Directional Multiscale Modeling of Images using the Contourlet Transform

Duncan D.-Y. Po and Minh N. Do, *Member, IEEE*,

***Abstract*— The contourlet transform is a new two-dimensional extension of the wavelet transform using multiscale and directional filter banks. The contourlet expansion is composed of basis images oriented at various directions in multiple scales, with flexible aspect ratios. Given this rich set of basis images, the contourlet transform effectively captures smooth contours that are the dominant feature in natural images. We begin with a detailed study on the statistics of the contourlet coefficients of natural images: using histograms to estimate the marginal and joint distributions, and mutual information to measure the dependencies between coefficients. This study reveals the highly non-Gaussian marginal statistics and strong inter-location, inter-scale, and inter-direction dependencies of contourlet coefficients. We also find that conditioned on the magnitudes of their generalized neighborhood coefficients, contourlet coefficients can be approximately modeled as Gaussian random variables. Based on these findings, we model contourlet coefficients using a hidden Markov tree (HMT) model with Gaussian mixtures that can capture all inter-scale, inter-direction, and inter-location dependencies. We present experimental results using this model in image denoising and texture retrieval applications. In denoising, the contourlet HMT outperforms other wavelet methods in terms of visual quality, especially around edges. In texture retrieval, it shows improvements in performance for various oriented textures.**

***Index Terms*— wavelets, contourlets, multiscale, multidirection, image modeling, statistical models, multiscale geometric analysis.**

I. Introduction

Image processing typically relies on simple statistical models to characterize images. Natural images tend to have certain common characteristics that make them look "natural." The aim of statistical modeling is to capture these defining characteristics in a small number of parameters so that they can be used as prior information in image processing tasks such as compression, denoising, feature extraction, and inverse problems. A simple, accurate and tractable model is an essential element in any successful image processing algorithm.

Images have effectively been modeled using the wavelet transform [1], [2], which offers a multiscale and time-frequency-localized image representation. Initially, the wavelet transform was considered to be a good decorrelator for images, and thus wavelet coefficients were assumed to be independent and were simply modeled using marginal statistics [3]. Later, it was realized that wavelet coefficients of natural images exhibit strong dependencies both across scales and between neighboring coefficients within a subband, especially around image edges. This gave rise to several successful joint statistical models in the wavelet domain [4], [5], [6], [7], [8], [9], as well as improved image compression schemes [10], [11], [12].

The major drawback for wavelets in two-dimensions is their limited ability in capturing directional information. To overcome this deficiency, researchers have recently considered multiscale and directional representations that can capture the intrinsic geometrical structures such as smooth contours in natural images. Some examples include the steerable pyramid [13], brushlets [14], complex wavelets [15], and the curvelet transform [16]. In particular, the curvelet transform, pioneered by Candès and Donoho, was shown to be optimal in a certain sense for functions in the continuous domain with curved singularities.

Inspired by curvelets, Do and Vetterli [17], [18] developed the *contourlet* transform based on an efficient two-dimensional multiscale and directional filter bank that can deal effectively with images having smooth contours. Contourlets not only possess the main features of wavelets (namely, multiscale and time-frequency localization), but also offer a high degree of directionality and anisotropy. The main difference between contourlets and other multiscale directional systems is that the contourlet transform allows for different and flexible number of directions at each scale, while achieving nearly critical sampling. In addition, the contourlet transform uses iterated filter banks, which makes it computationally efficient; specifically, it requires $O(N)$ operations for an N-pixel image.

In this work, we focus on image modeling in the contourlet domain. Our primary goal is to provide an extensive study on the statistics of contourlet coefficients in order to gain a thorough understanding of their properties. We then develop an appropriate model that can capture these properties, which can be useful in future contourlet applications including compression, denoising, and feature extraction. Similar to wavelet-based models, contourlet-based models need to take into account dependencies across *scales* and *space* (or *locations*). However, as a "true" two-dimensional representation, contourlets additionally allow us to model the dependency across *directions*. In other words, contourlet modeling allows us to jointly model all three fundamental parameters of visual information, namely *scale*, *space*, and *direction*.

The rest of the paper is organized as follows. Section II introduces the basics of contourlets including their transform algorithm, structure, properties, and coefficient relationships. In Section III, we study the marginal and joint statistics of

Duncan D.-Y. Po was with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign. He is now with The MathWorks, Natick, MA 01760.

M. N. Do is with the Department of Electrical and Computer Engineering, the Coordinated Science Laboratory, and the Beckman Institute, University of Illinois at Urbana-Champaign, Urbana IL 61801 (email: minhdo@uiuc.edu).

This work was supported by the National Science Foundation under Grant CCF-0237633 (CAREER).

contourlet coefficients of natural images using histograms. Section IV examines the dependencies between coefficients using mutual information. Inspired by these results, we develop a hidden Markov tree (HMT) model for the contourlet transform in Section V. In Section VI, we apply the contourlet HMT model in image denoising and texture retrieval. Finally, a conclusion is presented in Section VII.

II. Background

A. Contourlets

The primary goal of the contourlet construction [17], [18] was to obtain a sparse expansion for typical images that are piecewise smooth away from *smooth contours*. Two-dimensional wavelets, with tensor-product basis functions as shown in Figure 1(a), lack directionality and are only good at catching *point* discontinuities, but do not capture the *geometrical smoothness* of the contours.

Contourlets were developed as an improvement over wavelets in terms of this inefficiency. The resulting transform has the multiscale and time-frequency-localization properties of wavelets, but also offers a high degree of directionality and anisotropy. Specifically, contourlet transform involves basis functions that are oriented at any power of two's number of directions with flexible aspect ratios, with some examples shown in Figure 1(b). With such a rich set of basis functions, contourlets can represent a smooth contour with fewer coefficients compared with wavelets, as illustrated in Figure 1(c).

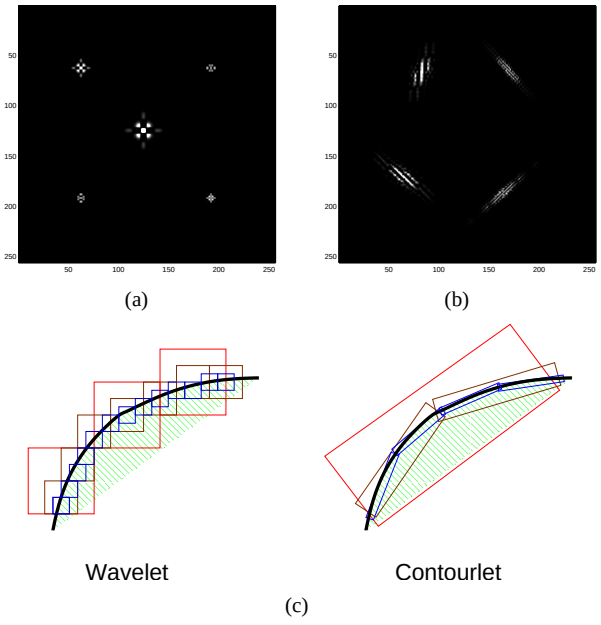


Fig. 1. Contourlet and wavelet representations for images. (a) Examples of five 2-D wavelet basis images. (b) Examples of four contourlet basis images. (c) Illustration showing how wavelets having square supports that can only capture point discontinuities, whereas contourlets having elongated supports that can capture linear segments of contours, and thus can effectively represent a smooth contour with fewer coefficients.

The contourlet transform is implemented via a two-dimensional filter bank that decomposes an image into several directional subbands at multiple scales. This is accomplished by combining the Laplacian pyramid [19] with a directional filter bank [20] at each scale. Due to this cascade structure, multiscale and directional decomposition stages in the contourlet transform are independent of each other. One can decompose each scale into any arbitrary power of two's number of directions, and different scales can be decomposed into different numbers of directions. This feature makes contourlets a unique transform that can achieve a high level of flexibility in decomposition while being close to critically sampled (up to 33% overcomplete, which comes from the Laplacian pyramid). Other multiscale directional transforms have either a fixed number of directions, or are significantly overcomplete (depending on the number of directions). Figure 2 shows an example frequency partition of the contourlet transform where the four scales are divided into four, four, eight, and eight directional subbands from coarse to fine scales, respectively.

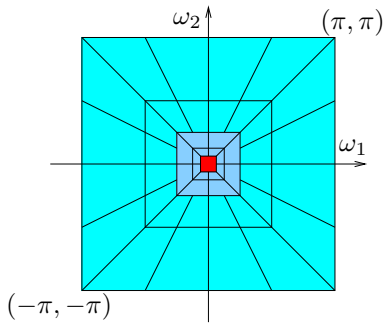


Fig. 2. An example frequency partition by the contourlet transform.

Figure 3 shows an example of the contourlet transform on the "Peppers" image. For the visual clarity, only two-scale decompositions are shown. The image is decomposed into a lowpass subband and several bandpass directional subbands. We notice that only contourlets that match with *both* location and direction of image contours produce significant coefficients. Thus, the contourlet transform effectively exploits the fact image edges are localized in both location and direction.

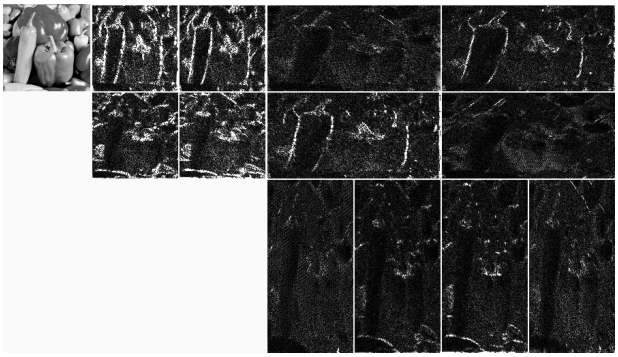


Fig. 3. Contourlet transform of the "Peppers" image. The image is decomposed into two pyramidal levels, which are then decomposed into four and eight directional subbands. Small coefficients are colored black while large coefficients are colored white.

B. Coefficient Relationships

We define some important contourlet coefficient relationships as depicted in Figure 4. For each contourlet coefficient

X, we define its eight adjacent coefficients in the same subband as its *neighbors* (NX). Next, the coefficient in the same spatial location in the immediately coarser scale is defined as its *parent* (PX), and those in the same spatial location in the immediately finer scale are its children. Note that each child has one parent and each parent has four children. We also define coefficients at the same scale and spatial location but in different directions as cousins (CX). This inter-direction relationship is more important for contourlets than for wavelets as contourlets have more directions. From Figure 4, it can be observed that there can be multiple cousins in certain directions. This is because the basis functions corresponding to the vertical and horizontal subbands are defined over different sampling lattices [21].

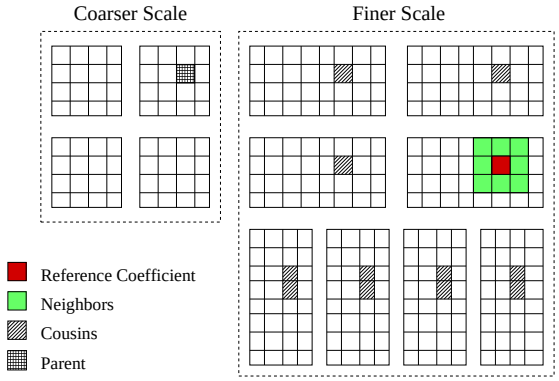


Fig. 4. Contourlet coefficient relationships.

Combining the relationships across scales, space, and directions, we refer to the collective set of all parent (PX), neighbors (NX), and cousins (CX) of each coefficient X as its *generalized neighborhood*. These relationships play an important role in contourlet modeling, as will be seen in subsequent sections.

C. Transform Setups

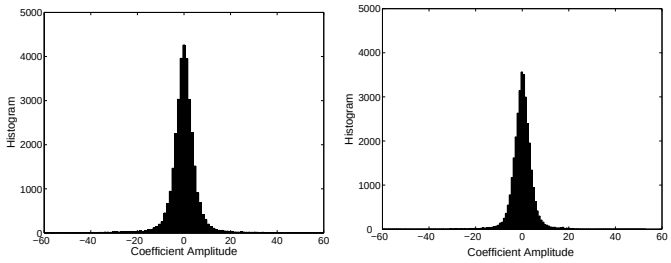
For the statistical studies in the next two sections, we experiment with various natural images of size 512×512. These images vary from simple edge-dominant images such as "Peppers" to highly textured images such as "Barbara." Unless stated otherwise, for the contourlet transform, we use the 9-7 biorthogonal filters (referred to as 9-7 filters) for the multiscale decomposition stage and the McClellan transformed directional filters from the 9-7 filters proposed by Cohen and Daubechies [22] (referred to as CD filters) for the multidirectional decomposition stage. We partition the finest and second finest scales into eight directional subbands, and the two next coarser scales into four directional subbands, and obtain a frequency partition as shown in Figure 2.

We would like to point out that in contrast to wavelets, where there exist many good wavelet filters, the filter design problem for contourlets is still an ongoing work. Thus, the results of the contourlet transform in this paper should be viewed as an indication of its features and potential rather as a comprehensive assessment.

III. Contourlet Statistics

A. Marginal Statistics

We first study the marginal statistics of the contourlet coefficients of natural images. Figure 5 plots the histograms of two finest subbands of the image "Peppers." These distributions exhibit a sharp peak at zero amplitude and heavy tails to both sides of the peak. This implies that the contourlet transform is sparse, as the majority of coefficients are close to zero. The kurtosis of the two shown distributions are 24.50 and 19.40, which are much higher than the kurtosis of 3 for Gaussian distributions. Similar distributions are also observed at all subbands of other test images. Thus, the subband marginal distributions of natural images in the contourlet domain are highly non-Gaussian.



(a) (b)

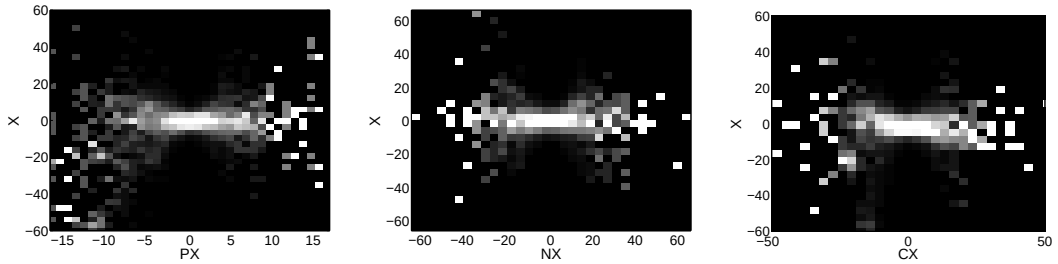
Fig. 5. Marginal statistics of two finest subbands of the image "Peppers." The kurtosis of the two distributions are measured at (a) 24.50 and (b) 19.40, showing that the coefficients are highly non-Gaussian.

B. Joint Statistics

Marginal statistics only describe individual behaviors of transform coefficients without accounting for their dependencies. It is clear that contourlet coefficients depend on each other since only contourlet functions that overlap and directionally align with image edges lead to significant coefficients. Figure 6 shows the conditional distributions of contourlet coefficients, conditioned on their parents (PX), neighbors (NX), and cousins (CX), using the "Peppers" image. First, we notice that all of these conditional distributions exhibit a "bow-tie" shape where the variance of the coefficients is related to the magnitude of the conditioned coefficient. Second, even though coefficients are correlated due to the slight overcompleteness of the contourlet transform, they are *approximately* decorrelated since conditional expectations $E[X \mid \cdot] \approx 0$. Again, similar behaviors are observed at all subbands of other test images. Therefore, we conclude that contourlet coefficients of natural images are approximately *uncorrelated* yet *dependent* on each other.

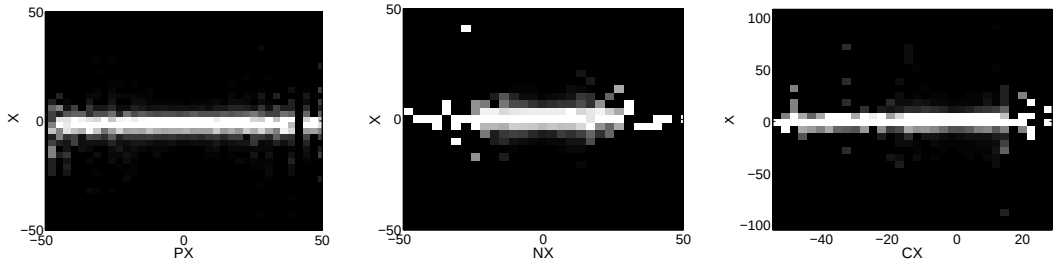
These dependencies, however, are *local*. Figure 7 shows the conditional distributions of contourlet coefficients conditioned on *distant* relatives and neighbors of the "Peppers" image. We observe that these conditional distributions are approximately invariant to the conditioned value, indicating independence.

Finally, we examine the conditional distributions $P(X \mid PX = px)$, $P(X \mid NX = nx)$ and $P(X \mid CX = cx)$, where px, nx, and cx are some fixed values, as shown in Figure 8. The kurtoses of the shown conditional distributions are



(a) (b) (c)

Fig. 6. Conditional distribution of a finest subband of "Peppers," conditioned on (a) parent $P(X \mid PX)$, (b) neighbor $P(X \mid NX)$, (c) cousin $P(X \mid CX)$.



(a) (b) (c)

Fig. 7. Distribution of a finest subband of "Peppers" conditioned on (a) ancestors, (b) neighbors, and (c) cousins, *all at distances of three coefficients away*.

3.90, 2.90, and 2.99, conditioned on the coefficients' parents, neighbors, and cousins, respectively. In contrast to the high kurtoses (around 20) of the marginal distributions in Figure 5, these conditional kurtoses are very close to the kurtosis of 3 for Gaussian distributions. Thus, we conclude that contourlet coefficients are *non-Gaussian but conditionally Gaussian*. In other words, the contourlet coefficients of natural images can be accurately modeled by mixtures of Gaussian distributions whose variances depend on their generalized neighborhood coefficients. We will explore this fact for developing contourlet-based models in Section V.

IV. Dependence Characterization via Information-Theoretic Analysis

In this section, we quantitatively study the joint statistics of contourlet coefficients in complement to the qualitative study in the previous section, using mutual information as a measure of dependencies [23]. Our goal is to compare the dependencies across scales, space, and directions in the contourlet domain.

A. Mutual Information and Estimation

Mutual information between two random variables X and Y with joint density $p(x, y)$ and marginal densities $p(x)$ and $p(y)$ is defined as [24]

$$I(X;Y) = \int\int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy. \qquad (1)$$

Mutual information can be interpreted as how much information one random variable contains about the other. Its value increases with increasing dependence between the two variables.

We resort to nonparametric histogram estimation of mutual information [25], [26] between contourlet coefficients and their generalized neighborhoods. Specifically, for a pair of contourlet coefficients X and Y (such as coefficients in a subband and their parents), we use the following estimator [25]

$$\hat{I}(X;Y) = \sum_{i,j} \frac{k_{ij}}{N} \log \frac{k_{ij} N}{k_i k_j} - \frac{(J-1)(K-1)}{2N}, \qquad (2)$$

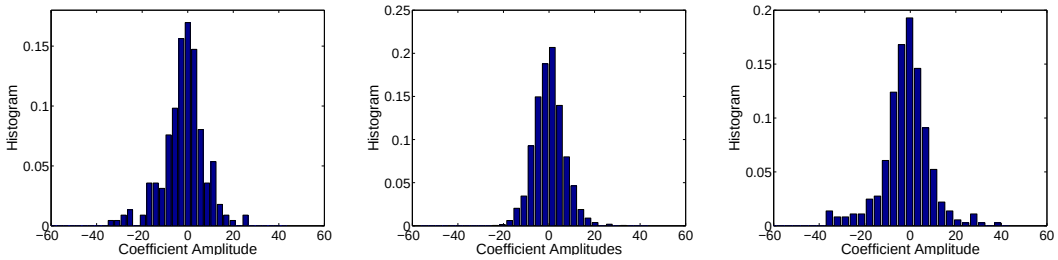
where k_{ij} is the number of coefficient pairs in the joint histogram cell (i, j), $k_i = \sum_j k_{ij}$ and $k_j = \sum_i k_{ij}$ are the marginal distribution histogram estimates, N is the total number of considered coefficient pairs, and J and K are the number of histogram bins along X and Y directions respectively. The first term in (2) is the mutual information histogram estimate, while the second term is a partial bias correction term. It can be shown that even after the bias is partially removed, the residual bias still causes the estimator to underestimate the mutual information and the estimate can only serve as a lower bound [25].

One way to tighten the estimation bound is by choosing J and K to give the maximum estimate in (2). Empirically, we found that using

$$J = K = \text{round}(N/3000) + 12, \qquad (3)$$

where round$(\cdot)$ denotes rounding to the nearest integer, yields good mutual information estimates for general natural images.

For mutual information involving a large set of variables, the estimator in (2) becomes inaccurate. For example, consider estimating the mutual information $I(X; Y_1, Y_2, \ldots, Y_M)$ between contourlet coefficients at a particular subband (denoted by X) and a subset of their generalized neighborhoods (denoted by $\{Y_i\}_{i=1}^M$). As the subset size M increases, mutual information estimation accuracy decreases exponentially [26].



(a) (b) (c)

Fig. 8. Conditional distribution of the coefficients of a finest subband of "Peppers" on (a) their parents $P(X \mid PX = px)$, (b) neighbors $P(X \mid NX = nx)$, and (c) cousins $P(X \mid CX = cx)$. The kurtosis of the distributions are measured at 3.90, 2.90, and 2.99, respectively.

In such cases, we follow the approach in [23] that replaces the high dimensional variable $\{Y_i\}_{i=1}^{M}$ by its sufficient statistic

$$T = \sum_{i=1}^{M} a_i |Y_i|, \tag{4}$$

where a_i are constant weights. Then $I(X;T)$ involves only two variables and can be accurately estimated as in (2). From the data processing inequality [24], we have

$$I(X;T) \leq I(X;Y_1,Y_2,\ldots,Y_M). \tag{5}$$

We obtain the tightest lower bound of $I(X;Y_1,Y_2,\ldots,Y_M)$ in (5) by choosing a_i to maximize $I(X;T)$, using standard optimization algorithms in MATLAB.

B. Estimation Results

We present the mutual information estimation results for three representative images "Lena," "Barbara," and "Peppers" in Table I, noting that other images give similar results. All images show significant mutual information across all of scales, space, and directions, which reinforces our observations in Section III that coefficients are dependent on their generalized neighborhoods. At fine scales, we empirically find that $I(X;NX)$ is higher than $I(X;CX)$, which is higher than $I(X;PX)$. Therefore, the eight neighbor coefficients contain the most information about the coefficients. This is especially true for highly textured images like "Barbara," since for such images large coefficients correspond to highly textured areas which span relatively large image areas with similar directional patterns.

TABLE I

MUTUAL INFORMATION ESTIMATES BETWEEN A CONTOURLET COEFFICIENT X AND ITS PARENT PX, ITS SPATIAL NEIGHBORS NX, AND ITS DIRECTIONAL COUSINS CX.

	Lena	Barbara	Peppers
$I(X;PX)$	0.11	0.14	0.10
$I(X;NX)$	0.23	0.58	0.17
$I(X;CX)$	0.19	0.39	0.14
$I(X;PX,NX)$	0.24	0.58	0.17
$I(X;NX,CX)$	0.26	0.59	0.20
$I(X;PX,CX)$	0.21	0.40	0.16
$I(X;PX,NX,CX)$	0.26	0.59	0.20

The mutual information estimates in Table I depend on the choice of filters and number of directions. Table II shows the mutual information estimates for the "Lena" image using different combinations of pyramidal (P) filters: the Haar and 9-7 filters, and directional (D) filters: the CD filters [22] and the ladder filters by Phoong et al. (referred to as PKVA filters) [27]. We find that replacing the Haar filters by the 9-7 filters for the multiscale decomposition stage significantly reduces all inter-scale, inter-location, and inter-direction mutual information of the contourlet coefficients. This suggests that the 9-7 filters are superior to the Haar filters in terms of whitening the contourlet coefficients. Similarly, replacing the CD filters by the PKVA filters for the directional decomposition stage reduces the inter-direction mutual information. This indicates the PKVA filters are more effective in localizing edge direction and should lead to better performance in applications. Again, we note that designing more effective directional filters for the contourlet transform is still an ongoing work, where better contourlet filters are expected to be found.

As a comparison, Table II also displays the mutual information estimates of wavelet coefficients for the image "Lena," using different wavelet filters. Compared with wavelet coefficients, contourlet coefficients exhibit similar inter-scale and inter-location dependencies but much higher inter-direction dependencies.

TABLE II

MUTUAL INFORMATION ESTIMATES FOR CONTOURLET AND WAVELET REPRESENTATIONS OF THE "LENA" IMAGE USING DIFFERENT FILTERS.

Contourlets			
P-filter; D-filter	$I(X;PX)$	$I(X;NX)$	$I(X;CX)$
Haar; CD	0.18	0.33	0.32
9-7; CD	0.11	0.23	0.19
9-7; PKVA	0.11	0.24	0.15
Wavelets			
Filter	$I(X;PX)$	$I(X;NX)$	$I(X;CX)$
Haar	0.20	0.27	0.14
Daubechies 4-taps	0.14	0.23	0.08
Daubechies 8-taps	0.11	0.20	0.05

Table III compares the inter-direction and inter-location mutual information for different directional partitioning schemes to partition the finest scale into 4, 8, and 16 directions. Each further directional partition increases the inter-direction dependency while decreasing the inter-location dependency.

C. Single Coefficient Estimates

Despite the observation that neighboring coefficients carry the strongest dependency, incorporating neighbors into statisti-

TABLE III
MUTUAL INFORMATION ESTIMATES FOR DIFFERENT DIRECTIONAL PARTITIONS OF THE FINEST SCALE. DATA ARE OBTAINED USING 9-7 PYRAMIDAL FILTERS AND CD DIRECTIONAL FILTERS ON THE "LENA" IMAGE.

	4 directions	8 directions	16 directions
$I(X;NX)$	0.26	0.23	0.20
$I(X;CX)$	0.14	0.19	0.19

cal models is not simple. Indeed, as every coefficient has eight neighbors, using a contextual model would result in a complex dependence network which is undesirable. To simplify the model, we want every coefficient to be modeled to depend on only a *single coefficient*. Thus in the next step, we measure and compare the mutual information between the coefficients and each of their neighbors and cousins individually rather than as a collective set.

Table IV presents estimation results of mutual information with *single* neighbor and cousin. To keep the results manageable, we reduce the data size by grouping the measurements and summarizing each group into an average. NX_i refers to i-th order neighbors where $i = 1$ denotes the four adjacent neighbors and $i = 2$ denotes the four diagonal neighbors. CX_j refers to the j-th order cousins, where $j = 1$ denotes the two immediate adjacent directions, $j = 2$ denotes the next two directions, and so on. $I(X;NX_i)$ and $I(X;CX_j)$ are the average of the mutual information contributed by each of the four i-th order neighbors and each of the two j-th order cousins, respectively. For edge-dominant images like "Lena" and "Peppers," we empirically find that $I(X;PX)$ is higher than $I(X;NX_i)$, which is higher than $I(X;CX_i)$, for all i. Thus, *the parent coefficient becomes the most significant predictor when generalized neighborhood coefficients are considered individually*. The exceptions are highly textured images like "Barbara," where single a neighbor coefficient remains the dominant predictor.

TABLE IV
AVERAGE MUTUAL INFORMATION ESTIMATES WITH A **SINGLE** PARENT, NEIGHBOR, AND COUSIN.

	Lena	Barbara	Peppers
$I(X;PX)$	0.11	0.14	0.08
$I(X;NX_1)$	0.09	0.31	0.07
$I(X;NX_2)$	0.07	0.27	0.05
$I(X;CX_1)$	0.08	0.20	0.06
$I(X;CX_2)$	0.06	0.17	0.05
$I(X;CX_3)$	0.06	0.20	0.04

D. Summary

Based on the measured statistics in Sections III and IV, contourlet coefficients of natural images possess the following properties:

- P1 Contourlet coefficients are marginally non-Gaussian and dependent on their generalized neighborhood coefficients.
- P2 Conditioned on the magnitudes of their generalized neighborhood coefficients, contourlet coefficient are zero-mean Gaussian distributed.
- P3 The parent coefficient is typically the most significant predictor when generalized neighborhood coefficients are considered individually.

V. IMAGE MODELING

We want to develop a statistical model that can incorporate the properties of contourlet coefficients summarized in Section IV-D while possessing other desirable characteristics. Specifically, an ideal statistical model should: (i) accurately model all properties P1–P3 of contourlet coefficients, (ii) have a simple structure to enable efficient training algorithms, and (iii) be defined on a small number of parameters to allow accurate training with limited data.

Properties P1 and P2 suggest that each contourlet coefficient can be accurately modeled by a mixture of Gaussian distributions where the condition for being in each Gaussian distribution depends on the generalized neighborhood coefficients. A good candidate for this is the hidden Markov model (HMM) which has been used effectively for the wavelet transform [7]. An N-state HMM associates each coefficient with a *hidden state* variable, randomly distributed over its N states. Conditioned on its state, each coefficient is modeled using a Gaussian distribution with parameters depending on the state. Therefore, each coefficient is characterized by an N-dimensional state probabilities vector $\boldsymbol{p}$ and an N-dimensional standard deviation vector $\boldsymbol{\sigma}$ (we assume contourlet coefficients have zero-mean since all contourlet basis functions have zero-sum)

$$\boldsymbol{p} = (p_1, p_2, \ldots, p_N)^T, \quad \text{and} \quad \boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)^T, \quad (6)$$

where $1, 2, \ldots, N$ denote the states.

The inter-coefficient dependencies in HMMs are established by links between the *hidden states* of dependent coefficients. Associated with each link between coefficients m and n is an $N \times N$ state transition probability matrix $\mathbf{A}_{m,n}$ where its (k,l) entry is the the probability that the coefficient m is in state k given the coefficient n is in state l. To reduce the number of parameters, like in the wavelet-domain HMM [7], we "tie" all contourlet coefficients in the same subband to share the same set of model parameters.

The hidden Markov tree (HMT) model [7] is an HMM that uses a quad-tree dependency structure. More precisely, the HMT model establishes links between the hidden state of each coefficient and those of its four children. The quad-tree structure of the HMT models directly model inter-scale dependencies, while indirectly modeling other dependencies between neighboring coefficients via their links to a common ancestor. The motivation for adapting the HMT model is because of the property P3 which states the parent coefficient is the most significant predictor when the generalized neighborhood coefficients are compared *individually*. Furthermore, the HMT model has a simple tree structure, which enables efficient training using the expectation maximization (EM) algorithm. In particular, with only two states ($N = 2$), the HMT model requires a relatively small number of parameters. Two-states mixture models also possess an intuitive appeal that their two states correspond to the two modes in natural

images: edges and smooth areas. We empirically verified that the marginal distributions of the contourlet coefficients in each subband (e.g. see Figure 5) can be accurately modeled by a mixture of two zero-mean Gaussian distributions.

Specifically, for a contourlet decomposition of J scales and m_j directional subbands within scale j ($j = 1, 2, \ldots J$, from coarse to fine), a contourlet HMT model contains the following parameters.

- $\boldsymbol{p}_{1,k}$ (where $k = 1, \ldots, m_1$): for the root state probability vector at each directional subband at the coarsest scale.
- $\mathbf{A}_{j,k}$ (where $j = 2, \ldots, J$, and $k = 1, \ldots, m_j$): for the state transition probability matrix to the directional subband k at scale j from its parent subband at scale $j-1$.
- $\boldsymbol{\sigma}_{j,k}$ (where $j = 1, \ldots, J$, and $k = 1, \ldots, m_j$): for the Gaussian standard deviation vector of the subband at scale j and direction k.

Compared with the wavelet HMT model, the contourlet HMT model has a major advantage is that it accounts for inter-direction dependencies while the wavelet HMT model does not. Figure 9 illustrates this difference. In the wavelet HMT model, the parent-children links are always in the same direction among three wavelet directions. As a result, the wavelet HMT models coefficients on each direction independently. In contrast, contourlet coefficients can have their four children in two separate directional subbands. As a result, the dependence tree in the contourlet HMT can span several adjacent directions in the finer scales, and thus inter-direction dependencies are modeled in a similar way as inter-location dependencies (see Figure 9(c)). In other words, the contourlet HMT model effectively captures all dependencies across scales, space, and directions.

Counting the number of free model parameters, where the number of states $N = 2$, we count one free parameter in each state probability vector $\boldsymbol{p}_{1,k}$ and two in each state transaction probability matrix $\mathbf{A}_{j,k}$ (the other parameters being constrained to sum with these parameters to unity), and two in each $\boldsymbol{\sigma}_{j,k}$. Therefore, the contourlet HMT model requires three parameters for each coarsest (root) subband and four parameters for each finer subband. For the subband partitioning scheme in Figure 2(b) with 4, 4, 8, and 8 directions from coarse to fine scales, the model has 92 free parameters in total.

It should be noted that there exist many other forms of Gaussian mixture and tree models outside the HMM family [9], [28], as well as variations of within the HMM family [8], [29], [30]. For example, the models in [29] are based on enlarged parameter sets that require more computation for training. With the goals of keeping the parameter set small and ensuring efficient algorithms, we choose the simple HMT model with two states.

Table V shows the HMT parameters of the "Peppers" image obtained via the EM training algorithm [7]. According to the standard deviations $\boldsymbol{\sigma_{j,k}}$, states 1 and 2 show two clearly different modes of coefficients being large and small, which correspond to edge and smooth parts of the image, respectively. The state transition matrices $\mathbf{A}_{j,k}$ capture the dependencies among coefficients. In particular, they show a high probability of a coefficient being small (i.e., in state 2) given its parent is also small. However, state 2 (i.e., coefficient being large) exhibits less persistent across scales. The reason for these is because small contourlet coefficients correspond to smooth image regions; if the parent is small, all of its children are likely to be small. Whereas, a large contourlet coefficient is obtained only when the corresponding contourlet basis function aligns with both location and direction of an image edge. Thus, on a quad-tree of contourlet coefficients, from coarse to fine scales the set of large coefficients becomes more localized to certain locations and directions.

VI. Applications

A. Denoising

We apply the contourlet HMT model developed in the previous section to denoising of zero-mean additive white Gaussian noise. In the contourlet domain, we can write

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{e}, \tag{7}$$

where $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{e}$ are contourlet coefficients of the clean, noisy, and additive noise images, respectively. The problem is to estimate $\boldsymbol{u}$ given $\boldsymbol{v}$.

We first fit an HMT model to the noisy contourlet coefficients $\boldsymbol{v}$ to obtain a parameter set $\boldsymbol{\theta_v}$. From this, we obtain a model for the clean contourlet coefficients $\boldsymbol{\theta_u}$ by subtracting the noise variances from the variances in the model $\boldsymbol{\theta_v}$:

$$(\sigma^{(\boldsymbol{u})}_{(j,k,n),m})^2 = \left((\sigma^{(\boldsymbol{v})}_{(j,k,n),m})^2 - (\sigma^{(\boldsymbol{e})}_{(j,k,n)})^2\right)_+, \tag{8}$$

where the triple (j, k, n) denotes the n-th contourlet coefficient in the subband indexed by scale j and direction k; the index m denotes the state; and $(x)_+ = x$ for $x \geq 0$ and $(x)_+ = 0$ for $x < 0$. Note that because of the "tying" of contourlet coefficients in each subband, $\boldsymbol{\sigma}^{(\boldsymbol{v})}_{j,k,n} = \boldsymbol{\sigma}^{(\boldsymbol{v})}_{j,k}$ for all n. The noise variance $(\sigma^{(\boldsymbol{e})}_{j,k,n})^2$ can be estimated using a Monte-Carlo method: by repeatedly generating random white noise images and averaging their variance in the contourlet domain.

With the HMT model $\boldsymbol{\theta_u}$, the denoising problem can then be formulated into a Bayesian estimation problem where we want to estimate the quantity $E[u_{j,k,n} \mid v_{j,k,n}, \boldsymbol{\theta_u}]$ for each contourlet coefficient (j, k, n). Note that given its states $S_{j,k,n}$, the contourlet coefficient is assumed to be Gaussian and the problem is reduced to estimating zero-mean Gaussian signal in zero-mean Gaussian noise. The solution to this problem is well known [7]

$$E[u_{j,k,n} \mid v_{j,k,n}, \boldsymbol{\theta_u}, S_{j,k,n} = m] = \frac{(\sigma^{(\boldsymbol{u})}_{(j,k,n),m})^2}{(\sigma^{(\boldsymbol{u})}_{(j,k,n),m})^2 + (\sigma^{(\boldsymbol{e})}_{j,k,n})^2} v_{j,k,n}. \tag{9}$$

From this, using the state probability $p(S_{j,k,n} = m \mid v_{j,k,n}, \boldsymbol{\theta_u})$, which can be obtained from the EM algorithm as a by-product during training, we obtain the final estimate for the clean coefficients

$$E[u_{j,k,n} \mid v_{j,k,n}, \boldsymbol{\theta_u}] = \sum_m p(S_{j,k,n} = m \mid v_{j,k,n}, \boldsymbol{\theta_u}) \times \frac{(\sigma^{(\boldsymbol{u})}_{(j,k,n),m})^2}{(\sigma^{(\boldsymbol{u})}_{(j,k,n),m})^2 + (\sigma^{(\boldsymbol{e})}_{j,k,n})^2} v_{j,k,n}. \tag{10}$$

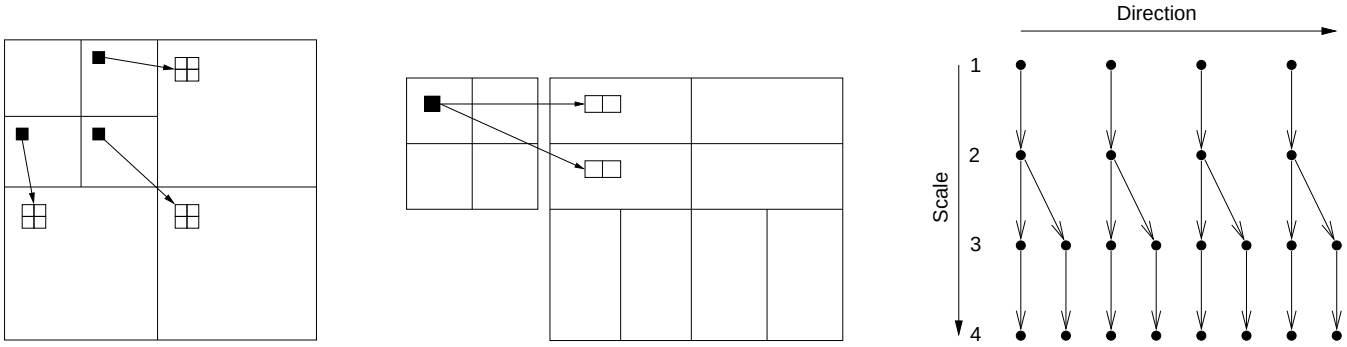


Fig. 9. Parent-children relationship for (a) wavelets and (b) a possible contourlet decomposition. Black squares represent parent coefficients with white squares as their children. Notice in the contourlet case, a parent coefficient can have its children spread over two subbands. (c) Dependency links between subbands (represented by black circles) of the contourlet decomposition with 4, 4, 8, and 8 directions from coarse to fine scales.

		$k=1$		$k=2$		$k=3$		$k=4$	
		state 1	state 2	state 1	state 2	state 1	state 2	state 1	state 2
$\boldsymbol{p}_{1,k}$		0.67	0.33	0.32	0.68	0.27	0.73	0.75	0.25
$\mathbf{A}_{2,k}$	state 1	0.87	0.02	0.83	0.04	0.74	0.05	0.72	0.03
	state 2	0.13	0.98	0.17	0.96	0.26	0.95	0.28	0.97
$\mathbf{A}_{3,2k-1}$	state 1	0.14	0.04	0.32	0.02	0.41	0.02	0.69	0.02
	state 2	0.86	0.96	0.68	0.98	0.59	0.98	0.31	0.98
$\mathbf{A}_{3,2k}$	state 1	0.22	0.08	0.29	0.02	0.52	0.01	0.65	0.02
	state 2	0.78	0.92	0.71	0.98	0.48	0.99	0.35	0.98
$\mathbf{A}_{4,2k-1}$	state 1	0.82	0.00	0.35	0.00	0.55	0.01	0.38	0.00
	state 2	0.18	1.00	0.65	1.00	0.45	0.99	0.62	1.00
$\mathbf{A}_{4,2k}$	state 1	0.36	0.00	0.44	0.00	0.40	0.01	0.41	0.00
	state 2	0.64	1.00	0.56	1.00	0.60	0.99	0.59	1.00
$\boldsymbol{\sigma_{1,k}}$		160.10	60.16	258.30	84.76	165.04	59.00	168.01	76.82
$\boldsymbol{\sigma_{2,k}}$		59.88	12.29	98.37	18.79	70.06	21.79	70.83	27.18
$\boldsymbol{\sigma_{3,2k-1}}$		7.71	7.68	76.11	11.62	20.62	7.90	39.47	6.90
$\boldsymbol{\sigma_{3,2k}}$		37.59	9.65	27.89	8.69	41.66	8.53	21.77	8.31
$\boldsymbol{\sigma_{4,2k-1}}$		14.69	3.20	33.65	5.26	16.05	3.57	17.76	4.38
$\boldsymbol{\sigma_{4,2k}}$		16.47	3.43	31.46	5.20	17.91	3.66	17.57	4.34

TABLE V
CONTOURLET HMT PARAMETERS OF THE IMAGE "PEPPERS."

We apply the above denoising algorithm to noisy versions of several test images and the results are shown in Table VI. Also in the table are the denoising results of the same noisy images using the Wiener filter ("wiener2"), the wavelet thresholding method with threshold $T = 3\sigma$, and the wavelet HMT denoising method [7]. In terms of peak signal-to-noise ratio (PSNR), the wavelet HMT and the contourlet HMT methods produce comparable results. Both HMT methods outperform "wiener2" and wavelet thresholding.

In terms of visual quality, the contourlet HMT produces superior denoised results. Figure 10 shows the denoising results of the "Zelda" image. It is clear that the contourlet HMT removes most of the noise and produces the smoothest image, even though it is not the best method in term of PSNR.

Recently, more complex wavelet-based denoising schemes (e.g., [31], [32], [28]), have been developed. While some have shown improvements over HMT algorithms, it is outside the scope of this paper to provide detailed comparisons with all denoising schemes. We only provide a comparison which demonstrates that using HMT models, contourlets achieve superior visual quality over wavelets.

TABLE VI
PEAK SIGNAL-TO-NOISE RATIO (IN DB) OF DENOISED IMAGES USING DIFFERENT METHODS.

image	noise level σ	noisy	wiener2 (5×5)	wavelet thresh.	wavelet HMT	contourlet HMT
Lena	30	18.88	27.40	26.67	**28.35**	28.18
	40	16.53	25.97	25.26	**27.21**	27.00
	50	14.63	24.75	24.20	25.89	**26.04**
Barbara	30	18.72	24.95	23.76	25.11	**25.27**
	40	16.38	23.59	22.66	**24.94**	24.79
	50	14.48	22.57	21.96	23.71	**23.74**
Zelda	30	18.83	28.67	28.24	**30.67**	30.00
	40	16.49	27.10	27.03	**29.27**	28.29
	50	14.61	25.78	26.05	**27.63**	27.07

B. Texture Retrieval

The contourlet HMT model is also applied to content-based texture retrieval. The problem is to find, given one query texture image, the texture images that are "most similar" to it within a large database of unlabeled texture images.

In the contourlet HMT texture retrieval system, each texture image is first transformed into contourlet coefficients. An

Fig. 10. Denoising results of "Zelda" image: (a) "Zelda" image, (b) noisy image (noise standard deviation = 50, PSNR = 14.61dB), (c) wiener2 (PSNR = 25.78dB), (d) wavelet thresholding (PSNR = 26.05dB), (e) wavelet HMT (PSNR = 27.63dB), and (f) contourlet HMT (PSNR = 27.07dB).

HMT model is then fit to the contourlet coefficients, and the HMT model parameters are extracted as image features. The Kullback-Liebler distances (KLD) between the query image model and each database image model are measured and the database images that give the smallest KLD are retrieved. The KLD is defined as [24]

$$D(p(X \mid \boldsymbol{\theta_q}) \parallel p(X \mid \boldsymbol{\theta_i})) = \int p(x \mid \boldsymbol{\theta_q}) \log \frac{p(x \mid \boldsymbol{\theta_q})}{p(x \mid \boldsymbol{\theta_i})} dx. \tag{11}$$

Retrieval based on KLD can be justified as retrieval by the maximum likelihood (ML) selection rule [33].

There is no closed form expression for the KLD between HMT models. To estimate the KLD between two HMT models, one can resort to a Monte-Carlo method that randomly generates multiple sets of data using the query model and then computes their likelihoods against each candidate model. An alternative method is to compute an upper bound of the KLD based on the log-sum inequality [34].

In the retrieval experiments, a subset of the Brodatz image database [35] was used. This subset contains 64 different textures of size 512×512 and is shown in Figure 11. Each of these textures was partitioned into 16 sub-images of size 128×128, and only 8 of the 16 were retained. Thus the database contained 512 texture images in total with 8 images from each of the 64 texture classes. We used each image in the database as the query image and retrieved the top 7 matches for each query. Table VII shows the average retrieval rates (i.e. percentage of correct images among the top 7 matches), where the contourlet HMT outperforms the wavelet HMT.

TABLE VII
AVERAGE RETRIEVAL RATES.

wavelet HMT	contourlet HMT
90.87%	**93.29**%

Comparing the retrieval rates for individual textures, contourlet and wavelet HMTs each gives better retrieval rates for certain textures. Figure 12(a) shows the texture images that are better retrieved by wavelets than by contourlets by at least 5%, and Figure 12(b) shows those better retrieved by contourlets by at least 5%. From these figures, it can be seen that the textures better retrieved by wavelets are all characterized by dominant vertical, horizontal, or $\pm 45^\circ$ directions, which are the only directions captured well by wavelets. In contrast, the textures better retrieved by contourlets exhibit more diverse directional components (such as circular or irregular shapes). This shows the superiority of contourlets in capturing directional information. For most textures, contourlet HMT texture retrieval gives satisfactory texture retrieval performance with retrieval rates above 80%.

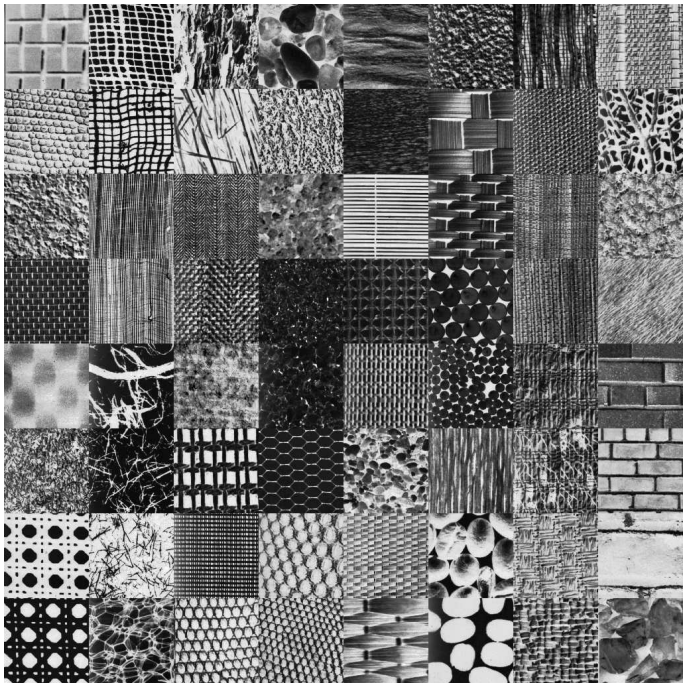
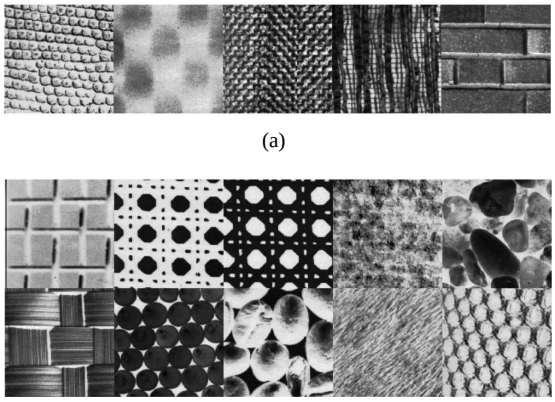


Fig. 11. Textures (64 images) from the Brodatz database used in our texture retrieval experiments.



(a)

(b)

Fig. 12. Texture retrieval results: (a) textures that are better retrieved by wavelets than by contourlets by at least 5%; (b) textures that are better retrieved by contourlets than by wavelets by at least 5%.

VII. Conclusion

We have studied the properties of the contourlet coefficients of natural images. It is found that similar to wavelets, contourlet coefficients are highly non-Gaussian and exhibit Markovian dependencies in the form of local clustering and persistence across scales. Moreover, coefficients across adjacent directions show more significant mutual dependencies compared to wavelets. Thus contourlet coefficients exhibit dependencies across all of scale, space, and direction. Conditioned on these generalized neighborhood coefficient magnitudes, contourlet coefficients are approximately Gaussian.

The dependencies across scales, space, and directions were quantitatively compared using mutual information measures. We have found that contourlet coefficients of natural images show high inter-location dependencies on all their eight neighbors, and also quite high inter-direction dependencies on their cousins. When comparing dependencies on each *single* coefficient, inter-scale dependencies become the highest for most images.

Based on these findings, we have developed a contourlet hidden Markov tree (HMT) model that captures the contourlet property of highly non-Gaussian but conditionally Gaussian distributions. Dependence on generalized neighborhood coefficient magnitudes is modeled through the links between the hidden states of the coefficients. The contourlet HMT model captures all inter-scale, inter-direction, and inter-location dependencies through an effective tree-structured dependence network.

We applied the contourlet HMT model to denoising and texture retrieval, and obtained promising results. In denoising, the contourlet HMT visually restores edges better than wavelet HMT and other classical methods. In texture retrieval, the contourlet HMT gives higher retrieval rates than wavelets for textures that show high directionality. Both results suggest that the contourlet transform and its proposed model capture directional information well and offer a valuable tool in image processing. Software for the implementation of the contourlet HMT model can be downloaded from MATLAB Central (`http://www.mathworks.com/matlabcentral/`).

References

[1] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Prentice-Hall, 1995.

[2] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, 1999.

[3] ——, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Patt. Recog. and Mach. Intell.*, vol. 11, no. 7, pp. 674–693, July 1989.

[4] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," in *Proc. SPIE 44-th Annual Meeting*, 1999, pp. 188–195.

[5] K. Mihçak, I. Kozintev, K. Ramchandran, and P. Moulin, "Low complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Proc. Letters*, pp. 300–303, Dec. 1999.

[6] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive image wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Proc.*, vol. 9, no. 9, pp. 1522–1531, Sep. 2000.

[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)*, vol. 46, pp. 886–902, Apr. 1998.

[8] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. Image Proc.*, vol. 10, no. 7, pp. 1056 –1068, Jul. 2001.

[9] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in modeling and analyzing natural images," *Journal of Appl. and Comput. Harmonic Analysis*, vol. 11, pp. 89–123, Jul. 2001.

[10] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing, Special Issue on Wavelets and Signal Processing*, vol. 41, no. 12, pp. 3445–3462, December 1993.

[11] S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation quantization framework," in *Proc. IEEE Data Compression Conf.*, 1997, pp. 221–230.

[12] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Proc.*, vol. 8, pp. 1688–1701, Dec. 1999.

[13] 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, no. 2, pp. 587–607, March 1992.

[14] F. G. Meyer and R. R. Coifman, "Brushlets: A tool for directional image analysis and image compression," *Journal of Appl. and Comput. Harmonic Analysis*, vol. 5, pp. 147–187, 1997.
[15] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Appl. and Comput. Harmonic Analysis*, vol. 10, pp. 234–253, 2001.
[16] E. J. Candès and D. L. Donoho, "Curvelets – a surprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting*, A. Cohen, C. Rabut, and L. L. Schumaker, Eds. Saint-Malo: Vanderbilt University Press, 2000.
[17] M. N. Do and M. Vetterli, "Contourlets," in *Beyond Wavelets*, G. V. Welland, Ed. New York: Academic Press, 2003.
[18] ——, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. Image Proc.*, to appear, `http://www.ifp.uiuc.edu/~minhdo/publications`.
[19] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, April 1983.
[20] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Trans. Signal Proc.*, vol. 40, no. 4, pp. 882–893, April 1992.
[21] M. N. Do, "Directional multiresolution image representations," Ph.D. dissertation, Swiss Federal Institute of Technology, Lausanne, Switzerland, December 2001, `http://www.ifp.uiuc.edu/~minhdo/publications`.
[22] A. Cohen and I. Daubechies, "Nonseparable bidimensional wavelet bases," *Rev. Mat. Iberoamericana*, vol. 9, no. 1, pp. 51–137, 1993.
[23] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Proc.*, vol. 10, no. 11, pp. 1647–1658, 2001.
[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: Wiley Interscience, 1991.
[25] R. Moddemeijer, "On estimation of entropy and mutual information of continuous distributions," *Signal Proc.*, vol. 16, no. 3, pp. 233–246, 1989.
[26] J. Beirlant, E. J. Dudewicz, L. Gyöfi, and E. van der Meulen, "Nonparametric entropy estimation: An overview," *International Journal of Mathematical and Statistical Sciences*, vol. 6, no. 1, pp. 17–39, 1997.
[27] S.-M. Phoong, C. W. Kim, P. P. Vaidyanathan, and R. Ansari, "A new class of two-channel biorthogonal filter banks and wavelet bases," *IEEE Trans. Signal Proc.*, vol. 43, no. 3, pp. 649–665, Mar. 1995.
[28] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussian in the wavelet domain," *IEEE Trans. Image Proc.*, vol. 12, no. 11, pp. 1338–1351, 2003.
[29] G. Fan and X. G. Xia, "Wavelet-based texture analysis and synthesis using hidden Markov models," *IEEE Trans. Circ. and Syst.*, vol. 50, no. 1, pp. 106–120, 2003.
[30] M. N. Do and M. Vetterli, "Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models," *IEEE Trans. Multimedia*, vol. 4, pp. 517–527, Dec. 2002.
[31] L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Proc. Letters*, vol. 9, no. 12, pp. 438–441, 2002.
[32] J. L. Starck, E. J. Candès, and D. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Proc.*, vol. 11, pp. 670–684, Jun. 2002.
[33] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Proc.*, vol. 11, pp. 146–158, Feb. 2002.
[34] M. N. Do, "Fast approximation of Kullback-Leibler distance for dependence trees and hidden Markov models," *IEEE Signal Proc. Letters*, vol. 10, pp. 115–118, Apr. 2003.
[35] P. Brodatz, *Textures: A Photographic Album for Artists & Designers*. Dover, 1966.

PLACE PHOTO HERE

Duncan D.-Y. Po was born in Hong Kong in 1978. He received the B.S. degree in Electrical and Computer Engineering from the University of Calgary, Canada, in 2001, and the M.S. degree in Electrical Engineering from the University of Illinois at Urbana-Champaign in 2003. He currently works for The Mathworks in Natick, Massachusetts.

PLACE PHOTO HERE

Minh N. Do was born in Thanh Hoa, Vietnam, in 1974. He received the B.Eng. degree in computer engineering 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, University of Illinois at Urbana-Champaign. His research interests include wavelets, image and multidimensional signal processing, multiscale geometric analysis, and visual information representation.

He received a Silver Medal from the 32nd International Mathematical Olympiad in 1991, a University Medal from the University of Canberra in 1997, the best doctoral thesis award from the Swiss Federal Institute of Technology Lausanne in 2001, and a CAREER award from the National Science Foundation in 2003.