Proc. of SPIE Conf. on Wavelet Applications in Signal and Image Processing VIII, San Diego, 2000

Rotation Invariant Texture Retrieval using Steerable Wavelet-domain Hidden Markov Models

Minh N. Do*, Aurélie C. Lozano* and Martin Vetterli*†

*Laboratory for Audio-Visual Communications (LCAV)
Swiss Federal Institute of Technology Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

†Department of Electrical Engineering and Computer Science
University of California Berkeley, Berkeley CA 94720, USA

Email: {Minh.Do, Aurelie.Lozano, Martin.Vetterli}@epfl.ch; Web: lcavwww.epfl.ch

ABSTRACT

A new statistical model for characterizing texture images based on wavelet-domain hidden Markov models and steerable pyramids is presented. The new model is shown to capture well both the subband marginal distributions and the dependencies across scales and orientations of the wavelet descriptors. Once it is trained for an input texture image, the model can be easily steered to characterize that texture at any other orientations. After a diagonalization operation, one obtains a rotation-invariant description of the texture image. The effectiveness of the new model is demonstrated in large test image databases where significant gains in retrieval performance are shown.

1. INTRODUCTION

With the explosive growth of multimedia databases and digital libraries, there is high demand for efficient tools which 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 well-known texture analysis methods in image retrieval application are wavelet-liked approaches.[1–4] Essentially, those methods measure energies (possibly weighted) at the output of filter banks as extracted features for texture discrimination. The basic assumption for those approaches is that the energy distribution in the frequency domain identifies a texture.

In a previous paper,[5] we proposed a general statistical framework for the image retrieval problem by jointly considering two related tasks: Feature Extraction (FE) and Similarity Measurement (SM); while keeping in mind the complexity constraint of such application. In this approach, 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 model parameters. In particular, by modeling the marginal distribution of wavelet coefficients at a given subband by a generalized Gaussian density (GGD) we obtained significant improvements in retrieval rates compared to the traditional, energy-based approaches. The motivation behind this approach comes from the finding of psychological research on human texture perception[6] which suggests that two homogeneous textures are often difficult to discriminate if they produce similar marginal distributions of responses from a bank of filters.

While having low complexity, the marginal distribution model ignores some important texture-specific information, notably the dependency 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 the same textures 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-band dependencies while being rotation invariant. The proposed model uses a wavelet domain hidden Markov tree[7] with steerable pyramids.[8] Rotation invariance is achieved via a diagonalization of the covariance matrices in the model.

1.1. Main contributions

Firstly, this work enhances the recent technique on wavelet-domain hidden Markov models (WD-HMM)[7] for better dealing with images by incorporating the dependency of wavelet coefficients across orientations. Furthermore, by replacing the standard wavelet transform with an overcomplete representation via steerable pyramids,[9] we obtain a steerable statistical model that can facilitate rotation invariant applications. WD-HMM's have been applied in the texture segmentation problem,[10] where only a small number of texture classes (typically a few regions) are classified. Our experiments with WD-HMM's in the image retrieval application provide a large scale test on their capacity in discriminating among many different texture classes.

1.2. Related works

Several authors have developed rotation invariant texture features, mostly for the classification problem – where the classes are defined *a priori*. Chen and Kundu[11] 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.[12] and Haley and Manjunath[13] used the magnitude of a discrete Fourier transform in the rotation dimension of features obtained from a multiresolution filtering. Wu and Wei[14] achieved rotation invariance by first converting 2-D texture images to 1-D signal via spiral sampling. A subband decomposition is applied to the 1-D signal and follows by an HMM on the subband features.

2. WAVELET-DOMAIN HIDDEN MARKOV MODELS

Recently, Crouse et. al.[7] proposed a new framework for statistical signal processing based on wavelet-domain hidden Markov models (WD-HMM's). 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 below together with the introduction of the notations.

In a WD-HMM, to each wavelet coefficient W_i there is an associated discrete hidden state S_i with the probability mass function $Pr(S_i = m) = p_i^m, form = 1, \dots, M$. Conditioning on $S_i = m$, W_i follows a Gaussian density with the mean $\mu_{i,m}$ and the variance $\sigma_{i,m}^2$. Since the wavelet coefficients are resulted from convolutions with filters that have zero sum (the wavelet high-pass filters), they can be considered 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. So, if we take $M = 2$, the marginal distribution wavelets coefficients at the j-th subband can be written as a mixture of two Gaussian densities

$$f_j(w) = p_j^1 g(w; \sigma_{j,1}) + p_j^2 g(w; \sigma_{j,2}), \tag{1}$$

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

In this hidden Markov modeling, p_j^1 and p_j^2 have physical interpretation as the probability that a wavelet coefficient W at the level j is in the state of being small and large, respectively. Small coefficients can be considered as outcomes of a small variance probability density function whereas large ones can be considered as outcomes of a large variance density.

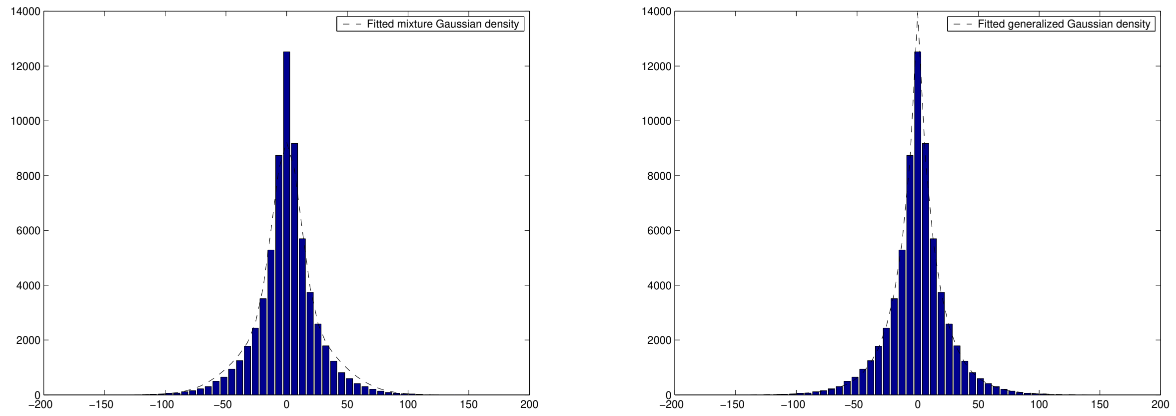
There is an inter-scale dependency, most notably between a wavelet coefficient at a coarse level (parent) to the four coefficients at the next intermediate level that correspond to the same location (children) in the image (see Figure 2(a)). In order to capture this persistence across scales, there are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

$$\mathsf{A}_j = \begin{bmatrix} p_j^{1\to1} & p_j^{1\to2} \\ p_j^{2\to1} & p_j^{2\to2} \end{bmatrix}, \qquad j = 2, 3, \dots, J. \tag{2}$$

Here $p_j^{m\to m'}$ is the probability that a child coefficient at the level j will be in state m' given its parent coefficient is in state m. It is clear that A_j has row sums equal to 1.

Denote $\rho(i)$ the parent of the node i in the wavelet coefficient tree then,

$$Pr(S_i = m) = \sum_{m'} Pr(S_{\rho(i)} = m') Pr(S_i = m | S_{\rho(i)} = m').$$



(a) WD-HMM. (b) GGD.

Figure 1. Example of wavelet subband coefficient histogram fitted with the marginal distribution curves by WD-HMM and GGD model.

Thus for the tied models (where all the nodes at the same level j share the same statistics) we have

$$p_j^m = \sum_{m'} p_{j-1}^{m'} p_j^{m' \to m}, \qquad j = 2, 3, \ldots, J. \tag{3}$$

Let's write $\boldsymbol{p}_j = [p_j^1, p_j^2]$, then (3) is equivalent to $\boldsymbol{p}_j = \boldsymbol{p}_{j-1}\mathsf{A}_j$. Hence we have

$$\boldsymbol{p}_j = \boldsymbol{p}_1 \mathsf{A}_2 \mathsf{A}_3 \ldots \mathsf{A}_j, \qquad \forall j = 2, \ldots, J. \tag{4}$$

So the WD-HMM for a tree of wavelet coefficients (also called hidden Markov tree model) is completely defined by a set of model parameter $\Theta = \{\boldsymbol{p}_1, \mathsf{A}_2, \ldots, \mathsf{A}_J;\ \sigma_{j,1}, \sigma_{j,2}\ (j = 1, \ldots, J)\}$. Here 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.[7]

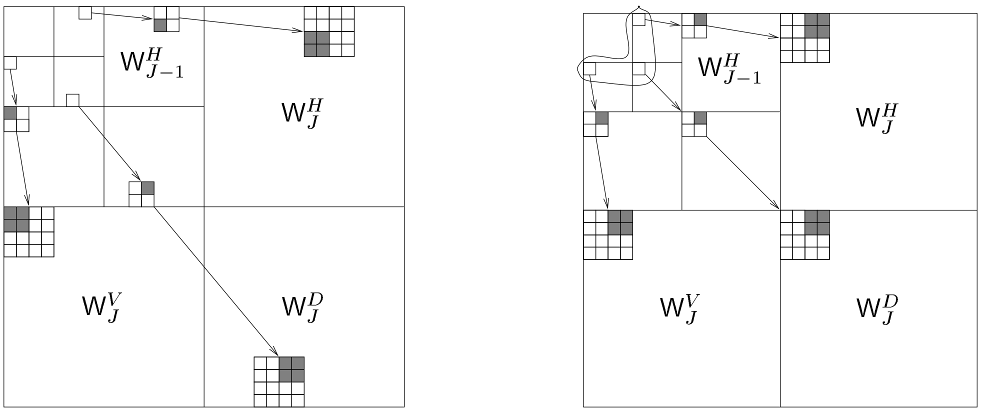
Figure 1 shows a typical example of the histogram of the wavelet coefficients from an image subband together with the plot of the subband marginal distribution density function obtained from the trained WD-HMM. For comparison we also show the fitted GGD using the ML estimator.[5] As can be seen from the figure, the WD-HMM provides a close match to the GGD in term of modeling the marginal distribution from a wavelet subband. However, the WD-HMM is more expressive than the GGD model by adding the dependencies between parent-child coefficients across scales.

3. VECTOR WD-HMM

Originally developed for 1-D signals, the WD-HMM has been generalized for images in segmentation[10] and denoising[15] 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[10,15] took the simple approach by considering these three orientations separately thus requires training three WD-HMM's to characterize an image, one for each orientation. We call those as *scalar* WD-HMM's.

The underlying assumption for the scalar WD-HMM approach is that wavelet coefficients at different orientations are independent. On the other hand, experiments in[16] show the importance of the cross-correlation of each subband with other orientations at the same scale in characterizing texture images.

Thus, 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 vectors and then model them by a single multidimensional WD-HMM (see Figure 2(b)). The result is one *vector* WD-HMM for an input image.



(a) Scalar WD-HMM. (b) Vector WD-HMM.

Figure 2. Tree structures on the WD-HMM's. 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.

More specifically, denote the wavelet coefficients at the orientation d ($d = 1, 2, 3$ for H, V, D respectively) and scale j as $w^d_{j,\boldsymbol{k}}$, $\boldsymbol{k} \in \mathbb{Z}^2$. The grouping operation will produce vectors of coefficients $\boldsymbol{w}_{j,\boldsymbol{k}} = [w^1_{j,\boldsymbol{k}}, w^2_{j,\boldsymbol{k}}, w^3_{j,\boldsymbol{k}}]^T$. 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 basis functions $\psi^d_{j,\boldsymbol{k}}$ at scale j and location $\boldsymbol{k}$,[17]

$$w^d_{j,\boldsymbol{k}} = \langle \psi^d_{j,\boldsymbol{k}}, \boldsymbol{x} \rangle \qquad d = 1, 2, 3.$$

The marginal distribution function of the wavelet coefficient vectors at the level j in the vector WD-HMM (with tying) is expressed as

$$f_j(\boldsymbol{w}) = p^1_j g(\boldsymbol{w}; \mathsf{C}_{j,1}) + p^2_j g(\boldsymbol{w}; \mathsf{C}_{j,2}). \tag{5}$$

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

$$g(\boldsymbol{w}; \mathsf{C}) = \frac{1}{\sqrt{(2\pi)^n |det(\mathsf{C})|}} \exp(-\boldsymbol{w}^T \mathsf{C}^{-1} \boldsymbol{w}), \tag{6}$$

where $n = 3$ is the number of orientations.

The wavelet coefficient vectors are then organized into the same quad-tree structure that connects each vector to its four children at the next intermediate level of the same location (see Figure 2(b)). The parent $\rightarrow$ child link relationships of these vectors are captured the same way as in (2) for the scalar WD-HMM.

So in the vector WD-HMM, wavelet coefficients at the same scale and location but different orientations are tied up to have same hidden state. The justification for this is that around edges in an image, wavelet coefficients at all orientations have high probability of being significant whereas in smooth regions, all wavelet coefficients are small. Additionally, in the vector WD-HMM the across orientation dependencies are captured via the non-diagonal entries of the covariance matrices in the multivariate Gaussian densities (6). It can be verified that the marginal distributions for each wavelet subband in the vector WD-HMM are mixtures of two zero-mean Gaussian densities. Thus one can expect that the vector WD-HMM also captures well the subband dependent marginal probability distributions of wavelet coefficients as the scalar WD-HMM.

3.1. Relations between models

In this section we draw the connections between the generalized Gaussian density (GGD) model used in[18] and the scalar and vector WD-HMM's. As discussed previously, all of these models capture very well 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.[6] In[18] by simply modeling those PDF's by GGD's, we have already obtained very good retrieval results. The scalar WD-HMM adds on an extra texture-specific information by capturing the inter-scale dependencies (via the A_j matrices). The vector WD-HMM furthermore adds on the inter-orientation dependencies information in characterizing textures.

Table 1 shows the number of free parameters needed to describe each image using different models, assuming the wavelet transform is decomposed with J levels. Note that due to the row sums property, each A_j has only 2 free parameters. The covariance matrices C are symmetric thus contain 6 free parameters each.

	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	$6J$	$12J-3$	$14J-1$

Table 1. Number of free parameters needed to specify different models for an image when the wavelet transform is taken to J levels.

4. STEERABLE WD-HMM

Both the scalar and the vector WD-HMM's 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 (except for the rotations of $k\pi/2,\ k \in \mathbb{Z}$). In fact, the wavelet coefficients of the rotated image are *not* just be simply rotated but also be modified.

One way to remedy this situation is to replace the standard wavelet decomposition with a steerable pyramid.[8,9] The steerable pyramid is a linear multi-scale, multi-orientation image decomposition where the basis functions are directional derivative operators. This transform satisfies the *shiftability* in orientation condition which states that at a fixed location and scale, one can compute responses at any orientation via a linear combination of coefficients corresponding to the oriented basis functions at that location and scale. As a bonus, the steerable pyramid representation is also translation-invariant.

Therefore, if we denote $\boldsymbol{w}$ and $\boldsymbol{w}^{\phi}$ as the vectors of steerable pyramid coefficients at fixed scale j and location $\boldsymbol{k}$ for an input image and its rotated copy by ϕ respectively then

$$\boldsymbol{w}^{\phi} = \mathsf{R}_{\phi}\boldsymbol{w}. \tag{7}$$

The columns of R_{ϕ} are a set of interpolation functions that depend only on the rotation angle ϕ and the steerable basis functions. Furthermore, orientation shiftability ensures the orientation invariance of response power,[8] i.e. $||\boldsymbol{w}^{\phi}||^2 = ||\boldsymbol{w}||^2$ for any $\boldsymbol{w}$. This is equivalent[19] to saying that R_{ϕ} is an orthogonal matrix, i.e. $\mathsf{R}_{\phi}^{-1} = \mathsf{R}_{\phi}^{T}$.

The following proposition shows that for the vector WD-HMM on a steerable pyramid, if the input image is rotated then the new model can be obtained by a simple transformation of the original WD-HMM.

PROPOSITION 1. *Suppose that* $\Theta = \{\boldsymbol{p}_1, \mathsf{A}_2, \ldots, \mathsf{A}_J;\ \mathsf{C}_{j,i}\ (j = 1, \ldots, J;\ i = 1,2)\}$ *is the vector WD-HMM on a steerable pyramid for a given homogeneous texture class. Then the corresponding model for the rotated version of that texture by* ϕ *is* $\Theta^{\phi} = \{\boldsymbol{p}_1, \mathsf{A}_2, \ldots, \mathsf{A}_J;\ \mathsf{C}^{\phi}_{j,i}\ (j = 1, \ldots, J;\ i = 1,2)\}$. *The only change are for the covariance matrices, precisely by*

$$\mathsf{C}^{\phi}_{j,i} = \mathsf{R}_{\phi}\mathsf{C}_{j,i}\mathsf{R}^{T}_{\phi}, \qquad j = 1, \ldots, J;\ i = 1, 2. \tag{8}$$

Proof: The marginal distribution function of the coefficient vectors $\boldsymbol{w}^\phi$ at the level j of the rotated texture can be written as

$$\begin{aligned} f_j^\phi(\boldsymbol{w}^\phi) &= \frac{1}{J(\boldsymbol{w}^\phi, \boldsymbol{w})} f(\mathsf{R}_\phi^{-1}\boldsymbol{w}^\phi) \\ &= p_j^1 g(\mathsf{R}_\phi^{-1}\boldsymbol{w}^\phi; \mathsf{C}_{j,1}) + p_j^2 g(\mathsf{R}_\phi^{-1}\boldsymbol{w}^\phi; \mathsf{C}_{j,2}), \end{aligned}$$

since the Jacobian $J(\boldsymbol{w}^\phi, \boldsymbol{w}) = |det(\mathsf{R}_\phi)| = 1$. Using the fact that R_ϕ is an orthogonal matrix again, by manipulating (6) we have

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

So $f_j^\phi(\boldsymbol{w}^\phi)$ 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 related by Equation (8). Combining this across scales we obtain the desired result. □

Thus 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 to describe that texture at any other orientations.

4.1. Rotation invariant model

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-HMM's Θ and Θ^ϕ of a given texture and its rotated version are among the covariance matrices. These covariance matrices are related by (8): $\mathsf{C}_{j,i}$ and $\mathsf{C}_{j,i}^\phi$ are said to be orthogonally equivalent.[19]

Using the Takagi's factorization,[19] we factor each covariance matrix in the steerable WD-HMM into a product

$$\mathsf{C}_{j,i} = \mathsf{U}_{j,i}\Lambda_{j,i}\mathsf{U}_{j,i}^T, \qquad j = 1, \dots, J;\ i = 1, 2, \tag{9}$$

where $\mathsf{U}_{j,i}$ is the orthogonal matrix whose columns are the normalized eigenvectors of $\mathsf{C}_{j,i}$ and $\Lambda_{j,i}$ is the diagonal matrix containing the real, nonnegative eigenvalues of $\mathsf{C}_{j,i}$ in descending order. This factorization is always possible since all covariance matrices are symmetric and positive semidefinite.

Let λ be an eigenvalue of $\mathsf{C}_{j,i}^\phi$. That means there exists a vector $\boldsymbol{v}$ satisfies $\mathsf{C}_{j,i}^\phi \boldsymbol{v} = \lambda \boldsymbol{v}$. Using (8) we have

$$\mathsf{R}_\phi \mathsf{C}_{j,i} \mathsf{R}_\phi^T \boldsymbol{v} = \lambda \boldsymbol{v} \qquad \text{or} \qquad \mathsf{C}_{j,i} \mathsf{R}_\phi^T \boldsymbol{v} = \lambda \mathsf{R}_\phi^T \boldsymbol{v}.$$

So if we let $\boldsymbol{v}' = \mathsf{R}_\phi^T \boldsymbol{v}$ then $\mathsf{C}_{j,i}\boldsymbol{v}' = \lambda \boldsymbol{v}'$ hence λ is also an eigenvalue of $\mathsf{C}_{j,i}$. Thus the diagonalization operation on the rotated model Θ^ϕ leads to

$$\Lambda_{j,i}^\phi = \Lambda_{j,i}, \qquad \forall j = 1, \dots, J;\ i = 1, 2.$$

In summary, given a steerable WD-HMM, if we factorize the covariance matrices into the form (9) then the $\mathsf{U}_{j,i}$ matrices are responsible for the orientation of the input image while the $\Lambda_{j,i}$ matrices contain rotation-invariant texture information. Thus we have the following result.

PROPOSITION 2. *The* diagonalized *steerable WD-HMM* $\bar{\Theta} = \{\boldsymbol{p}_1, \mathsf{A}_2, \dots, \mathsf{A}_J;\ \Lambda_{j,i}\ (j = 1, \dots, J;\ i = 1, 2)\}$ *is a rotation-invariant model.*

Remark: 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.

5. TEXTURE SIMILARITY MEASUREMENT USING WD-HMM'S

In a simple term, the task of the image retrieval problem is given a query image Q to search for the top M most similar images from an image database $\{I_1, ..., I_N\}$. This requires the computation of the (dis-)similarity measurements $d(Q, I_i)$ between the query Q and each image I_i in the database. All of the statistical models described above allow us to represent compactly each texture image I_i in the database by its model parameter Θ_i. Those model parameters can be computed beforehand (off-line) – for example when the images are inserted into the database. In this statistical setting, there are two ways of computing the similarity measurements.

5.1. Likelihood method

Under the multiple hypothesis setting, the optimal maximum likelihood selection rule relates the measurement $d(Q, I_i)$ with the likelihood of the model Θ_i of I_i given the data from Q. That means we compute the likelihoods of the wavelet coefficients of the query image against all WD-HMM's Θ_i in the database; the images correspond to the models with highest likelihoods are returned as retrieved images. Thus we rank the relevance of the image I_i in the database to the query Q base on

$$d(Q, I_i) = -Pr(Q|\Theta_i). \tag{10}$$

The main drawback of this method is that it uses the raw data of Q despite the fact that a much more compact representation, namely its model, may be already computed. This is often the case when the query image is one of the images in the database. This consideration leads to the use of a second method.

5.2. Kullback-Leibler distance

In[5] we show that the likelihood method is asymptotically equivalent to use the Kullback-Leibler distance (KLD) between the models of the query and the candidate image as,

$$d(Q, I_i) = D(\Theta_q||\Theta_i) = \int f(\boldsymbol{w}|\Theta_q) \log\left(\frac{f(\boldsymbol{w}|\Theta_q)}{f(\boldsymbol{w}|\Theta_i)}\right) d\boldsymbol{w} \tag{11}$$

An additional advantage of the KLD method is that since this similarity measurement is defined directly on the extracted model's parameters so for the rotation-invariant models it leads to a rotation-invariant image retrieval system.

Unfortunately, we have not yet found a closed form expression for the KLD between WD-HMM's. Thus we have to resort to a Monte-Carlo method. In fact, from the model Θ_q we randomly generate data $\tilde{Q}$ and compute the likelihood as done in (10).

6. EXPERIMENTAL RESULTS

We use two texture collections in our experiments. In a first series of experiments, we evaluate the retrieval effectiveness of both scalar and multidimensional WD-HMM against the GDD method in a large database. We used 40 textures obtained from the MIT Vision Texture (VisTex) database.[20] The names of the chosen textures are listed in Table 2. Each of these 512 × 512 images was divided into sixteen 128 × 128 non-overlapping subimages, thus creating a test database of 640 texture images. Only gray-scale levels of the images (computed from the luminance component) were used in the experiments.

The second image collection is used to test the rotation-invariant property of different WD-HMM's. It consists of thirteen 512 × 512 Brodatz textures images that were rotated to various degrees *before* being digitized.[21] Figure 3 displays the original textures at the non-rotated position. From these images we first construct the *non-rotated* image database by dividing each 0° version of the original textures into sixteen 128 × 128 non-overlapping subimages. Next we construct the *rotated* image database by taking four non-overlapping 128 × 128 subimages each from the original images at 0, 30, 60, and 120 degrees. Both databases contain 208 images that come from 13 texture classes. The *"non-rotated set"* serves as the ideal case, where all images in a same class have same orientation, for the *"rotated set"*.

In our retrieval experiments, each image in the database is simulated as a query. The relevant images for each query are defined as the other 15 subimages from the same original VisTex image. Following,[4] the retrieval performance is evaluated in term of the percentage of relevant images among the top 15 retrieved images.

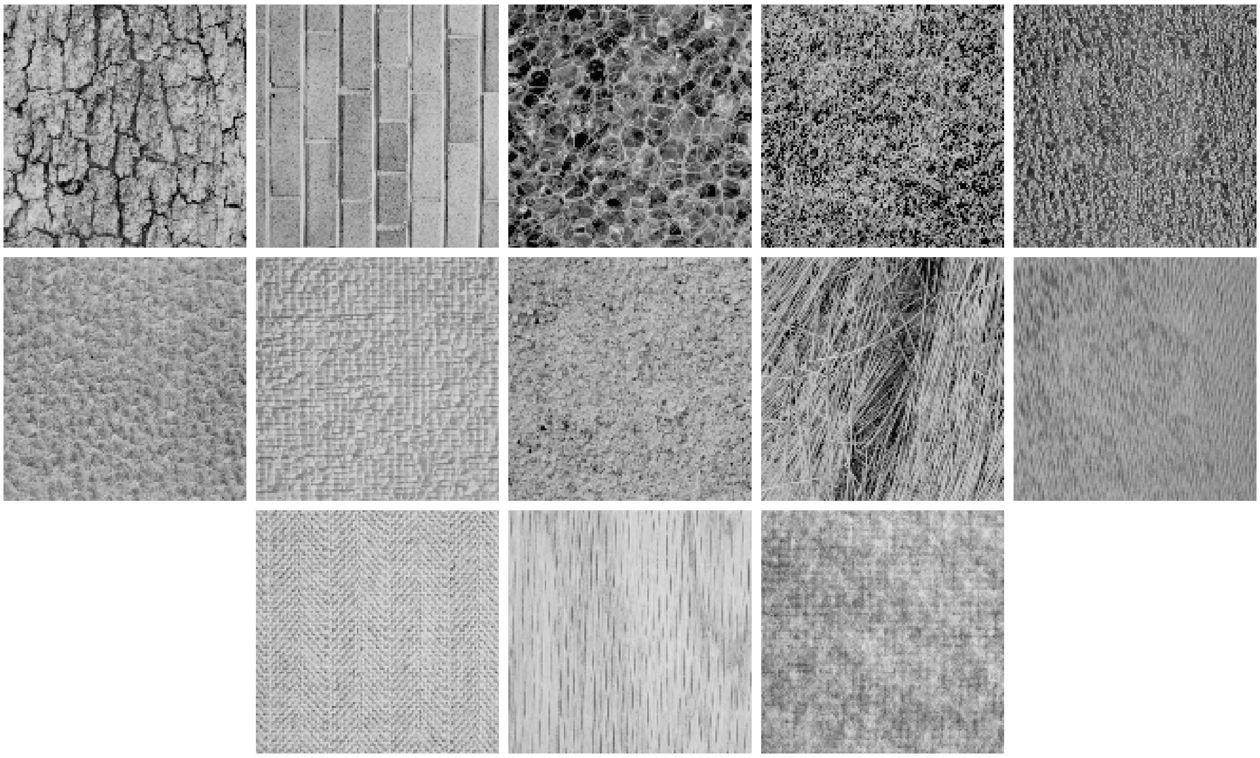


Figure 3. Texture images from the rotate collection that are used in the experiments. The images are at 0° degree. From left to right and top to bottom are: Bark, Brick, Bubbles, Grass, Leather, Pigskin, Raffia, Sand, Straw, Water, Weave, Wood, and Wool.

6.1. Effectiveness of WD-HMM's

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

We start first with the database of 640 textures images from VisTex. By looking at Table 2 we notice that the scalar WD-HMM method gives compatible result with the GGD method whereas the vector WD-HMM method significantly improves the retrieval rates in many texture classes as well on average. If we focus on the WD-HMM methods, we can see that the vector model almost always outperforms the scalar one. In fact it is consistent with the argument in Section 3: the vector parameters are more precise in characterizing textures since they add on the consideration of the inter-orientation dependency information.

However, when using the "non-rotated" database, we can see in Figure 4 that for Brick the vector WD-HMM method gives worse results than the scalar WD-HMM method. Our explanation for this is that Brick images have mainly vertical textures patterns. Therefore,in this case, there is no correlation across orientations. Only the vertical wavelet subbands contain significant coefficients corresponding to the "stripes" of Brick, whereas wavelet coefficients at all other subbands are very small. By tying the three subbands together, we are dramatically reducing the impact of the coefficients of the vertical subband. Therefore, tying within orientation could reduce the performances.

Figure 5 shows the KLD between two models according to the number of randomly generated trees used to compute the distance. The KLD converge clearly and the variance decreases as the percentage of nodes used increases. Therefore it is sufficient to use a certain part of trees to have accurate results. In our experiments we noticed that after 64 random generated trees used, the results did not much vary anymore.

Texture Class	Methods			Texture Class	Methods		
	GGD	scalar WD-HMM	vector WD-HMM		GGD	scalar WD-HMM	vector WD-HMM
Bark0	57.03	56.67	64.58	Food8	91.80	78.75	95.83
Bark6	55.86	51.25	62.50	Grass1	71.48	74.58	65.83
Bark8	84.38	90.00	94.17	Leaves8	51.56	70.42	68.75
Bark9	70.70	53.75	70.00	Leaves10	32.81	55.42	39.17
Brick1	76.95	79.17	85.42	Leaves11	67.97	77.50	83.75
Brick4	71.88	61.67	75.00	Leaves12	75.78	80.83	82.92
Brick5	81.25	90.83	90.83	Leaves16	73.83	81.25	72.50
Buildings9	92.58	81.67	85.00	Metal0	75.78	80.00	80.83
Fabric0	98.83	95.83	88.75	Metal2	96.88	100.00	100.00
Fabric4	60.16	60.42	64.17	Misc2	77.34	89.58	87.92
Fabric7	97.27	95.42	100.00	Sand0	77.73	70.83	85.83
Fabric9	82.03	89.17	85.00	Stone1	33.59	53.33	56.67
Fabric11	78.52	87.50	85.83	Stone4	81.25	88.75	90.00
Fabric14	100.00	100 00	100.00	Terrain10	54.30	45.83	47.92
Fabric15	98.05	100.00	88.33	Tile1	51.56	58.33	72.92
Fabric17	68.36	96.67	100.00	Tile4	95.31	88.33	93.75
Fabric18	98.83	97.08	100.00	Tile7	100.00	42.08	91.67
Flowers5	55.47	63.75	71.67	Water5	91.41	92.08	95.83
Food0	96.09	97.92	96.67	Wood1	35.16	35.83	25.00
Food5	80.47	84.58	94.17	Wood2	89.06	63.33	62.92
				Average	**75.73**	**76.51**	**80.05**

Table 2. Average retrieval rates for individual texture class using standard wavelet transform with Haar filters and 3 decomposition levels.

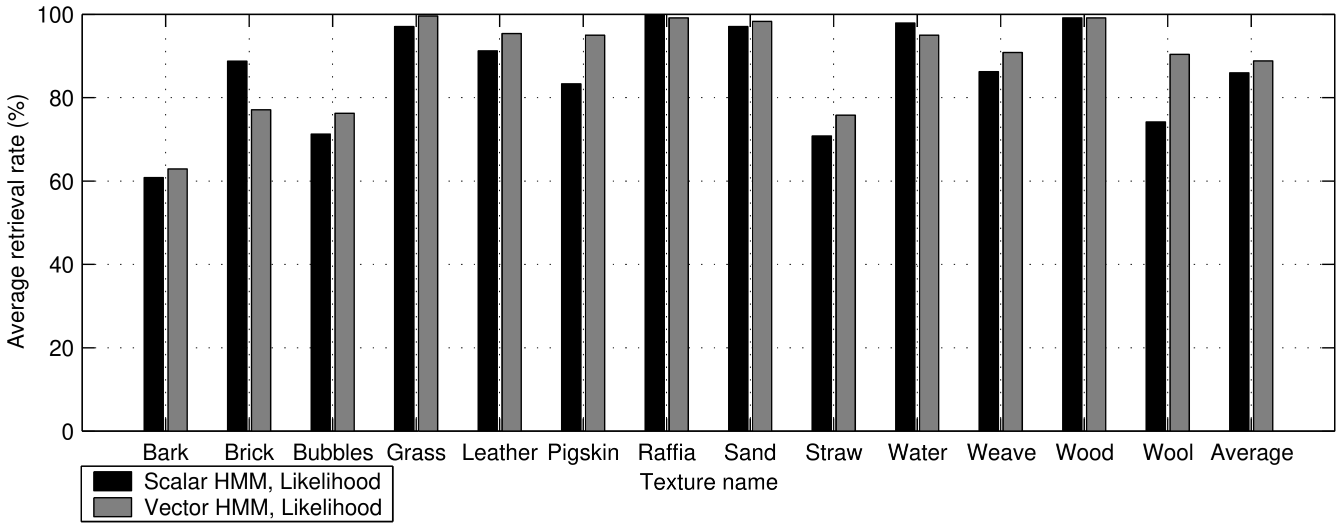


Figure 4. Average retrieval rates for individual texture class using Haar filter, 3 decomposition level and 25% of random generated trees for scalar and vector WD-HMM

6.2. Effect on the choice of the wavelet and the similarity measurements

Table 3 summarizes the performance when different filters and similarity measurements are used in the experiment. We restrain our experiments to Haar filters, since Table 3 shows that the smallest the filter length is, the better are the results. One can see that KLD performances are very close to those obtained by using the likelihood, despite the fact we have not found yet a closed form to compute the KLD and we are using an approximation instead.

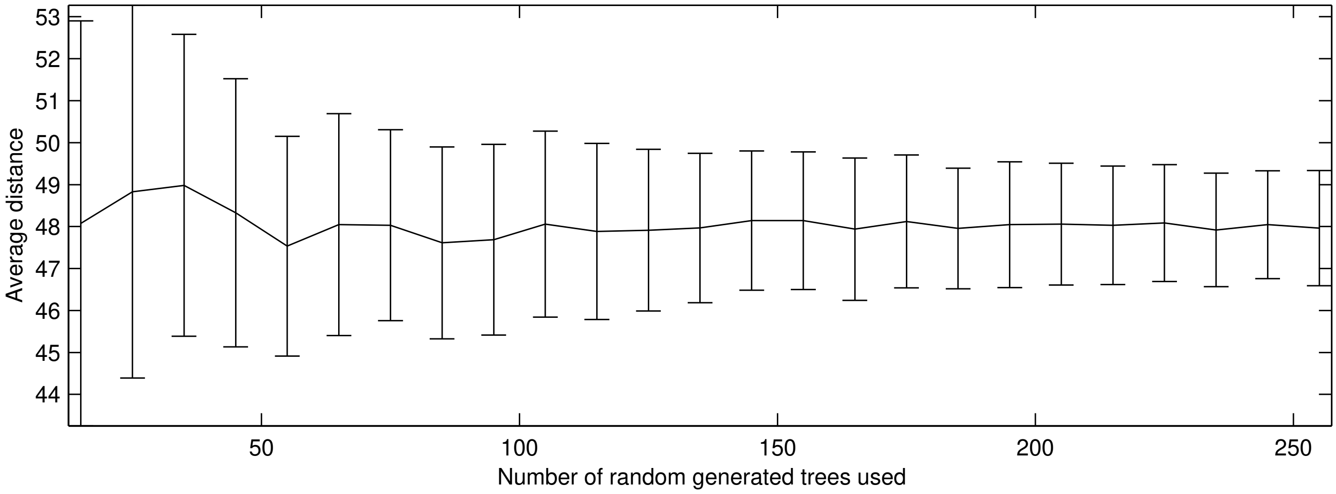


Figure 5. Average distance between leather01 and raffia01 with error bars specified by the standard deviation, using KLD.

Texture	scalar methods		vectorial methods	
Filter	*Likelihood*	*KLD*	*Likelihood*	*KLD*
Daub2	85.99	85.71	88.85	86.28
Daub4	83.27	82.66	85.32	83.11
Daub6	79.42	77.18	82.76	77.37
Daub8	76.31	75.96	78.30	74.17

Table 3. Average retrieval rates for "non rotated" set using standard wavelet transform with Daubechies filters of length 2, 4, 6 and 8, with 3 decomposition levels and 64 random generated trees (or 25% from the query image).

6.3. Effectiveness of rotation invariance

In the last series of experiments, we used steerable pyramids to benefit from the rotation property that they offer (refer to Section 4). We chose steerable pyramid having two directions and three levels of decomposition.

Figure 6 shows comparison of the performances in average percentages of retrieving relevant images for the non rotated set, the rotated set without rotation invariance and the rotated set with rotation invariance. Firstly, one can analyze the retrieval results obtained for the rotated set without rotation invariance and for the non-rotated set. We can make the following remarks. Textures that give quite the same results for both sets are those that have no strong direction (Bark, Bubbles, Grass, Weave). Therefore, those textures are not affected by rotation. Moreover all have very distinct texture patterns. Textures that are most seriously affected by rotation are the one which are strongly directional (Raffia, Wood, Water, Straw).

Secondly if we have a look at the retrieval results obtained for the rotated set with rotation invariance, as we expected, results are quite similar to those for the non-rotated set. The improvement is more striking for the strong directional textures. So 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.

However some details need to be underlined. For the textures Bark and Weave, we notice that the results obtained for the rotated image without rotation invariance are better than those of the non rotated set. Moreover, the average result for the rotated set with rotation invariance is better than for the non rotated set. We must keep in mind that each image of the non-rotated set is non-homogeneous. Therefore, as the images are divided into sub-images, it can happen that on sub-image can have a different direction than another even if both come from a unique texture. This can be easily observed by looking at Straw for instance. The rotated set with rotation invariance does not suffer from those differences in direction, therefore the results are better.

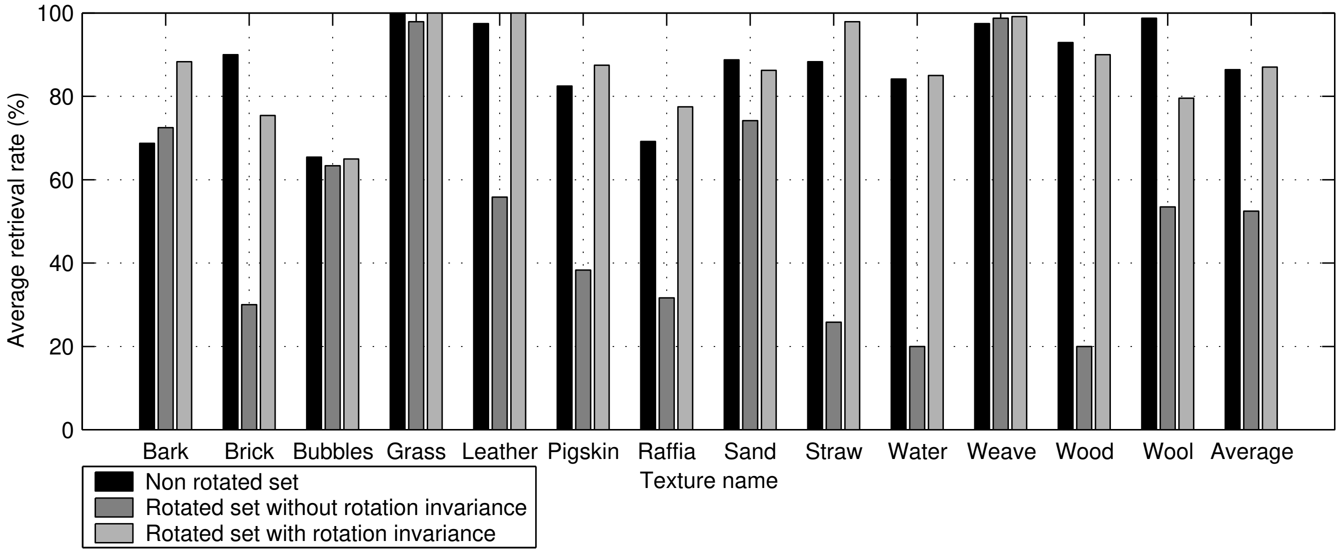


Figure 6. Average retrieval rates for individual texture class using order 1 steerable filters, 3 decomposition levels and 25% of random generated trees for respectively non-rotated set, rotated set without rotation invariance and rotated set with rotation invariance.

7. CONCLUSION AND FURTHER WORK

We have introduced a new texture model that is based on WD-HMM. The model is shown to capture well both the subband marginal distributions and the dependencies of wavelet descriptors across scales and orientations. Furthermore unlike its predecessor, the new model is rotation invariant. This is an useful property for the image retrieval applications and also for many other applications.

Experimental results indicate that the new method indeed gives higher texture retrieval performance than other wavelet-based approaches. The rotation invariant property was also tested and gave results consistent with the theory.

One of the main disadvantage of the WD-HMM for the image retrieval applications is the missing of a closed form for the KLD between two WD-HMM's. One way we plan to overcome this is to derive some lower bounds for the KLD. Those lower bounds can be used to quickly eliminate a large set of textures that are very different with the query image. Then with the remaining candidate images, finer estimation of KLD can be used.

Another remark is to notice a slight decrease in performance of the WD-HMM using steerable pyramid compared with the wavelet transform using short Daubechies filters. An explanation for this is that the used filters in the steerable pyramids are too smooth so that most of the texture information is down to the low-pass subband which is ignored in our models. With new steerable filters that are designed specifically for our application, higher performance could be obtained.

Acknowledgment

This work was supported in part by a Department of Communication Systems, EPFL PhD Fellowship and the Swiss National Science Foundation grant number 21-52439.97.

REFERENCES

1. A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Patt. Anal. Machine Intell.* **15**, pp. 1186–1191, 1993.
2. 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.
3. M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE. Trans. Image Proc.* **4**, pp. 1549–1560, 1995.

4. B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Patt. Recog. and Mach. Intell.* **18**, pp. 837–842, August 1996.
5. M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance." submitted to IEEE Trans. on Image Proc., December 1999.
6. J. R. Bergen and E. H. Adelson, "Theories of visual texture perception," in *Spatial Vision*, D. Regan, ed., CRC press, 1991.
7. M. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based signal processing using hidden Markov models," *IEEE Trans. Signal Proc. (Special Issue on Wavelets and Filterbanks)* , pp. 886–902, April 1998.
8. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Info. Theory* **38**, pp. 587–607, Mar. 1992.
9. E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. of IEEE Int'l Conf. on Image Processing*, (Washington DC), 1995.
10. H. Choi and R. G. Baraniuk, "Image segmentation using wavelet-domain classification," in *Proc. of SPIE Conf. on Mathematical Modeling, Bayesian Estimation, and Inverse Problems*, vol. 3816, pp. 306–320, 1999.
11. J.-L. Chen and A. Kundu, "Rotational and gray-scale transform invariant texture identification using wavelet decomposition and hidden Markov model," *IEEE Trans. Patt. Recog. and Mach. Intell.* **16**, pp. 208–214, Feb. 1994.
12. H. Greenspan, S. Belongie, R. Goodman, and P. Perona, "Rotation invariant texture recognition using a steerable pyramid," in *Proc. Int. Conf. on Pattern Recognition*, pp. 162–167, 1994.
13. G. M. Haley and B. S. Manjunath, "Rotation-invariant texture classification using a complete space-frequency model," *IEEE Trans. Image Proc.* **8**, pp. 255–269, Feb. 1999.
14. 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 Proc.* **5**, pp. 1423–1434, Oct. 1996.
15. J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden markov models," in *Proc. of SPIE Conf. on Mathematical Modeling, Bayesian Estimation, and Inverse Problems*, vol. 3816, pp. 306–320, 1999.
16. 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.
17. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2nd ed., 1999.
18. M. N. Do and M. Vetterli, "Texture similarity measurement using Kullback-Leibler distance on wavelet subbands," in *Proc. IEEE Int. Conf. on Image Proc. (ICIP)*, (Vancouver, Canada), Sep. 2000.
19. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
20. MIT Vision and Modeling Group, "Vision texture." `http://vismod.www.media.mit.edu`, 1995.
21. University of Southern California, Signal and Image Processing Institute, "Rotated textures." `http://sipi.usc.edu/services/database/Database.html`.