WAVELET-DOMAIN COMPRESSIVE SIGNAL RECONSTRUCTION USING A HIDDEN MARKOV TREE MODEL

Marco F. Duarte,[r] Michael B. Wakin,[m] and Richard G. Baraniuk [r]

[r] Department of Electrical and Computer Engineering, Rice University
[m] Department of Electrical Engineering and Computer Science, The University of Michigan at Ann Arbor

ABSTRACT

Compressive sensing aims to recover a sparse or compressible signal from a small set of projections onto random vectors; conventional solutions involve linear programming or greedy algorithms that can be computationally expensive. Moreover, these recovery techniques are generic and assume no particular structure in the signal aside from sparsity. In this paper, we propose a new algorithm that enables fast recovery of *piecewise smooth signals*, a large and useful class of signals whose sparse wavelet expansions feature a distinct "connected tree" structure. Our algorithm fuses recent results on iterative reweighted ℓ_1-norm minimization with the wavelet Hidden Markov Tree model. The resulting optimization-based solver outperforms the standard compressive recovery algorithms as well as previously proposed wavelet-based recovery algorithms. As a bonus, the algorithm reduces the number of measurements necessary to achieve low-distortion reconstruction.

Index Terms— Compressive sensing, wavelet transforms, data compression, signal reconstruction, Hidden Markov Models.

1. INTRODUCTION

Compressive sensing (CS) is a new approach to simultaneous sensing and compression that enables a potentially large reduction in the sampling and computation costs at a sensor for signals having a sparse representation in some basis. CS builds on the work of Candès, Romberg, and Tao [1] and Donoho [2], who showed that a signal having a sparse representation in one basis can be reconstructed from a small set of projections onto a second, measurement basis that is *incoherent* with the first.[1] *Random projections* play a central role as a *universal* measurement basis in the sense that they are incoherent with any fixed basis with high probability. The CS measurement process is nonadaptive, and the reconstruction process is nonlinear. A variety of reconstruction algorithms have been proposed [1–4].

These recovery algorithms are generic in the sense that they do not exploit any particular structure in the signal besides its sparsity in some basis. However, for some signals we have additional a priori information that we should be able to exploit for improving performance. For example, the *piecewise smooth signals* [5] that feature prominently in a wide range of applications are not only sparse in the wavelet domain, but also sport wavelet coefficients that cluster

around a connected subtree in the wavelet domain [6, 7]. The goal of this paper is to design a new CS signal recovery algorithm that takes full advantage of this structure.

Previous work in this vein has developed tailored combinations of measurement/reconstruction algorithms [8]. Additionally, modified greedy algorithms have been proposed [9, 10] to exploit the parent-child wavelet coefficient relationships of piecewise smooth signals. To date, these algorithms suffer from mixed, signal-dependent performance. Furthermore, they do not exploit additional structures, such as the magnitude decay of the wavelet coefficients across scale.

In this paper, we endow a recently introduced iterative reweighted ℓ_1-norm minimization algorithm [11] with a weighting scheme based on the Hidden Markov Tree model (HMT) [6,7] to enforce the additional tree-structured dependencies exhibited by piecewise smooth signals. The combination of these two approaches provides better reconstruction performance over standard algorithms; this improvement is significant when a limited number of measurements is available. The resulting scheme consists of a initial training stage, in which the model parameters are estimated, followed by alternating iterations of a reconstruction stage based on a linear and a weight update stage based on the Viterbi algorithm.

This paper is organized as follows. Section 2 provides the necessary background on CS, and Section 3 overviews work on CS for wavelet-sparse signals. Section 4 reviews the HMT and its properties, and Section 5 introduces our new algorithm. Section 6 contains experimental results, and Section 7 concludes.

2. COMPRESSIVE SENSING BACKGROUND

Let $x \in \mathbb{R}^N$ be a signal and let the matrix $\Psi := [\psi_1, \psi_2, \ldots, \psi_N]$ be a basis for $\mathbb{R}^N$. We say that x is *K-sparse* if it can be expressed as a linear combination of K vectors from Ψ; that is, $x = \sum_{i=1}^{K} \theta_{n_i} \psi_{n_i}$ with $K \ll N$.[2] A basis that provides a sparse representation for a class of signals effectively captures the structure inherent in the class; for example, the Fourier basis sparsifies smooth signals, and wavelet bases sparsify piecewise smooth signals.

2.1. Incoherent measurements

Consider a signal x that is K-sparse in Ψ. Consider also an $M \times N$ *measurement matrix* Φ, $M \ll N$, where the rows of Φ are incoherent with the columns of Ψ. For example, let Φ contain i.i.d. Gaussian entries; with high probability, such a matrix is incoherent with any fixed Ψ (universality). Compute the *measurements* $y = \Phi x$ and note that $y \in \mathbb{R}^M$ with $M \ll N$. The CS theory states that there exists an overmeasuring factor $c > 1$ such that only $M := cK$ incoherent measurements y are required to reconstruct x with high

[1]Roughly speaking, *incoherence* means that no element of one basis has a sparse representation in terms of the other basis.

[2]Much of the CS theory extends gracefully to compressible signals, which are well-approximated by sparse signals.

probability [1, 2]. That is, just cK incoherent measurements encode all of the salient information about the K-sparse signal x.

2.2. Reconstruction from incoherent projections

The overmeasuring factor c required depends on the (nonlinear) reconstruction algorithm. Under the sparsity assumption, one can search for the sparsest signal that agrees with the obtained measurements; by using the ℓ_0-norm $\|\theta\|_0 = \#\{n : \theta_n \neq 0\}$, the reconstruction algorithm can be expressed as

$$\widehat{\theta} = \arg\min_\theta \|\theta\|_0 \quad \text{subject to} \quad \mathbf{\Phi\Psi}\theta = y.$$

While this algorithm demands the smallest possible overmeasuring factor ($c = 2$), its computational complexity renders it unfeasible. Most of the existing literature on CS [1, 2, 4] has concentrated on optimization-based methods for signal recovery, in particular ℓ_1-norm minimization. The ℓ_1-norm approach seeks a set of sparse coefficients $\widehat{\theta}$ by solving the linear program

$$\widehat{\theta} = \arg\min_\theta \|\theta\|_1 \quad \text{subject to} \quad \mathbf{\Phi\Psi}\theta = y; \quad (1)$$

the reconstruction of sparse signals via ℓ_1-norm minimization is typically exact, provided that $c = O(\log(N/K))$. Other algorithms, including greedy algorithms [3], have also been proposed for CS reconstruction and require similar oversampling factors. However, all of these algorithms are generic in the sense that, aside from sparsity, they assume no particular structure within the signal coefficients.

2.3. Iterative reweighted ℓ_1-norm minimization

When the complexity of the signal is measured using the ℓ_1-norm, individual signal coefficients are penalized according to their magnitude; in contrast, when the ℓ_0-norm is used to measure the signal complexity, the penalty for a nonzero coefficient is independent of its magnitude. The effect of this disparity is reflected in the increase of the oversampling factor between the two algorithms.

A small variation to the ℓ_1-norm penalty function has been suggested to rectify the imbalance between the ℓ_0-norm and ℓ_1-norm penalty functions [11]. The basic goal is to minimize a weighted ℓ_1-norm penalty function $\|W\theta\|_1$, where W is a diagonal "weighting" matrix with entries $W_{n,n}$ approximately proportional to $1/|\theta_n|$. This creates a penalty function that achieves higher magnitude-independence. Since the true values of θ_n are unknown (indeed they are sought), however, an iterative reweighted ℓ_1-norm minimization (IRWL1) algorithm is suggested.

The algorithm starts with the solution to the unweighted ℓ_1-norm minimization algorithm (1), which we name $\widehat{\theta}^{(0)}$. The algorithm then proceeds iteratively: on iteration $i > 0$, it solves the optimization problem

$$\widehat{\theta}^{(i)} = \arg\min_\theta \|W^{(i)}\theta\|_1 \quad \text{subject to} \quad \mathbf{\Phi\Psi}\theta = y, \quad (2)$$

where $W^{(i)}$ is a diagonal reweighting matrix with entries

$$W_{n,n}^{(i)} = \left(\left| \widehat{\theta}_n^{(i-1)} \right| + \epsilon \right)^{-1},$$

and ϵ is a small regularization constant. The algorithm can be terminated when the change between consecutive solutions is smaller than an established threshold or after a fixed number of iterations. Each iteration of this algorithm can be posed as a linear program, for which there exist efficient solvers.

3. CS FOR WAVELET-SPARSE SIGNALS

A widely used sparse representation in signal and image processing is the wavelet transform. Since piecewise polynomial signals have sparse wavelet expansions [5] and since many real-world signals describe punctuated, piecewise smooth phenomena, it follows that many real-world signals have sparse or compressible wavelet expansions. However, the significant wavelet coefficients in general do not occur in arbitrary positions. Instead they exhibit a characteristic signal-dependent structure.

Without loss of generality, we focus on 1D signals, although similar arguments apply for 2D and higher dimensional data in the wavelet or curvelet domains. In a typical 1D wavelet transform, each coefficient at scale $j \in \{1, \dots, J := \log_2(N)\}$ describes a portion of the signal of size $O(2^{-j})$. With 2^{j-1} such coefficients at each scale, a binary tree provides a natural organization for the coefficients. Each coefficient at scale $j < \log_2(N)$ has 2 *children* at scale $j + 1$, and each coefficient at scale $j > 1$ has one *parent* at scale $j - 1$.

Due to the analysis properties of wavelets, coefficient values tend to persist through scale. A large wavelet coefficient (in magnitude) generally indicates the presence of a singularity inside its support; a small wavelet coefficient generally indicates a smooth region. Thanks to the nesting of child wavelets inside their parents, edges in general manifest themselves in the wavelet domain as chains of large coefficients propagating across scales in the wavelet tree; we call this phenomenon the *persistence property*. Additionally, wavelet coefficients also have *exponentially decaying magnitudes* at finer scales [5]. This causes the significant wavelet coefficients of piecewise smooth signals to concentrate within a connected subtree of the wavelet binary tree.

This connected subtree structure was exploited by previous CS reconstruction algorithms known as Tree Matching Pursuit (TMP) and Tree Orthogonal Matching Pursuit (TOMP) [9, 10]. TMP and TOMP are modifications to the standard greedy algorithms matching pursuit and orthogonal matching pursuit, with the proviso that each selection made by the greedy algorithm build upon a connected tree. Such techniques enable faster algorithms due to the restriction in the greedy search and lower reconstruction distortion due to the inherent regularization.

However, for real-world piecewise smooth signals, the nonzero coefficients generally do not form a perfectly connected subtree. The reasons for this are twofold. First, since wavelets are band-pass functions, wavelet coefficients oscillate between positive and negative values around singularities. Second, due to the linearity of the wavelet transform, two or more singularities in the signal may cause destructive interference among coarse scale wavelet coefficients; that is, the persistence of the wavelets across scale is *weaker at coarser scales*. Either of these factors may cause the wavelet coefficient corresponding to a discontinuity to be small yet have large children, yielding a non-connected set of meaningful wavelet coefficients. TMP and TOMP used heuristic rules to ameliorate the effect of this phenomenon. However, this considerably increases the computational complexity, and the success of such heuristics varies markedly between different signals in the proposed class.

In summary, we have identified several properties of wavelet expansions:

- large/small values of wavelet coefficients generally persist across the scales of the wavelet tree;

- persistence becomes stronger as we move to finer scales; and

- the magnitude of the wavelet coefficients decreases exponentially as we move to finer scales.

We also note that the sparsity of the wavelet transform causes the coefficients to have a peaky, non-Gaussian distribution.

4. HIDDEN MARKOV TREE MODELS

The properties identified in Section 3 induce a joint statistical structure among the wavelet coefficients that is far stronger than simple sparsity or the simple connectivity models used in the TMP and TOMP algorithms. HMT [6] offers one modeling framework that succinctly and accurately captures this joint structure. HMT modeling has been used successfully to improve performance of denoising, classification, and segmentation algorithms for wavelet-sparse signals.

The HMT models the probability density function of each wavelet as a Gaussian mixture density with a hidden binary state that determines whether the coefficient is large or small. The persistence across scale is captured by a tree-based Markov model that correlates the states of parent and children coefficients. The following properties are captured by the HMT.

Non-Gaussianity: Sparse coefficients can be modeled probabilistically using a mixture of Gaussians: one component features a large variance that models large nonzero coefficients and receives a small weight (to encourage few such coefficients), while a second component features a small variance that models small and zero-valued coefficients and receives a large weight. We distinguish these two components by associating to each wavelet coefficient θ_n an unobserved hidden state $S_n \in \{S, L\}$; the value of S_n determines which of the two components of the mixture model is used to generate θ_n. Thus we have

$$
\begin{aligned}
f(\theta_n | S_n = S) &= \mathcal{N}(0, \sigma_{S,n}^2), \\
f(\theta_n | S_n = L) &= \mathcal{N}(0, \sigma_{L,n}^2),
\end{aligned}
$$

with $\sigma_{L,n}^2 > \sigma_{S,n}^2$. To generate the mixture, we apply a probability distribution to the available states: $p(S_n = S) = p_n^S$ and $p(S_n = L) = p_n^L$, with $p_n^S + p_n^L = 1$.

Persistence: The perpetuation of large and small coefficients from parent to child is well-modeled by a Markov model that links coefficient states. This induces a Markov tree where the state S_n of a coefficient θ_n is affected only by the state $S_{\mathcal{P}(n)}$ of its parent $\mathcal{P}(n)$. The Markov model is then completely determined by the set of state transition matrices for the different coefficients θ_n at wavelet scales $1 < j \leq J$:

$$
A_n = \begin{bmatrix} p_n^{S \to S} & p_n^{S \to L} \\ p_n^{L \to S} & p_n^{L \to L} \end{bmatrix}.
$$

The persistence property implies that the values of $p_n^{L \to L}$ and $p_n^{S \to S}$ are significantly larger than their complements. If we are provided the hidden state probabilities for the wavelet coefficient in the coarsest scale p_1^S and p_1^L, then the probability distribution for any hidden state can be obtain recursively:

$$
p(S_n = L) = p_{\mathcal{P}(n)}^S p_n^{S \to L} + p_{\mathcal{P}(n)}^L p_n^{L \to L}.
$$

As posed, the HMT parameters include the probabilities for the hidden state $\{p_1^S, p_1^L\}$, the state transition matrices A_n, and Gaussian distribution variances $\{\sigma_{L,n}^2, \sigma_{S,n}^2\}$ for each of the wavelet coefficients θ_n. To simplify the model, the coefficient-dependent parameters are made equal for all coefficients within a scale; that is, the new model has parameters A_j for $1 < j \leq J$ and $\{\sigma_{L,j}^2, \sigma_{S,j}^2\}$ for $1 \leq j \leq J$.

Magnitude decay: To enforce the decay of the coefficient magnitudes across scale, the variances $\sigma_{L,j}^2$ and $\sigma_{S,j}^2$ are modeled so that they decay exponentially as the scale becomes finer [7]:

$$
\begin{aligned}
\sigma_{L,j}^2 &= C_{\sigma_L} 2^{-j\alpha_L}, \\
\sigma_{S,j}^2 &= C_{\sigma_S} 2^{-j\alpha_S}.
\end{aligned}
$$

Since the wavelet coefficients that correspond to signal discontinuities decay slower than those representing smooth regions, the model sets $\alpha_S \geq \alpha_L$.

Scale-dependent persistence: To capture the weaker persistence present in the coarsest scales, the values of the state transition matrices A_j follow a model that strengthens the persistence at finer scales [7]. Additionally, the model must reflect that in general, any large parent generally implies only one large child (that which is aligned with the discontinuity). This implies that the probability that $S_n = L$, given that $S_{p(n)} = L$, should be roughly $1/2$. HMT accounts for both factors by setting

$$
p_j^{L \to L} = \frac{1}{2} + C_{LL} 2^{-\gamma_L j}, \quad p_j^{L \to S} = \frac{1}{2} - C_{LL} 2^{-\gamma_L j}
$$

$$
p_j^{S \to S} = 1 - C_{SS} 2^{-\gamma_S j}, \quad \text{and } p_j^{S \to L} = C_{SS} 2^{-\gamma_S j}.
$$

Estimation: We can obtain estimates of all parameters
$$\Theta = \{p_1^S, p_1^L, \alpha_S, \alpha_L, C_{\sigma_L}, C_{\sigma_S}, \gamma_L, \gamma_S, C_{LL}, C_{SS}\}$$
for a set of coefficients θ using maximum likelihood estimation:

$$
\Theta_{ML} = \arg \max_{\Theta} f(\theta | \Theta). \tag{3}
$$

The expectation-maximization (EM) algorithm in [6] efficiently performs this estimation. Similarly, one can obtain the state probabilities $p(S_n = S | \theta, \Theta)$ using the Viterbi algorithm; the state probabilities for a given coefficient will be dependent on the states and coefficient values of all of its predecessors in the wavelet tree.

5. HMT-BASED WEIGHTS FOR IRWL1

The IRWL1 algorithm described in Sec. 2.3 provides an opportunity to implement flexible signal penalizations while retaining the favorable computational complexity of ℓ_1-norm minimizations.

We now pose a new weight rule for the IRWL1 algorithm that integrates the HMT model to enforce the wavelet coefficient structure during CS reconstruction. Our weighting scheme, dubbed HMT+IRWL1, employs the following weighting scheme:

$$
W_{n,n}^{(i)} = \left(p \left(S_n = L | \widehat{\theta}^{(i-1)}, \Theta \right) + \delta \right)^{-q}.
$$

In words, for each wavelet coefficient in the current estimate we obtain the probability that the coefficient's hidden state is large; in the next iteration, we apply to that coefficient a weight that is inversely proportional to that probability. The parameter δ is a regularization parameter for cases where $p(S_n = L | \widehat{\theta}^{(i-1)})$ is very small, and the exponent q is a parameter that regulates the strength of the penalization for small coefficients. The goal of this weighting scheme is to penalize coefficients with large magnitudes that have low likelihood of being generated by a wavelet sparse signal; these coefficients are often the largest contributors to the reconstruction error.

The first step of HMT+IRWL1 consists of an initial training stage in which an EM algorithm solves (3) to estimate the values of the parameters for a representative signal; additionally, the solution $\widehat{\theta}^{(0)}$ for the standard formulation (1) is obtained. Subsequently, we proceed iteratively with two alternating steps: a weight update step in which the Viterbi algorithm for state probability calculations is executed for the previous solution $\widehat{\theta}^{(i-1)}$, and a reconstruction
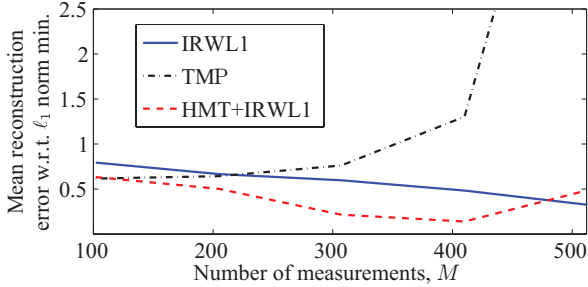
Fig. 1. *Performance of IRWL1 algorithm, normalized by the performance of ℓ_1-norm minimization. Since all values are less than 1, IRWL1 and HMT+IRWL1 consistently outperforms ℓ_1-norm minimization.*

Fig. 2. *Example outputs for the reconstruction algorithms.*

step in which the obtained weights are used in (2) to obtain an updated solution $\widehat{\theta}^{(i)}$. The convergence criterion for this algorithm is the same as for the IRWL1 algortihm.

Other probabilistic models for wavelet-sparse signals can also be used in combination with the IRWL1 algorithm, including generalized Gaussian densities [12], Gaussian scales mixtures [13], and hierarchical Dirichlet processes [14].
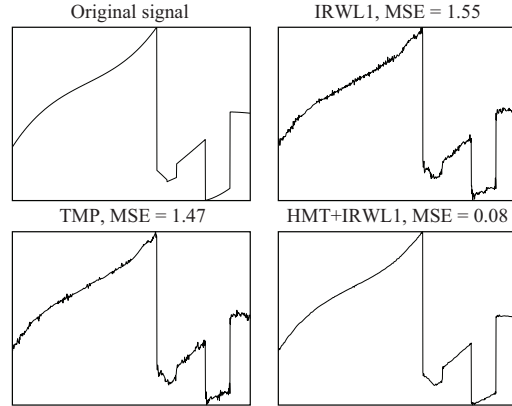
6. SIMULATIONS

We now compare the IRWL1 and HMT+IRWL1 algorithms. We use piecewise-smooth signals of length $N = 1024$, with 5 randomly placed discontinuities and cubic polynomial pieces with random coefficients. Daubechies-4 wavelets are used to sparsify the signals. Measurements are obtained using a matrix with i.i.d. Gaussian entries. For values of M ranging from 102 to 512, we test the ℓ_1-norm minimization and the IRWL1, TMP [9] and HMT+IRWL1 algorithms. We fix the number of iterations for IRWL1 and HMT+IRWL1 to 10. The parameters are set for best performance to $\epsilon = 0.2$, $q = 0.1$, and $\delta = 10^{-10}$. For each M we perform 100 simulations using different randomly generated signals and measurement matrices.

Figure 1 shows the magnitude of the reconstruction error for each of the algorithms, normalized by the error of the unweighted ℓ_1-norm minimization reconstruction, as a function of the iteration count. Figure 2 shows a reconstruction example. TMP performs well for smaller numbers of measurements M. IRWL1 consistently outperforms ℓ_1 minimization. Our proposed HMT+IRWL1 algorithm outperforms IRWL1 for most values of M. For large M near $N/2$, HMT+IRWL1 becomes less efficient than IRWL1; we speculate that at this stage the recovered signal has roughly equal numbers of large and small wavelet coefficients, which begins to violate the HMT model. Figure 2 plots the various reconstructed signals for one realization of the experiment, with $M = 300$.

7. CONCLUSIONS

This paper introduced a wavelet-tailored weighting scheme for the iterative reweighted ℓ_1-norm minimization algorithm for compressive sensing recovery. The HMT+IRWL1 scheme penalizes large wavelet coefficients that are isolated in the wavelet tree, which are unlikely to be generated by a wavelet-sparse signal. We employ the Hidden Markov Tree model to determine the probabilities of the wavelet coefficients being significant; the model provides inherent regularization by promoting several properties characteristic of sparse wavelet decompositions. For piecewise smooth signals,

the HMT+IRWL1 scheme reduces the artifacts of the ℓ_1-norm minimization reconstruction over fewer iterations than the standard iterative reweighted ℓ_1 minimization algorithm. Our scheme can be used as well in the extensions of the iterative reweighted ℓ_1-norm minimization scheme, such as reconstruction from noisy measurements.

8. REFERENCES

[1] E. J. Candès, "Compressive sampling," in *Proc. International Congress of Mathematicians*, Madrid, Spain, 2006, vol. 3, pp. 1433–1452.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, September 2006.

[3] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Info. Theory*, 2007, To appear.

[4] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, Sept. 2006.

[5] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, USA, 1999.

[6] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using Hidden Markov Models," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

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

[8] D. L. Donoho and Y. Tsaig, "Extensions of compressed sensing," *Signal Processing*, vol. 86, no. 3, pp. 549–571, March 2006.

[9] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Fast reconstruction of piecewise smooth signals from random projections," in *Proc. SPARS05*, Rennes, France, Nov. 2005.

[10] C. La and M. N. Do, "Signal reconstruction using sparse tree representation," in *Proc. Wavelets XI at SPIE Optics and Photonics*, San Diego, CA, Aug. 2005.

[11] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by weighted ℓ_1 minimization," Tech. Rep., California Institute of Technology, Pasadena, CA, Oct. 2007.

[12] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Processing*, vol. 11, no. 2, pp. 146–158, 2002.

[13] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Neural Information Processing Systems (NIPS)*, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds., Cambridge, MA, Dec. 2000, vol. 12, pp. 855–861, MIT Press.

[14] J. Kivinen, E. Sudderth, and M. Jordan, "Image denoising with nonparametric Hidden Markov Trees," in *IEEE International Conference on Image Processing (ICIP)*, San Antonio, TX, Sept. 2007.