

Directional Multiscale Modeling of Images using the Contourlet Transform

Duncan D.-Y. Po and Minh N. Do

Coordinated Science Lab and Beckman Institute

University of Illinois at Urbana-Champaign

Urbana IL 61801

Email: duncanpo@ifp.uiuc.edu, minhdo@uiuc.edu

Abstract

The contourlet transform is a new extension to the wavelet transform in two dimensions using nonseparable and directional filter banks. The contourlet expansion is composed of basis images oriented at varying directions in multiple scales, with flexible aspect ratios. With this rich set of basis images, the contourlet transform can effectively capture the smooth contours that are the dominant features in natural images with only a small number of coefficients. We begin with a detailed study on the statistics of the contourlet coefficients of natural images, using histogram estimates of the marginal and joint distributions, and mutual information measurements to characterize the dependencies between coefficients. The study reveals the non-Gaussian marginal statistics and strong intra-subband, cross-scale, and cross-orientation dependencies of contourlet coefficients. It is also found that conditioned on the magnitudes of their generalized neighborhood coefficients, contourlet coefficients can approximately be modeled as Gaussian variables. Based on these statistics, we model contourlet coefficients using a hidden Markov tree (HMT) model that can capture all of their inter-scale, inter-orientation, and intra-subband dependencies. We experiment this model in the image denoising and texture retrieval applications where the results are very promising. In denoising, contourlet HMT outperforms wavelet HMT and other classical methods in terms of visual quality. In particular, it preserves edges and oriented features better than other existing methods. In texture retrieval, it shows improvements in performance over wavelet methods for various oriented textures.

I. INTRODUCTION

In image processing, it has been a common practice to use simple statistical models to describe 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 and denoising. A simple, accurate and tractable model is an essential element in any successful image processing algorithm.

Images can be better modeled in the *wavelet* transform [1],[2] domain, which shows multiresolution and time-frequency localization properties, and in which energy density has a more local structure than in the image spatial domain. Initially, wavelet transform was considered to be a good decorrelator for images, and thus wavelet coefficients were assumed to be independent and were simply modeled by marginal statistics [3]. Later it was realized that wavelet coefficients for natural images exhibit strong dependencies both across scales and between neighbor coefficients within a subband, especially around image edges. This gave rise to several successful joint statistical models in the wavelet domain [4]–[10], as well as improved image compression schemes [11]–[13].

The major drawback for wavelets in 2-D is their limited ability in capturing directional information. To counter this deficiency, researchers have most recently shifted their attention to multiscale and directional representations that can capture the intrinsic geometrical structures such as smooth directional contours in natural images. Some examples include the steerable pyramid [14], brushlets [15], complex wavelets [16], and the curvelet transform [17]. 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 [18]–[20] developed the *contourlet* representation based on an efficient two-dimensional nonseparable filter bank that can deal effectively with images having smooth contours. Contourlets not only possess the main features of wavelets (namely, multiresolution and time-frequency localization), but also show a high degree of directionality and anisotropy. The main difference between contourlets and other multiscale directional systems is that contourlets allow for a different and flexible number of directions at each scale, while achieving nearly critical sampling. In addition, contourlet transform employs iterated filter banks, which makes it computationally efficient.

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. Then we 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 the coefficients' dependencies across *scale* and *space*. However, as a “true” two-dimensional representation, contourlets allow us to also model the coefficients' dependencies 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 contourlet coefficients of natural images via 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 denoising and texture retrieval. Finally, a conclusion is presented in Section VII.

II. BACKGROUND

A. Contourlets

Do and Vetterli developed contourlets in [18]–[20]. Their primary aim was to construct a sparse efficient decomposition for two-dimensional signals that are piecewise smooth away from smooth contours. Such signals resemble natural images of ordinary objects and scenes, with the discontinuities as boundaries of objects. These discontinuities, referred to as *edges*, are gathered along one-dimensional smooth contours. Two-dimensional wavelets, with basis functions shown in Figure 1(a), lack directionality and are only good at catching zero-dimensional or point discontinuities, resulting in largely inefficient decompositions. For example, as shown in Figure 1(c), it would take many wavelet coefficients to accurately represent even one simple one-dimensional curve.

Contourlets were developed as an improvement over wavelets in terms of this inefficiency. The resulting transform has the multiresolution and time-frequency localization properties of wavelets, but also shows a very high degree of directionality and anisotropy. Precisely, 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 richness in the choice of basis functions, contourlets can represent any one-dimensional smooth edges with close to optimal efficiency. For instance, Figure 1(d) shows that compared with wavelets, contourlets can represent a smooth contour with much fewer coefficients.

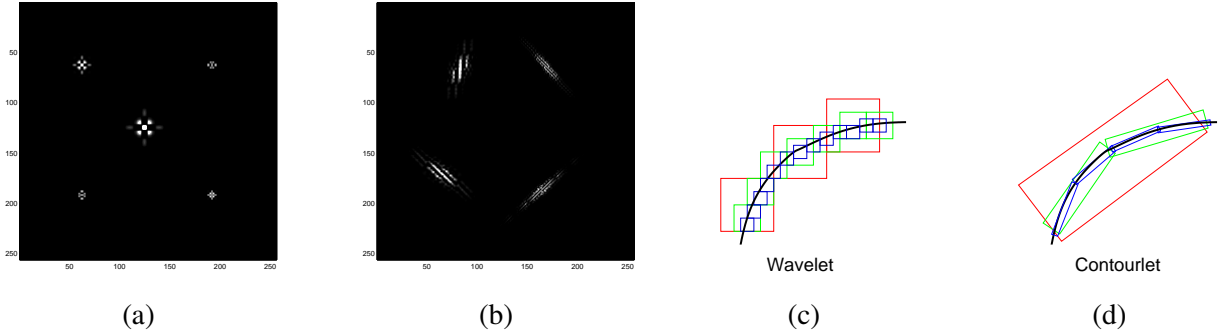


Fig. 1. Contourlet and wavelet representation for images. (a) Basis functions of 2-D wavelets (b) Basis functions of contourlets (c) Wavelets have square supports and can only capture points. (d) Contourlets have elongated supports and can capture line segments. Contourlets thus can effectively represent a smooth contour with fewer coefficients.

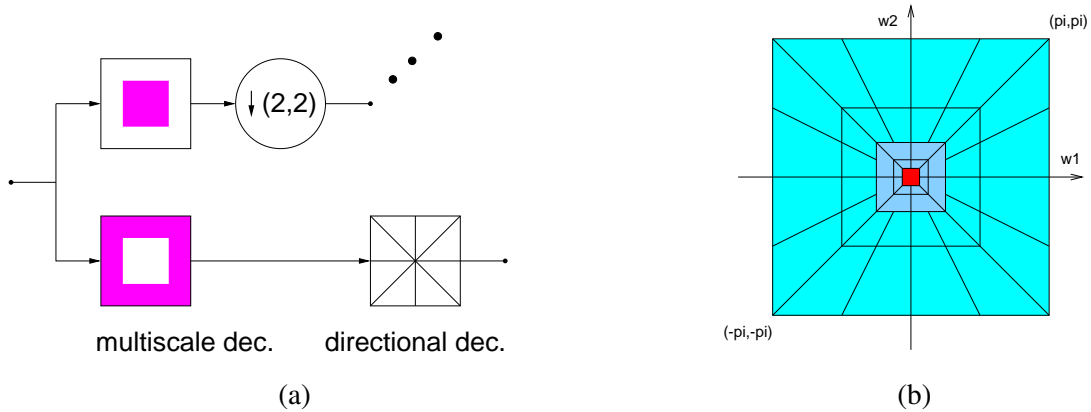


Fig. 2. (a) Pyramidal directional filter bank structure that implements the discrete contourlet transform. (b) A typical contourlet frequency partition scheme.

Contourlets are implemented by the pyramidal directional filter bank (PDFB) which decomposes images into directional subbands at multiple scales [18]–[20]. The PDFB is a cascade of a Laplacian pyramid [21] and a directional filter bank [22] as shown in Figure 2(a). The directional filter bank is a critically sampled filter bank that decomposes images into any power of two's number of directions. Due to the PDFB's cascaded structure, the multiscale and directional decompositions are independent of each other. One can decompose each scale into any arbitrary power of two's number of orientations and different scales can be divided into different numbers of orientations. This decomposition property makes contourlets a unique transform that can achieve a high level of flexibility in decomposition while being

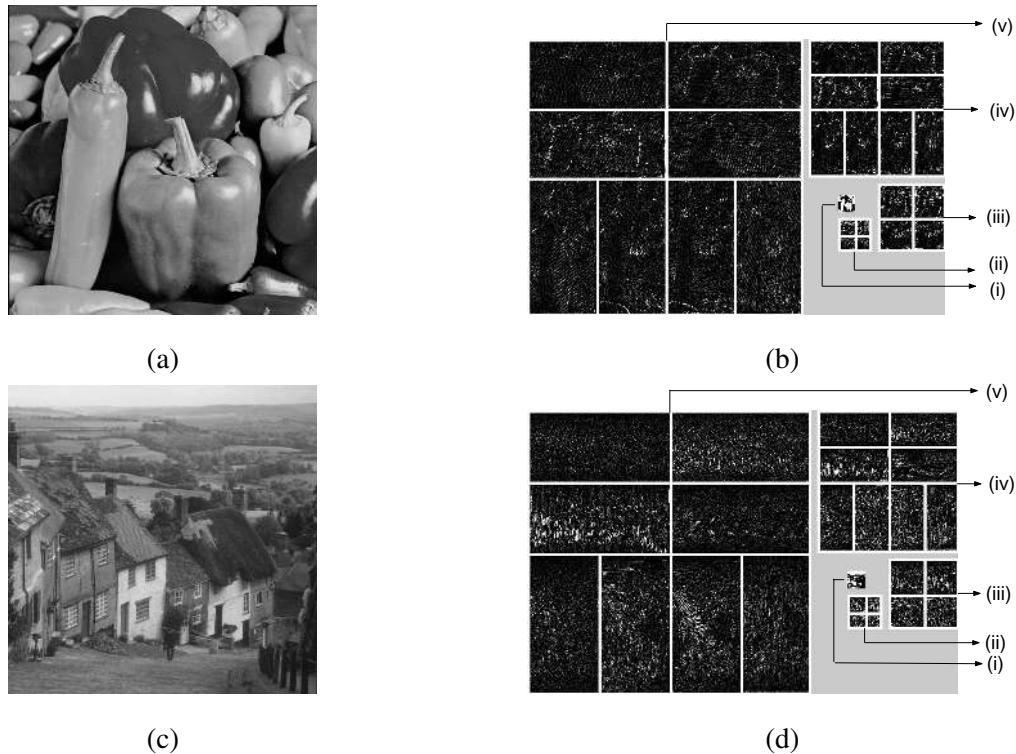


Fig. 3. (a) The “Peppers” image. (b) Contourlet representation of (a). (c) The “Goldhill” image. (d) Contourlet representation of (c). (i)–(v) represents coarse to fine scales respectively. Small coefficients are colored black while large coefficients are colored white.

close to critically sampled (up to 33% overcomplete, which comes from the Laplacian pyramid)¹. Other multiscale directional transforms have either a fixed number of directions, such as complex wavelets [16], or are significantly overcomplete (depending on the number of directions), such as the steerable pyramid [14]. Figure 2(b) shows a typical frequency division of the contourlet transform where the four scales are divided into four, four, eight and eight subbands from coarse to fine scales respectively. The fact that contourlets are close to critically sampled makes them especially promising in image compression.

Figure 3(a) shows the image “Peppers” and Figure 3(b) shows its contourlet representation. Similarly, Figure 3(c) shows the image “Goldhill” and Figure 3(d) shows its contourlet representation. In this particular decomposition, the image is divided into an approximation image (i) and four detail scales (ii), (iii), (iv), (v) from coarse to fine. Each detail scale is further partitioned into directional subbands according to the scheme in Figure 2(b). The two coarser scales are partitioned into four directional

¹Recently, a modified version of the contourlet scheme that is critically sampled was developed [23].

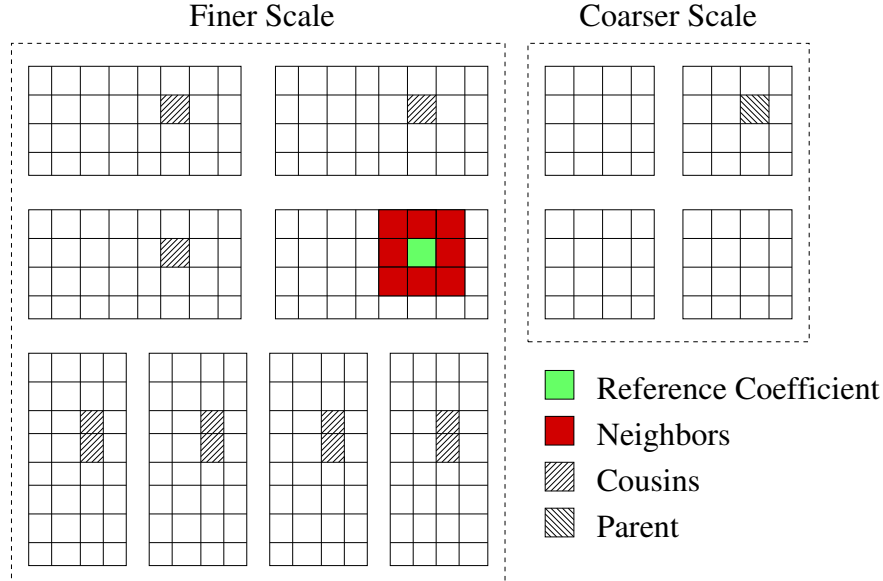


Fig. 4. Contourlet coefficient relationships.

subbands, and the two finer scales are partitioned into eight subbands. In the case of partitioning into eight or more directions, the directional subbands that correspond to horizontal directions differ in shape from the vertical direction subbands.

B. Coefficient Relationships

We define some important contourlet coefficient relationships 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 coefficient has one parent and four children. We also define coefficients in the same scale and same spatial location but different directions as *cousins* (CX). This cross-orientation relationship is more important for contourlets than for wavelets as contourlets have more directions.

From Figure 4, it can be observed that in certain directions there can be multiple cousins. Specifically, coefficients in the horizontal subbands have multiple cousins in the vertical subbands, and those in the vertical subbands have multiple cousins in the horizontal subbands. The reason is that the basis functions corresponding to the vertical and horizontal subbands are defined over different supports [19]. Figure 5 shows some typical contourlet bases oriented at certain vertical and horizontal directions, respectively. With different sampling grids, their basis functions are also defined on different supports. Thus, each

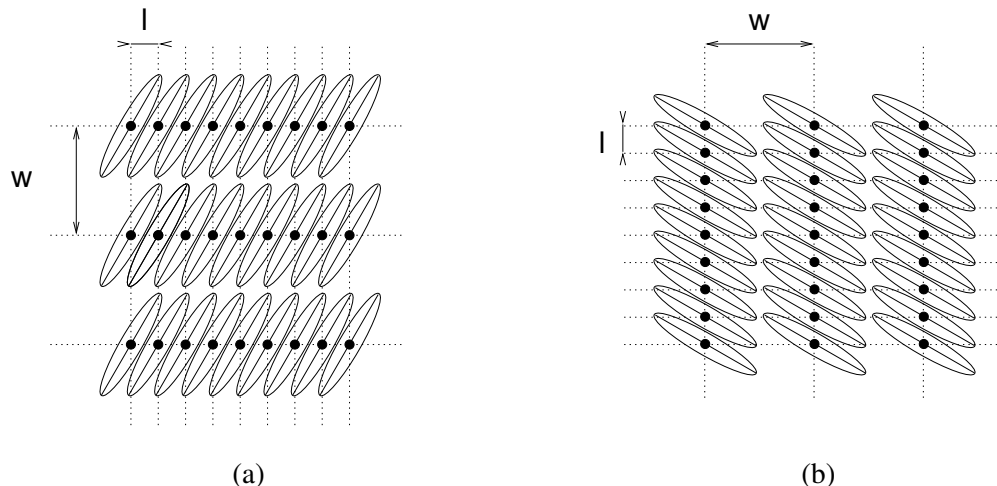


Fig. 5. Vertical and horizontal contourlet basis functions with their different sampling grids and supports: (a) Vertical bases and (b) horizontal bases.

vertical basis overlaps with multiple horizontal bases, and vice versa. We refer to all such bases with overlapped supports as cousins.

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

III. CONTOURLET STATISTICS

A. Marginal Statistics

We first study the marginal statistics of the contourlet coefficients of natural images. Figure 6 plots the histograms of two finest subbands of the image “Peppers.” The distributions are characterized by a very sharp peak at zero amplitude and extended tails to both sides of the peak. This implies that the contourlet transform is very sparse, as the majority of coefficients have amplitudes close to zero. The kurtoses of the two distributions are measured at 24.50 and 19.40, which are much higher than the kurtosis of 3 for Gaussian distributions. Similar distributions are also observed for all subbands of other test images. Thus, the marginal distributions of natural images in contourlet domain are highly non-Gaussian.

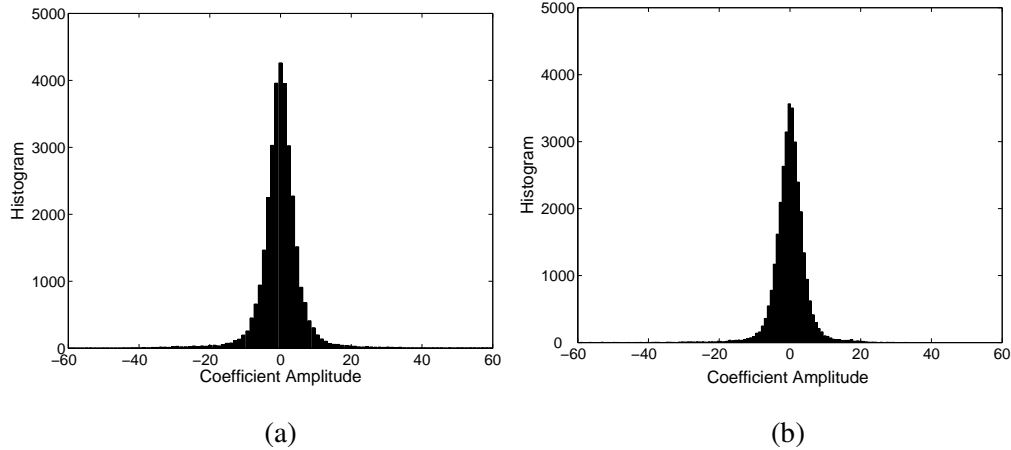


Fig. 6. Marginal statistics of two finest subbands of the image “Peppers.” The kurtoses of the two distributions are measured at (a) 24.50 and (b) 19.40, showing that the coefficients are highly non-Gaussian.

B. Joint Statistics

Marginal statistics only describe individual behaviours of transform coefficients without accounting for their dependence. Indeed, it is clear that contourlet coefficients are dependent on each other since basis functions that overlap and align with image edges lead to significant coefficients. A simple inspection of Figure 3(b) and (d) confirms this dependence. Significant coefficients tend to cluster spatially around the edges of objects in the images, and also persist to other scales and other orientations.

We characterize coefficient dependencies by measuring joint statistics. Joint statistics are particularly important because in the wavelet case, image processing algorithms exploiting joint statistics of coefficients [5]–[7],[11],[12] show significant improvements in performance over those that exploit marginal statistics alone [3],[8]. As the contourlet transform is similar to the wavelet transform, it is natural to extend this assumption to the contourlet case as well.

Figure 7 shows the conditional distributions of contourlet coefficients, conditioned on their parents $P(X | PX)$, neighbors $P(X | NX)$, and cousins $P(X | CX)$ for the image “Peppers.” All three of the conditional distributions show a common “bow-tie” shape where the variance of the coefficients is directly related to the conditioned coefficient magnitude. This indicates two properties of the coefficients. First, even though the coefficients are correlated due to the overcomplete nature of the contourlet transform, they are *approximately* decorrelated with the conditional expectation $E[X | \cdot] \approx 0$. Second, the distribution of the coefficients varies with the conditioned magnitude and therefore the coefficients are dependent

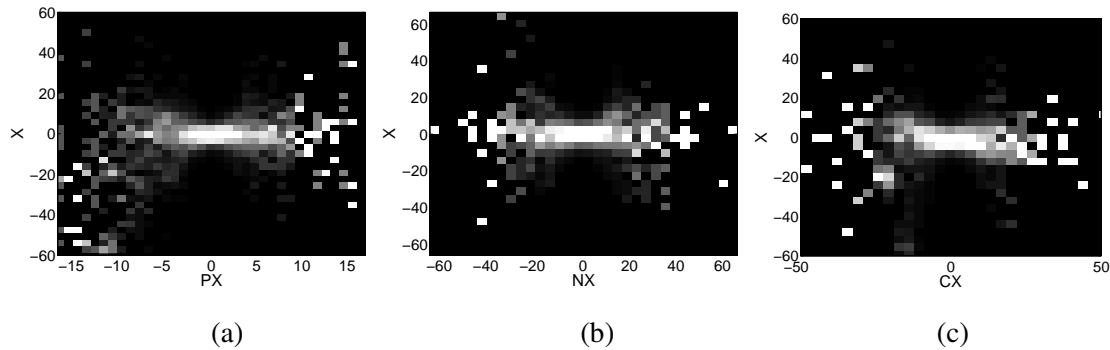


Fig. 7. Conditional distribution of a finest subband of “Peppers,” conditioned on (a) parent $P(X | PX)$, (b) neighbor $P(X | NX)$, (d) cousin $P(X | CX)$.

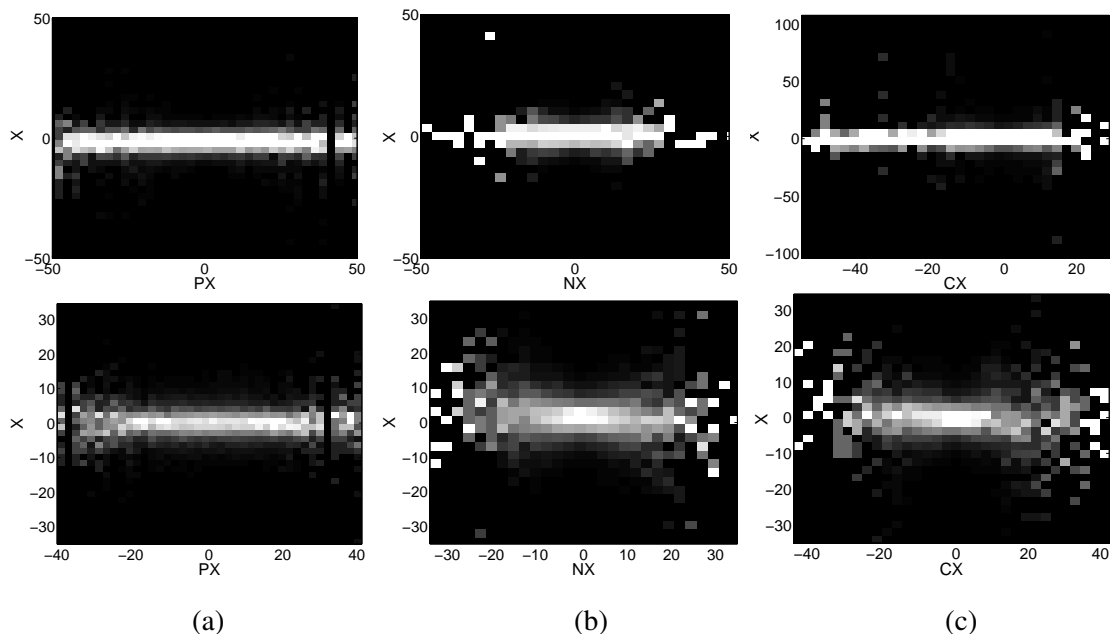


Fig. 8. Distribution of a finest subband of “Peppers” (top row) and “Goldhill” (bottom row), conditioned on (a) ancestors, (b) neighbors, and (c) cousins, all at distances of three coefficients away .

on one another. Contourlet coefficients of natural images are therefore approximately *uncorrelated yet dependent* on each other.

The dependence, however, is local and does not extend to farther coefficients. Figure 8 shows the conditional distributions of coefficients on distant coefficients, for the two images “Peppers” and “Goldhill.” In all cases for “Peppers,” the conditional distributions are approximately invariant to the conditioned

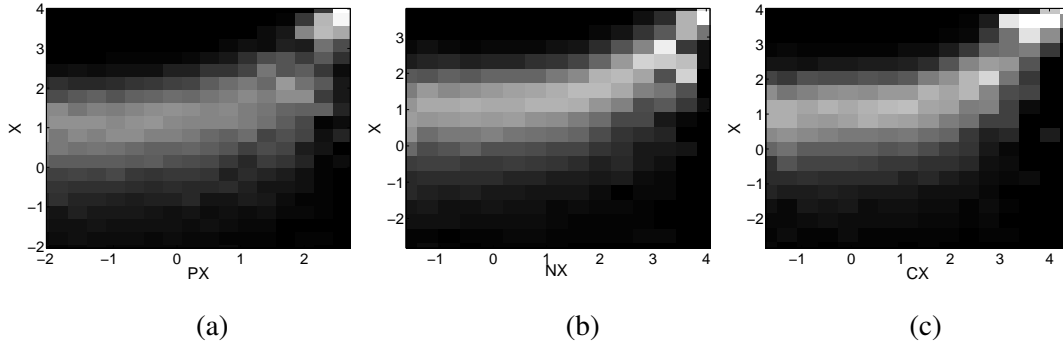


Fig. 9. Distribution of log magnitude of a finest subband of the image “Peppers” conditioned on log magnitude of (a) parent, (b) neighbor, (c) cousin.

value, indicating independence. For “Goldhill,” dependence still exists conditioned on distant neighbors and cousins. The difference between the two images is that “Peppers” is an edge-dominant image with very little texture, while “Goldhill” is a more textured image. A large coefficient in “Peppers” most likely corresponds to an edge in the image. If the edge is not perfectly straight, distant neighbors will be small because the same edge at distant locations do not align along the same direction. Meanwhile, a large coefficient in “Goldhill” is most likely due to its texture. Textures usually have similar directional patterns over relatively large areas of space. Thus, we can assume Markovian behaviors for the contourlet coefficients of smooth images that do not contain excessive textures or details.

Figure 9 shows the conditional distributions in log-log scale. The right sides of these plots all show a unit slope distribution, suggesting a linear relationship between the magnitudes of the coefficients and their generalized neighborhoods when their generalized neighborhoods’ magnitudes are sufficiently large. The left sides of all the plots display a horizontal line, indicating independence when the magnitudes of the coefficients’ generalized neighborhoods are small. This conditional distribution is also observed in the studies of wavelet coefficients [4],[13]. If we assume that the horizontal distributions of the left sides of the plots are due to quantization errors and other sources of uncertainties dominating at small coefficient magnitudes, contourlet coefficients of natural images can be modeled according to some distributions with variances directly related to any linear combination of the magnitudes of their generalized neighborhoods.

It is also of interest to investigate the conditional distributions $P(X | PX = px)$, $P(X | NX = nx)$ and $P(X | CX = cx)$, where px , nx , and cx are some fixed values, as shown in Figure 10. These distributions are the vertical cross-sections of the two-dimensional distributions in Figure 7. The kurtoses of the conditional distributions are measured at 3.90, 2.90, and 2.99, conditioned on the coefficients’

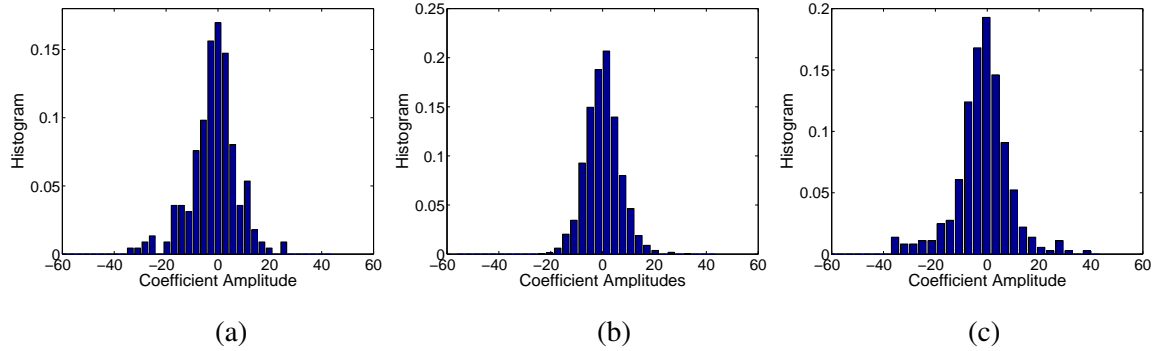


Fig. 10. Conditional distribution of the coefficients of a finest subband of “Peppers” on (a) their parents $P(X | PX = px)$, (b) neighbors $P(X | NX = nx)$, and (c) cousins $P(X | CX = cx)$. The kurtoses of the distributions are measured at 3.90, 2.90, and 2.99, respectively.

parents, neighbors, and cousins, respectively. The kurtoses are very close to the kurtosis of 3 for Gaussian distributions, indicating that the conditional distributions of contourlet coefficients can be well-modeled by Gaussian distributions. Compared to the unconditional distributions in Figure 6, for which the kurtoses are highly non-Gaussian at around 20, it can be concluded that contourlet coefficients are *non-Gaussian but conditionally Gaussian*. Since in Figure 7, the vertical cross-section of each column can be approximated as a Gaussian distribution, contourlet coefficients can be accurately modeled using a mixture of Gaussian distributions of different variances.

IV. DEPENDENCE CHARACTERIZATION VIA INFORMATION-THEORETIC ANALYSIS

In this section, we propose a quantitative study on the joint statistics of contourlet coefficients in complement to the qualitative study in the previous section, using mutual information as a simple quantitative measure of dependencies [24].

A. Mutual Information and Estimation

Mutual information between two random variables X and Y is defined as [25]

$$I(X; Y) = \int \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (1)$$

where X and Y can be scalars or vectors. In the vector case, a density function of three or higher dimensions is required to evaluate the mutual information, and the integrals are defined over the entire domain spanned by the vectors. Qualitatively, mutual information can be interpreted as how much

information one random variable contains about another. Its value increases with increasing dependence between the two variables. It is zero when the two variables are independent and is positive infinity when they are functions of each other.

We resort to nonparametric histogram estimation of mutual information between contourlet coefficients and their generalized neighborhoods. For general mutual information and entropy estimation, readers are referred to [26],[27]. Specifically, between any two contourlet coefficients X and Y , we use the estimator [26]

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

where k_{ij} is the number of coefficients in a subband observed in 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 coefficients in the subband considered, and J and K are the number of histogram bins along the X and Y directions respectively. Note that the first term is the mutual information histogram estimate and 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 lowerbound [26].

One way to tighten the estimation bound is by choosing J and K to give the maximum estimate in (2). In the case of infinite samples, then

$$\lim_{J \rightarrow \infty, K \rightarrow \infty} \hat{I}(X; Y) = I(X; Y) \quad (3)$$

However, in dealing with images, we can only have finite samples (i.e. the finite sizes of the images), and (3) is no longer true. Figure 11 plots the histogram estimate of the inter-scale mutual information of the image ‘‘Peppers’’ as a function of J and K . The plot shows that with increasing J and K , the estimate first increases to a global maximum point and then decays rapidly. The maximum point corresponds to the histogram resolution of the best estimator. However, locating this unique point is difficult as it varies with different numbers of samples and different images. Empirically, we found that using

$$J = K = \text{round}\left(\frac{N}{3000}\right) + 12 \quad (4)$$

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

A scenario where the estimator in (2) is inaccurate is the estimation of high dimensional mutual information. For instance, consider estimating the mutual information between coefficients and any subset of their generalized neighborhoods $I(X; \widetilde{N}X_1, \widetilde{N}X_2, \dots, \widetilde{N}X_M)$, where $\widetilde{N}X_i$ is a generalized

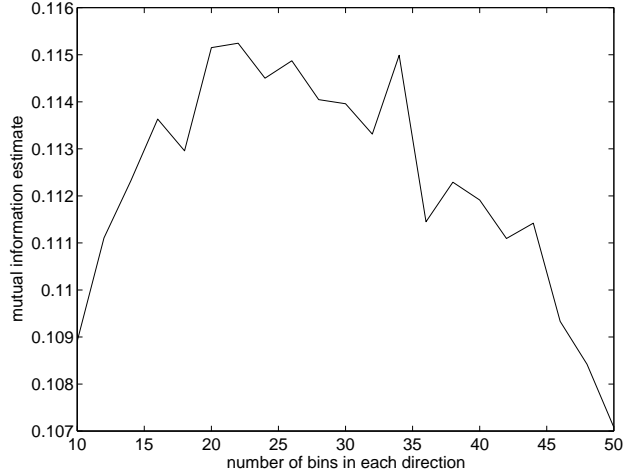


Fig. 11. Mutual information estimate for the “Peppers” image as a function of the number of bins used in the estimate for each dimension.

neighborhood coefficient for each $i = 1, 2, 3, \dots, M$ and M is the subset size. As the dimensionality M increases, estimation accuracy decreases exponentially [28]. In such cases, we follow the approach in [24] to replace the high dimensional variable with its sufficient statistic

$$T = \sum_{i=1}^M a_i | \widetilde{N\tilde{X}}_i | \quad (5)$$

where a_i is the weight of the i^{th} generalized neighbor coefficient. Note that $I(X; T)$ is only a two-dimensional statistic and therefore can be accurately estimated. From the data processing inequality, $I(X; T)$ bounds $I(X; \widetilde{N\tilde{X}}_1, \widetilde{N\tilde{X}}_2, \dots, \widetilde{N\tilde{X}}_M)$ from below [25]

$$I(X; T) \leq I(X; \widetilde{N\tilde{X}}_1, \widetilde{N\tilde{X}}_2, \dots, \widetilde{N\tilde{X}}_M) \quad (6)$$

Therefore, we can obtain the tightest bound in (6) by choosing a_i to maximize $I(X; T)$, using optimization algorithms such as the Matlab Optimization Toolbox.

B. Estimation Results

We estimated the mutual information for various 512×512 natural images. These images vary from simple edge-dominant images such as “Peppers” to highly textured images such as “Goldhill” and “Barbara.” Unless stated otherwise, we used the 9-7 pyramidal filters (9-7 filters) and the McClellan transformed directional filters from the 9-7 filters, proposed by Cohen and Daubechies (CD filter) [29] for the contourlet transform. We partitioned both the finest and second finest scales into eight directional

TABLE I

MUTUAL INFORMATION ESTIMATE. PX REFERS TO THE PARENT OF X ; NX IS A LINEAR COMBINATION OF THE SET OF EIGHT SPATIAL NEIGHBORS WITHIN THE SAME SUBBAND; CX IS A LINEAR COMBINATION OF SEVEN COUSINS IN DIFFERENT DIRECTIONAL SUBBANDS.

	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

subbands. Therefore, each finest subband was of size 256×128 or 128×256 and each second finest subband was of size 128×64 or 64×128 , depending on its direction. Furthermore, as stated in Section II, each coefficient can have a unique cousin in some directions but multiple cousins in others. To avoid bias in our calculations towards the directions with multiple cousins, we used only the average magnitude of the multiple cousins in each of those directions. We present the results for three representative images “Lena,” “Barbara,” and “Peppers,” noting that the other images give similar results.

Table I shows the mutual information estimation results. Note that all images show significant mutual information across all of scale, space and directions, and reinforces our observation in Section III that coefficients are dependent on their generalized neighborhoods. For all images 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. At fine scales, most edges are approximately straight lines and overlap and align with the basis support of adjacent neighbor coefficients. Therefore, information about edges is shared by neighbor coefficients and neighbor coefficients can be accurate predictors of each other. This is especially true for highly textured images such as “Barbara” since for such images, large coefficients correspond to highly textured areas which span relatively large image areas with similar directional patterns.

The mutual information estimates in Table I are highly dependent on the choice of filters and the directional partition scheme used. Table II shows the mutual information estimates for “Lena” using different combinations of the 9-7 or Haar pyramidal filters and two directional filters: the CD filter or

TABLE II
MUTUAL INFORMATION ESTIMATE FOR CONTOURLET AND WAVELET REPRESENTATION OF “LENA” 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

the ladder filter by Phoong et al. (PKVA filter) [30]. From the table, the inter-scale, intra-subband, and inter-orientation mutual information are all much higher for the contourlet coefficients obtained using the Haar pyramidal filter than for those obtained using the 9-7 filters. This suggests that the 9-7 filters are superior to the Haar filter in terms of whitening the contourlet coefficients. Similarly, using different directional filters yields different inter-orientation mutual information, with the CD filter resulting in higher mutual information than the PKVA filter. Therefore, the PKVA filter whitens the coefficients more effectively and should give good performance in applications. It should be noted that there is ongoing effort to design more effective directional filters and mutual information is expected to decrease further.

As a comparison, Table II also displays the mutual information estimates of wavelet coefficients for the image “Lena,” using different wavelet filters. Comparing wavelets and contourlets, wavelet coefficients exhibit similar inter-scale and intra-subband dependencies but much lower inter-orientation dependencies. Thus for contourlets, cousin coefficients carry more significant information than in the case of wavelets. Again note that as more effective directional filters for contourlets are developed in the future, it is expected that for contourlets, $I(X; NX)$ and $I(X; CX)$ should further decrease.

Table III compares the inter-orientation and intra-subband mutual information for different directional partitioning schemes to partition the finest scale into 4, 8, and 16 orientations. Each further partition increases the inter-orientation dependency and decreases the intra-subband dependency.

Despite the observation where neighboring coefficients are the most important in terms of joint

TABLE III

MUTUAL INFORMATION ESTIMATE FOR DIFFERENT PARTITIONS OF THE FINEST SUBBANDS. DATA ARE OBTAINED USING 9-7 PYRAMIDAL FILTERS AND CD DIRECTIONAL FILTER 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

statistics, incorporating neighbors into models is not straightforward. Indeed, as every coefficient has eight neighbors, using a contextual model would result in a complex dependence network which is undesirable. Ideally, to simplify the model, we want every coefficient to be modeled to depend on only a single coefficient. We measure and compare the mutual information between the coefficients and each of their neighbors and cousins individually rather than as a collective set. To avoid keeping track of excessive data, we reduce the data size by grouping the measurements and summarizing each group into an average. We group the neighbors into first (directly adjacent) and second (diagonally adjacent) order neighbors and cousins into first, second and third order cousins (where the order refers to the distance between the two orientation subbands).

The results are presented in Table IV. When comparing individually, for “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 . The parent becomes the most significant predictor when generalized neighborhood coefficients are considered individually. The exception is “Barbara,” where $I(X; NX_i) > I(X; CX_i) > I(X; PX)$, for all i . In general, for images with high levels of textures, most large coefficients correspond to image areas with high image activities. Neighbor coefficients remain the dominant predictors in such image areas.

V. IMAGE MODELING

Based on the statistics measured in Sections III and IV, we propose models for contourlet coefficients of natural images. Recall from Section III and IV that contourlet coefficients of natural images possess the following properties:

- (a) Contourlet coefficients are marginally non-Gaussian.
- (b) They are dependent on their generalized neighborhood coefficients.
- (c) Conditioned on any linear combination of the magnitudes of their generalized neighborhood coefficients, they are Gaussian distributed.

TABLE IV

AVERAGE MUTUAL INFORMATION ESTIMATE FOR EACH PARENT, NEIGHBOR, AND COUSIN. NX_i REFERS TO i^{th} ORDER NEIGHBORS WHERE $i = 1$ MEANS THE FOUR ADJACENT NEIGHBORS AND $i = 2$ ARE THE FOUR DIAGONAL NEIGHBORS; CX_j REFERS TO THE j^{th} ORDER COUSINS WHERE $j = 1$ REFERS TO THE TWO IMMEDIATE ADJACENT ORIENTATIONS, $j = 2$ ARE THE NEXT TWO ORIENTATIONS, 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.

	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

We wish to develop a model that can incorporate the above properties while possessing other desirable characteristics. Specifically, the ideal model should

- (i) accurately model properties (a) to (c) listed above,
- (ii) have a simple structure to enable fast and efficient operations including fast training algorithms,
- (iii) be defined on a small number of parameters to allow accurate training with finite data.

It is suggested in Section III that contourlet coefficients can be accurately modeled using a mixture of Gaussian distributions of different variances. We consider the hidden Markov model (HMM) family [7], which is one of the most well-known and widely used family of Gaussian mixture models.

A zero-mean 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 normal distribution with the corresponding variance of the state. This satisfies property (c) of contourlet coefficients. Note that the unconditional distribution is non-Gaussian (property (a)). Therefore, each coefficient is characterized by the set θ of state probabilities p and standard deviations σ

$$\theta = \{p^1, p^2, \dots, p^N, \sigma_1, \sigma_2, \dots, \sigma_N\} \quad (7)$$

where $1, 2, \dots, N$ denote the state. To reduce the number of parameters, we “tie” all coefficients in the same subband to use the same set of statistics θ .

The HMM models the inter-coefficient dependencies by establishing links between the *hidden states* of dependent coefficients, satisfying property (b). Associated with each link is a state transition probability matrix $\mathbf{A}_{m,n}$

$$\mathbf{A}_{m,n} = \begin{pmatrix} a_{m,n}^{1,1} & a_{m,n}^{1,2} & \dots & a_{m,n}^{1,N} \\ a_{m,n}^{2,1} & a_{m,n}^{2,2} & \dots & a_{m,n}^{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,n}^{N,1} & a_{m,n}^{N,2} & \dots & a_{m,n}^{N,N} \end{pmatrix} \quad (8)$$

where $a_{m,n}^{k,l}$ is the probability that the coefficient m is in state k given the coefficient n is in state l .

It should be noted that Gaussian mixture models outside the HMM family exist [10],[31]. However, these models, by using continuous mixtures of Gaussian distributions, suffer from increased complexity and are therefore disadvantaged according to criterion (ii) of the ideal model.

Within the HMM family, many forms and variations exist [7],[9],[32]–[35]. With criteria (i)–(iii) of the ideal model in mind, we choose the two-states hidden Markov tree (HMT) model [7].

The HMT model is an HMM that adopts a quad-tree dependency structure. Each coefficient is modeled to depend on its parent. In other words, the model establishes links between the hidden state of each coefficient and those of its parent and its four children, resulting in coefficients linked as a quad-tree. Associated with each link is a state transition matrix specified in (8).

Note that the quad-tree structure of the HMT model allows for direct modeling of inter-scale dependencies, while also modeling indirectly the dependencies between neighboring coefficients within the same subband, via their links to their common ancestors. This is consistent with the observation in Section IV that for most images, the parent coefficient is the most significant predictor when the generalized neighborhood coefficients are compared *individually*. Although the neighbor coefficients can become more significant when combined, direct modeling the dependencies on all the neighbors require a one-to-many dependence relationship which will significantly increase model complexity, thereby violating criterion (ii).

Criterion (ii) states that the ideal model should have a simple structure. The HMT model has a simple tree structure, which enables efficient training using the expectation maximization (EM) algorithm. Moreover, with only two states, the model requires a relatively small number of parameters, thus also satisfying criterion (iii). In particular it has been shown that using only a mixture of two Gaussians, the HMT can already achieve satisfactory accuracy in wavelet coefficient modeling [7]. Two-states mixture models also possess an intuitive appeal that their two states can match to the two features of natural images, namely edges and smooth areas.

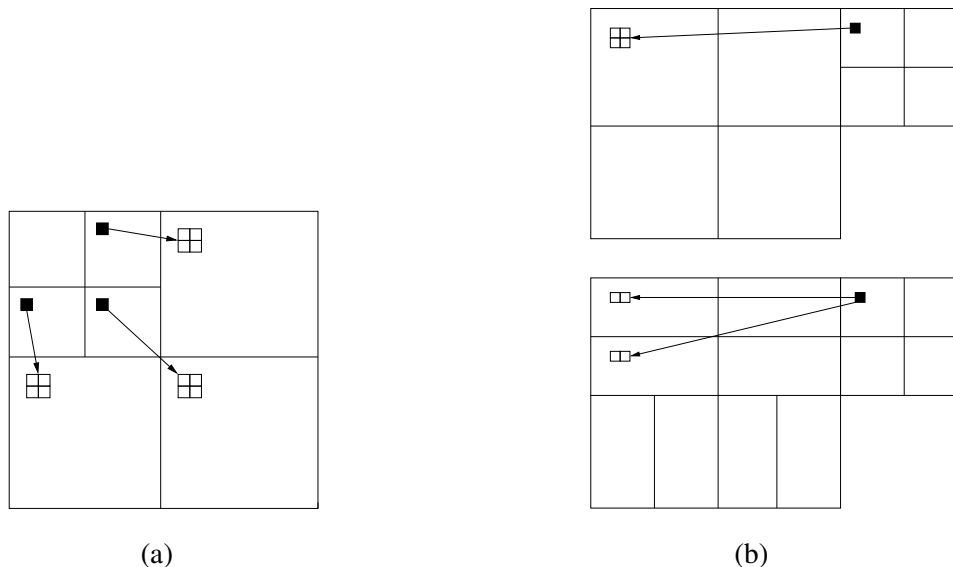


Fig. 12. Parent-children relationship for (a) wavelets and (b) two possible contourlet structures. Black squares represent parent coefficients with white squares as their children. Notice for the bottom contourlet structure, a parent coefficient can have its children spread over two subbands.

Other forms and variations of HMT such as those found in [9],[32]–[35], are usually based on more complex structures than the HMT to gain improved accuracy. For example, the models in [34] and [35] are based on enlarged parameter sets that they have to rely on an extra initialization step in training. Again, we base our decision on simplicity (criteria (ii) and (iii)) and choose the HMT model.

One major difference between the wavelet HMT and the contourlet HMT models is that the contourlet HMT model also accounts for inter-orientation dependencies while the wavelet HMT does not. Figure 12 illustrates this difference. In wavelet decompositions, the four children of any parent coefficient are always in the same directional subband. As a result, each tree within the model is entirely in one direction. As the HMT model models each tree independently, dependencies across directions are not modeled. Conversely, contourlet coefficients can have their four children in two separate directional subbands. Thus, contourlet HMTs can span several adjacent directions in the finer scales and therefore, inter-orientation dependencies are modeled. In other words, the contourlet HMT model effectively captures the dependencies across all of scale, space and orientation. The software for the implementation of this model can be downloaded from <http://www.ifp.uiuc.edu/~minhdo>.

For a contourlet decomposition of J scales and m_1, m_2, \dots, m_J directional subbands within each scale $1, 2, \dots, J$ respectively, the contourlet HMT parameter set of each tree contains:

- $\vec{p}_{1,1}$ for the root state probabilities
- $\mathbf{A}_{j,k}$ where $j = 2, \dots, J$, $k = 1, \dots, n_J$, and $n_x = \frac{m_x}{m_1}$; for the state transition probability matrix between subband k in scale j and its parent subband in scale $j - 1$
- $\vec{\sigma}_{p,q}$ where $p = 1, \dots, J$, $q = 1, \dots, n_J$, and $n_x = \frac{m_x}{m_1}$; for the Gaussian standard deviations of subband q in scale p

Each image is then characterized by the HMT of each root subband

$$\theta_{image} = \{\theta_1, \theta_2, \dots, \theta_{m_1}\} \quad (9)$$

where m_1 is the number of root subbands.

As in the wavelet HMT model, the parameters can be trained using the EM algorithm that iterates upwards and downwards through the trees to maximize the likelihood function $E[x | \theta]$ [7]. Counting the number of free model parameters, we count one free parameter in each $\vec{p}_{1,1}$ and two in each matrix $\mathbf{A}_{a,b}$ (the other parameters being constrained to sum with these parameters to unity), and two in each $\vec{\sigma}_{a,b}$. Therefore, this model requires four parameters for each subband and three for each root subband. For the subband partitioning scheme in Figure 2(b) where the coarse to fine scales are divided into 4, 4, 8, and 8 subbands respectively, the model uses 92 free parameters in total.

Table V shows the HMT parameters of the image ‘‘Peppers.’’ As shown in the table, for typical natural images, the two states of each subband generally have very different variances, where the small variance state models the smooth parts of the image and the large variance state models the image edges. The state transition matrices also reveal the persistence property, where there is a high probability that parents and their children are all in the same large or small state.

VI. APPLICATIONS

A. Denoising

We apply the contourlet HMT model developed in the previous section in denoising of zero-mean additive white Gaussian noise. In theory, since contourlets are more sparse in representing natural images than wavelets, they should provide larger separations between images and noise, and thereby enable more effective denoising.

In the contourlet domain, we express

$$\vec{v} = \vec{u} + \vec{n} \quad (10)$$

where \vec{v} is the noisy contourlet coefficients, \vec{u} is the clean contourlet coefficients, and \vec{n} is the contourlet transform of white Gaussian noise. The problem is to estimate \vec{u} given \vec{v} . We experiment with noise of

TABLE V

HMT PARAMETERS OF THE IMAGE "PEPPERS." THE ELEMENTS OF THE MATRICES $\mathbf{A}_{x,y}$ ARE STATE TRANSITION PROBABILITIES FROM THE COLUMN STATE TO THE ROW STATE.

		direction 1		direction 2		direction 3		direction 4	
		state 1	state 2	state 1	state 2	state 1	state 2	state 1	state 2
$\vec{p}_{1,1}$		0.67	0.33	0.32	0.68	0.27	0.73	0.75	0.25
$\mathbf{A}_{2,1}$	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
$\mathbf{A}_{3,1}$	state 1	0.86	0.96	0.32	0.02	0.41	0.02	0.69	0.02
	state 2	0.14	0.04	0.68	0.98	0.59	0.98	0.31	0.98
$\mathbf{A}_{3,2}$	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
$\mathbf{A}_{4,1}$	state 1	0.00	0.82	0.35	0.00	0.55	0.01	0.38	0.00
	state 2	1.00	0.18	0.65	1.00	0.45	0.99	0.62	1.00
$\mathbf{A}_{4,2}$	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
$\vec{\sigma}_{1,1}$		160.10	60.16	258.30	84.76	165.04	59.00	168.01	76.82
$\vec{\sigma}_{2,1}$		59.88	12.29	98.37	18.79	70.06	21.79	70.83	27.18
$\vec{\sigma}_{3,1}$		7.71	7.68	76.11	11.62	20.62	7.90	39.47	6.90
$\vec{\sigma}_{3,2}$		37.59	9.65	27.89	8.69	41.66	8.53	21.77	8.31
$\vec{\sigma}_{4,1}$		14.69	3.20	33.65	5.26	16.05	3.57	17.76	4.38
$\vec{\sigma}_{4,2}$		16.47	3.43	31.46	5.20	17.91	3.66	17.57	4.34

various standard deviations as shown in Table VI. Note that since contourlet transform is an overcomplete transform, the transformed noise \vec{n} is no longer white.

We first fit an HMT model to the noisy contourlet coefficients \vec{v} to obtain the parameter set θ_v . From this model, we obtain the model of the clean contourlet coefficients θ_u by subtracting the power of the transformed noise, $\sigma_{i,j,k}^{2,(n)}$, from the variance $\sigma_{i,j,m}^{2,(v)}$ in θ_v , for each of the k^{th} coefficient in the j^{th} subband of the i^{th} scale, for the m^{th} state,

$$\sigma_{i,j,k,m}^{2,(u)} = (\sigma_{i,j,m}^{2,(v)} - \sigma_{i,j,k}^{2,(n)})_+ \quad (11)$$

where $\sigma_{i,j,k,m}^{2,(u)}$ is the variance of clean contourlet coefficients and $(x)_+ = x$ for $x \geq 0$ and $(x)_+ = 0$ for $x < 0$. The transformed noise power $\sigma_{i,j,k}^{2,(n)}$ can be estimated using the Monte-Carlo method: by repeatedly generating random white noise images and averaging their variance in the contourlet domain.

TABLE VI
PEAK SIGNAL-TO-NOISE RATIO (IN DECIBELS) OF DENOISED IMAGES.

image	noise standard deviation	noisy	wiener2 (5 × 5)	wavelet thresholding	wavelet HMT	contourlet HMT
Lena	10	28.13	33.03	32.00	33.84	33.38
	20	22.25	29.47	28.63	30.39	28.89
	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	10	28.11	31.38	29.90	31.36	29.18
	20	22.13	27.08	25.81	27.80	25.82
	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	10	28.13	34.06	33.37	35.33	33.45
	20	22.25	30.86	30.29	32.46	31.39
	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

With the HMT model parameters, the denoising problem can then be formulated into a Bayesian estimation problem where we try to estimate the quantity $E[u_{i,j,k} | v_{i,j,k}, \theta_{\mathbf{u}}]$ for all of k^{th} coefficients in the j^{th} subbands in the i^{th} scales. Note that given their states $S_{i,j,k}$, the coefficients are 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]

$$E[u_{i,j,k} | v_{i,j,k}, \theta_{\mathbf{u}}, S_{i,j,k} = m] = \frac{\sigma_{i,j,k,m}^{2,(u)}}{\sigma_{i,j,k}^{2,(n)} + \sigma_{i,j,k,m}^{2,(u)}} v_{i,j,k} \quad (12)$$

Equation (12) is the expression for the conditional mean of the contourlet coefficients conditioned on their states. We can thus introduce the state probability term $p(S_{i,j,k} = m | v_{i,j,k}, \theta_{\mathbf{u}})$ to obtain the optimal coefficient estimate unconditional on the coefficient states

$$E[u_{i,j,k} | v_{i,j,k}, \theta_{\mathbf{u}}] = \sum_m p(S_{i,j,k} = m | v_{i,j,k}, \theta_{\mathbf{u}}) \times \frac{\sigma_{i,j,k,m}^{2,(u)}}{\sigma_{i,j,k}^{2,(n)} + \sigma_{i,j,k,m}^{2,(u)}} v_{i,j,k} \quad (13)$$

The likelihood of the states $p(S_{i,j,k} = m | v_{i,j,k}, \theta_{\mathbf{u}})$ can be obtained from the EM algorithm as a byproduct during training [7].



Fig. 13. 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).

We applied the denoising algorithm described above to noisy versions of the test images and the results are shown in Table VI. Also in the table are the denoising results of the same noisy images using wavelet HMT and two other classical denoising methods, Matlab’s “wiener2” and wavelet thresholding (with threshold set to three times the noise standard deviation). In terms of peak signal-to-noise ratio (PSNR), wavelet HMT produces the best denoised results, followed closely by contourlet HMT. Both HMT methods easily outperform “wiener2” and wavelet thresholding.

A more useful and representative criterion for “goodness” of denoised images is by simply observing their visual quality. In terms of this more important quality, contourlet HMT produces superior denoised results. Figure 13 shows the denoising results of the “Zelda” image. It is clear that contourlet HMT removes most of the noise and produces the smoothest image. Figure 14 shows another example in a subimage of the “Barbara” image. It is shown in this result that contourlet HMT best preserves edges and

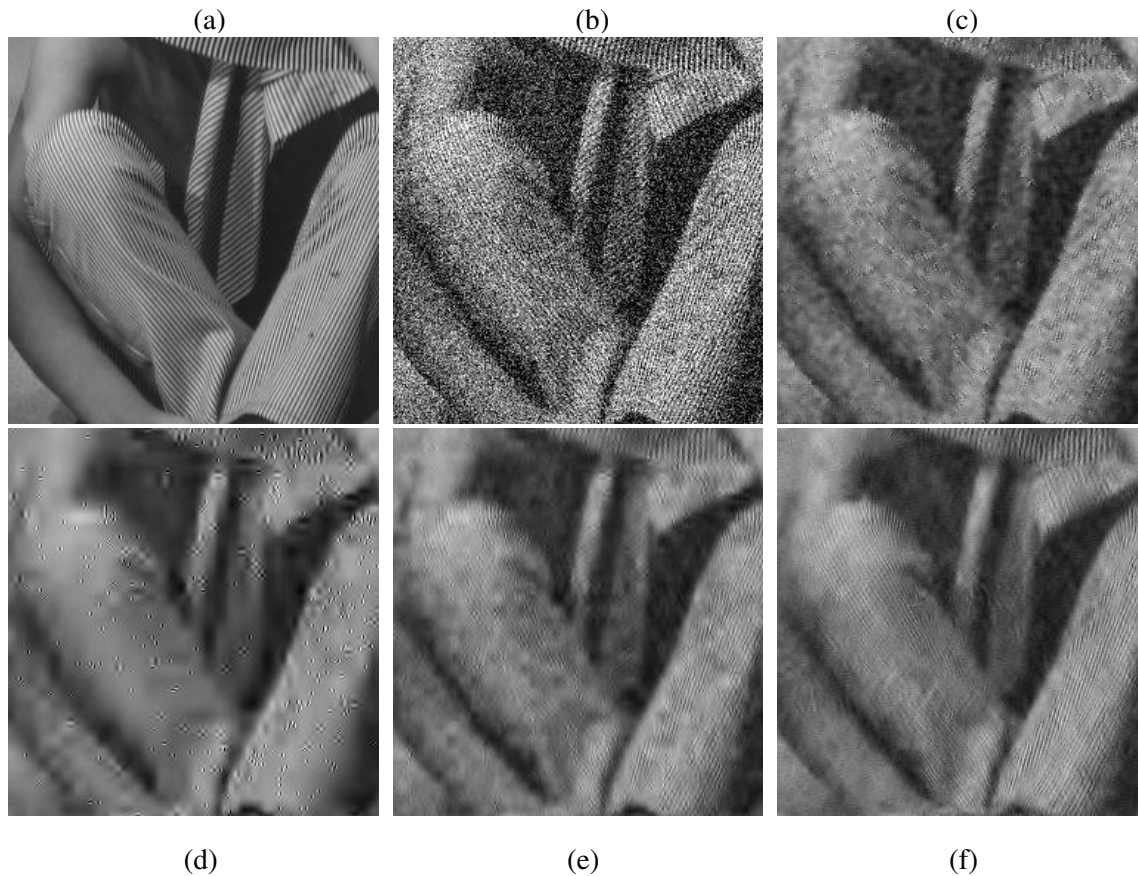


Fig. 14. Denoising results of “Barbara” image: (a) Barbara sub-image, (b) noisy image (noise standard deviation = 50, PSNR = 14.48dB), (c) wiener2 (PSNR = 22.57dB), (d) wavelet thresholding (PSNR = 21.96dB), (e) wavelet HMT (PSNR = 23.71dB), and (f) contourlet HMT (PSNR = 23.74dB).

oriented features. In the denoised results of all of “wiener2,” wavelet thresholding, and wavelet HMT, even though a large portion of noise is removed, the lines in the original image are also severely distorted. However, in the contourlet HMT denoised result, the lines are still clearly visible.

Recently, more complex wavelet-based denoising schemes [31],[33],[35]–[37] have been developed. While some have shown significant improvements in PSNR over the HMT algorithms, it is out of the scope of this paper to provide detailed comparisons with each denoising scheme. This paper has provided a comparison which shows that using the HMT model, contourlets achieve superior visual quality over wavelets.

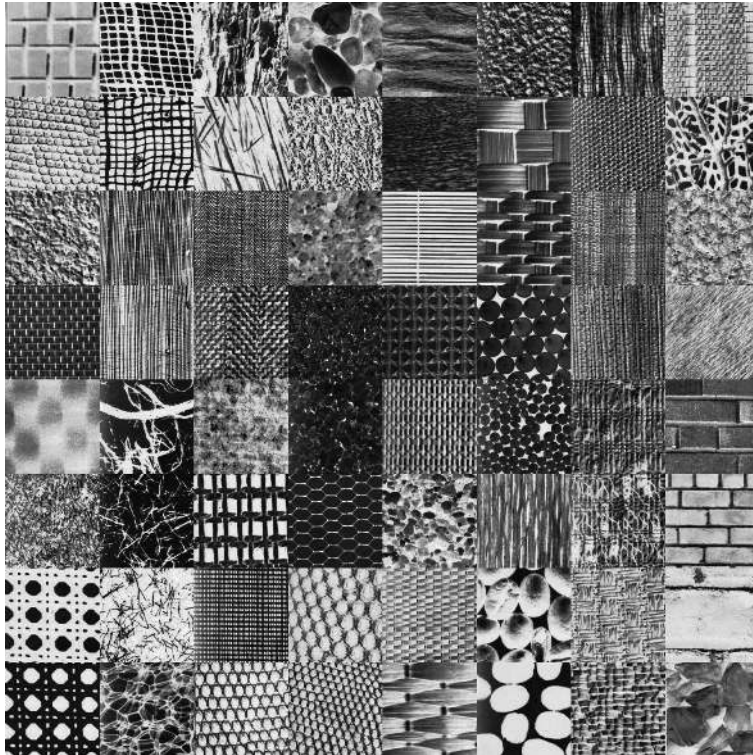


Fig. 15. Textures from the Brodatz database used in our texture retrieval experiments.

B. Texture Retrieval

The contourlet HMT model is then applied to content-based texture retrieval. The problem is to find, given one query homogeneous texture image, the texture images that are “most similar” to it within a large database collection of unlabelled texture images such as Figure 15, in a timely manner.

In the contourlet HMT texture retrieval system, the texture images are first transformed into contourlet coefficients. An HMT model is then fit to the contourlet coefficients, and the HMT model parameters are extracted as features. The Kullback-Liebler distances (KLD) between the query image model and each database image model are then measured and the database images that give the smallest KLD are retrieved. The KLD is defined as [25]

$$D(p(X | \boldsymbol{\theta}_q) \| p(X | \boldsymbol{\theta}_i)) = \int p(X | \boldsymbol{\theta}_q) \times \log \frac{p(X | \boldsymbol{\theta}_q)}{p(X | \boldsymbol{\theta}_i)} dx \quad (14)$$

Retrieval by minimization of KLD can be justified as retrieval by the maximum likelihood (ML) selection rule [20]. However, there is no closed form expression for the KLD between HMT models.

To estimate the KLD between two HMT models, one can resort to the Monte-Carlo method that

TABLE VII

AVERAGE RETRIEVAL RATES. ALL FIGURES ARE PERCENTAGES.

wavelet HMT	contourlet HMT
90.87	93.29

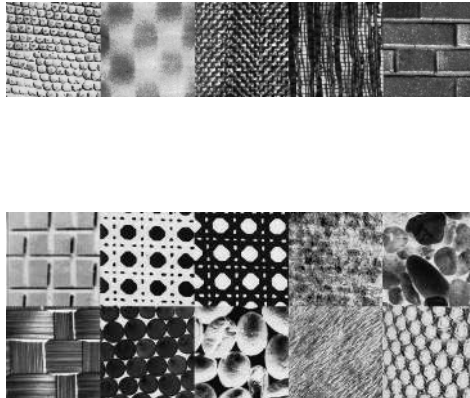


Fig. 16. Texture retrieval results. The top row shows the textures that are better retrieved by wavelets than by contourlets by at least 5%. The bottom two rows show the textures that are better retrieved by contourlets than by wavelets by at least 5%.

randomly generates multiple sets of data using the query model and then computes their likelihoods against each candidate model [20]. An alternative method is proposed by Do [38], which estimates the KLD using an upperbound based on the log-sum inequality. In this work, we only consider the Monte-Carlo method for its higher accuracy despite its slower speed.

In the retrieval experiments, a subset of the Brodatz image database [39] was used. This subset contains 64 different textures of size 512×512 and is shown in Figure 15. 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 15 matches for each query. We compare our results with wavelet HMT texture retrieval [20] and the results are shown in Table VII. Note that contourlet HMT gives an average retrieval rate approximately 2.5% higher than that of wavelet HMT.

Comparing the retrieval rates for individual textures, contourlet and wavelet HMT each give better retrieval rates for different textures. The top part of Figure 16 shows the texture images that are better

retrieved by wavelets than by contourlets by at least 5%, and the bottom part shows those better retrieved by contourlets by at least 5%. From the figure, it can be seen that the textures better retrieved by wavelets are all characterized by dominant vertical or horizontal directional components. In contrast, the textures better retrieved by contourlets generally demonstrate more diverse directional components (such as circular or irregular shapes). This shows the superior quality of contourlets in capturing directional information. The difference in retrieval rates between wavelet and contourlet HMT retrieval can vary from fractions of a percent to as large as near 20%. In any case, for most textures, contourlet HMT texture retrieval gives satisfactory texture retrieval performance with retrieval rates typically above 80%.

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 show Markovian dependencies in the form of local clustering and persistence across scales. Moreover, coefficients across adjacent orientations show more significant mutual dependencies compared to wavelets. Thus contourlet coefficients exhibit dependencies across all of scales, space, and orientations. Conditioned on these generalized neighborhood coefficient magnitudes, contourlet coefficients are approximately Gaussian.

These properties are verified quantitatively using mutual information measures. We have found that contourlet coefficients of natural images show very high intra-subband dependencies on all their eight neighbors, and also a rather high level of inter-orientation dependencies on their cousins. When comparing dependencies on each single coefficient, interscale dependencies become the highest for most images.

Based on the above properties, 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 HMT model exploits all of cross-scale, cross-orientation, and intra-subband dependencies through a tree-structured dependence network.

We applied the contourlet HMT model on denoising and texture retrieval. The results are highly promising. In denoising, contourlet HMT can preserve edges and lines better than wavelet HMT and other classical methods, and also produces superior visual quality in the denoised images. In texture retrieval, contourlet HMT gives higher retrieval rates than wavelets for textures that show high directionality. Both results suggest that contourlets can capture directional information very well, which is a highly valuable property in image processing.

REFERENCES

- [1] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Upper Saddle River, NJ: Prentice Hall, 1995.
- [2] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic Press, 1999.
- [3] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, July 1989.
- [4] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," in *Proceedings of the SPIE 44th Annual Meeting*, 1999, pp. 188–195.
- [5] M. K. Mihçak, I. Kozintsev, and K. Ramchandran, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, December 1999.
- [6] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1522–1531, September 2000.
- [7] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Transactions on Information Theory*, vol. 46, pp. 886–902, April 1998.
- [8] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proceedings of the Third International Conference on Image Processing*, 1996, pp. I. 379–382.
- [9] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, July 2001.
- [10] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Applied and Computational Harmonic Analysis*, vol. 11, no. 1, pp. 89–123, July 2001.
- [11] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3445–3462, December 1993.
- [12] 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 *Proceedings of IEEE Data Compression Conference*, 1997, pp. 221–230.
- [13] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 8, pp. 1688–1701, December 1999.
- [14] 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.
- [15] F. G. Meyer and R. R. Coifman, "Brushlets: a tool for directional image analysis and image compression," *Journal of Applied and Computational Harmonic Analysis*, vol. 5, pp. 147–187, 1997.
- [16] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis*, vol. 10, pp. 234–253, 2001.
- [17] E. J. Candès and D. L. Donoho, "Curvelets – a surprisingly effective non-adaptive representation for objects with edges," in *Curve and Surface Fitting*, A. Cohen, C. Rabut, and L. L. Schumaker, Eds. Nashville, TN: Vanderbilt University Press, 1999, pp. 105–120.
- [18] M. N. Do and M. Vetterli, "Contourlets," in *Beyond Wavelets*, J. Stoeckler and G. V. Welland, Eds. San Diego, CA: Academic Press, 2003. To Appear.
- [19] M. N. Do and M. Vetterli, "Contourlet: A computational framework for directional multiresolution image representation," submitted to *IEEE Transactions on Image Processing*, 2003.

- [20] M. N. Do, "Directional multiresolution image representations," Ph.D. dissertation, Swiss Federal Institute of Technology, Lausanne, 2001.
- [21] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, April 1983.
- [22] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: theory and design," *IEEE Transactions on Signal Processing*, vol. 40, pp. 882–893, April 1992.
- [23] Y. Lu and M. N. Do, "Crisp-contourlet: A critically sampled directional multiresolution image representation," in *SPIE Wavelet X Conference*, August 2003.
- [24] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1647–1658, November 2001.
- [25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: John Wiley and Sons, Inc., 1991.
- [26] R. Moddemeijer, "On estimation of entropy and mutual information of continuous distributions," *Signal Processing*, vol. 16, no. 3, pp. 233–246, March 1989.
- [27] J. Beirlant, E. J. Dudewicz, L. Gyöfi, and E. van der Meulen, "Non-parametric entropy estimation: An overview," *International Journal of Mathematical and Statistical Sciences*, vol. 6, no. 1, pp. 17–39, 1997.
- [28] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, UK: Chapman and Hall, 1986.
- [29] A. Cohen and I. Daubechies, "Nonseparable bidimensional wavelet bases," *Revista Matematica Iberoamericana*, vol. 9, no. 1, pp. 51–137, December 1993.
- [30] 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 Transactions on Signal Processing*, vol. 43, no. 3, pp. 649–665, March 1995.
- [31] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussian in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.
- [32] M. S. Crouse and R. G. Baraniuk, "Contextual hidden Markov models for wavelet-domain signal processing," in *Proceedings of 31st Asilomar Conference on Signal, Systems, and Computers*, November 1997.
- [33] G. Fan and X. -G. Xia, "Improved hidden Markov models in the wavelet-domain," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 115–120, January 2001.
- [34] G. Fan and X. -G. Xia, "Wavelet-based texture analysis and synthesis using hidden Markov models," *IEEE Transactions on Circuits and Systems – Part I*, vol. 50, no. 1, pp. 106–120, January 2003.
- [35] G. Fan and X. -G. Xia, "Image denoising using local contextual hidden Markov model in the wavelet domain," *IEEE Signal Processing Letter*, vol. 8, no. 5, pp. 125–128, May 2001.
- [36] L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Letters*, vol. 9, no. 12, pp. 438–441, December 2002.
- [37] H. Choi, J. Romberg, R. Baraniuk, and N. Kingsbury, "Hidden Markov tree modeling of complex wavelet transforms," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, June, 2000.
- [38] M. N. Do, "Fast approximation of Kullback-Leibler distance for dependence trees and hidden Markov models," *IEEE Signal Processing Letters*, vol. 10, pp. 115–118, April 2003.
- [39] "Brodatz database," 2002, <http://www.ux.his.no/~tranden/index.html> .