warmuth, "Batch and on-line parameter estimation of Gaussian mixtures based on the joint entropy," in *Advances in Neural Information Processing Systems 11 (NIPS'98)*, 1998, pp. 578–584.

[34] MIT Vision and Modeling Group, Cambridge, MA. Vision texture. [Online]. Available: http://vismod.www.media.mit.edu.

[35] Univ. Southern California, Signal and Image Processing Institute, Los Angeles, MA. Rotated textures. [Online]. Available: http://sipi.usc.edu/services/database/Database.html.

**Minh N. Do** (S'01–M'02) was born in Thanh Hoa, Vietnam, in 1974. He received the B.Eng. degree in computer engineering (with first class honors) from the University of Canberra, Australia, in 1997, and the Dr.Sci. degree in communication systems from the Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland, in 2001.

Since 2002, he has been an Assistant Professor with the Department of Electrical and Computer Engineering and a Research Assistant Professor with the Coordinated Science Laboratory and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign. His research interests include wavelets and harmonic analysis, multirate systems, image and multidimensional signal processing, and visual information representation and retrieval.

Dr. Do received a Silver Medal from the 32nd International Mathematical Olympiad in Sweden (1991), a University Medal from the University of Canberra (1997), the best doctoral thesis award from the Swiss Federal Institute of Technology Lausanne (2001), and a Career award from the National Science Foundation (2003).
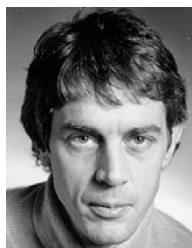
**Martin Vetterli** (M'86–SM'90–F'95) received the Dipl. El.-Ing. degree from ETH Zürich (ETHZ), Switzerland, in 1981, the M.S. degree from Stanford University, Stanford, CA, in 1982, and the Dr. Sci. degree from EPF Lausanne (EPFL), Switzerland, in 1986.

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, 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 Department 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 (1996–1997), and heads of the Audio-Visual Communications Laboratory. He held visiting positions at ETHZ (1990) and Stanford (1998). He is on the editorial boards of *Annals of Telecommunications, Applied and Computational Harmonic Analysis*, and *The Journal of Fourier Analysis and Applications*. He and is the co-author, with J. Kovacevic, of the book *Wavelets and Subband Coding* (Englewood Cliffs, NJ: Prentice-Hall, 1995). He has published about 85 journal papers on a variety of topics in signal and image processing and holds five patents. His research interests include wavelets, multirate signal processing, computational complexity, signal processing for telecommunications, digital video processing and compression and wireless video communications.

Dr. Vetterli is a member of SIAM, and was the Area Editor for Speech, Image, Video, and Signal Processing of the IEEE TRANSACTIONS ON COMMUNICATIONS. 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 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).