Wavelet Statistical Models and Besov Spaces

Hyeokho Choi and Richard Baraniuk

Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA

ABSTRACT

We discover a new relationship between two seemingly different image modeling methodologies; the Besov space theory and the wavelet-domain statistical image models. Besov spaces characterize the set of real-world images through a deterministic characterization of the image smoothness, while statistical image models capture the probabilistic properties of images. By establishing a relationship between the Besov norm and the normalized likelihood function under an independent wavelet-domain generalized Gaussian model, we obtain a new interpretation of the Besov norm which provides a natural generalization of the theory for practical image processing. Based on this new interpretation of the Besov space, we propose a new image denoising algorithm based on projections onto the convex sets defined in the Besov space. After pointing out the limitations of Besov spaces, we propose possible generalizations using more accurate image models.

Keywords: Besov spaces, wavelets, generalized Gaussian distribution, signal estimation

1. INTRODUCTION

1.1. Model-based statistical image processing

Statistical modeling of signals and images has led to many successful processing algorithms in areas such as speech coding, spectral estimation, and image compression. The model approximates the joint probability density function (pdf) of the random process underlying the generation of the signals of interest, and the observed signal is considered to be a realization of the modeled random process.

A good model captures the key properties of the signals of interest; that is, the model provides *a priori* knowledge of the signal to be processed. *Bayesian* statistical inference techniques directly lead to successful signal processing algorithms for such problems as estimation, detection/classification, and compression. For example, in signal compression, the model characterizes the statistics of the source, enabling the reduction of the amount of the data down to the entropy level of the source. A bad (wrong or inaccurate) model will convey incorrect information, resulting in poor performance of the subsequent processing.

In statistical image modeling, we regard the image $\boldsymbol{X}$ as a random realization from a family or distribution of images with joint pdf $f(\boldsymbol{x})$.* If the model describes the set of images of interest correctly, the likelihood $f(\boldsymbol{x})$ will be large for an image $\boldsymbol{x}$ of interest and small for an image of less interest. Thus, the model distinguishes the images of interest, and the likelihood is a measure of this "degree of interest." The power of the model-based image processing stems from this ability of models to differentiate the images of interest and adapt the processing algorithms to match those images.

In this paper, we investigate statistical image models and the set (or space) of images of interest described by a given model and the measure of the degree of interest based on likelihood. While the models are general and can be used for a variety of statistical image processing tasks, we will concentrate on the problem of image estimation from noisy observations.

Email: choi@ece.rice.edu, richb@rice.edu. Web: www.dsp.rice.edu. This work was supported by the National Science Foundation, grant MIP-9457438, DARPA/AFOSR, grant F49620-97-1-0513, Texas Instruments, and the Rice Consortium for Computational Seismic/Signal Interpretation.

*We assume that $\boldsymbol{X}$ is a discrete sample of a 2-D function of continuous variables. We refer to the original image before sampling as the *continuous-time image*. All results also hold for 1-D signals.

1.2. Statistical models and wavelets

Ideally a model describes the complete pdf of the random process of which the given signal is a realization. However, the complete pdf is typically unavailable in practice or overly complicated, and so models are typically designed to focus on just the key properties.

Transform-domain image models are based on the idea that often a linear, invertible transform will "restructure" the image, leaving transform coefficients whose structure is "simpler" to model. Most real-world images, especially texture images, are well characterized by their *singularity* (edge and ridge) structure. The wavelet transform provides a powerful transform domain for modeling real-world images containing singularities.[1]

The wavelet transform can be interpreted as a multiscale edge detector representing the singularity content of the image at multiple scales at three different orientations – horizontal, vertical and diagonal. Each singularity in the image is represented by a cascade of large wavelet coefficients across scale.[1] If the singularity is within the support of a wavelet basis function, the corresponding wavelet coefficient is large. Hence the wavelet coefficients at the singularity location tend to be large. Likewise, a smooth image region is represented by a cascade of small wavelet coefficients across scale. Thus, the wavelet coefficients represent the edges in the image at multiple scales. The wavelet transform of an $N \times N$ discrete image can be efficiently computed using a filter bank or lifting algorithm in $O(N^2)$ computations.[2,3]

The following salient features of wavelet transforms make the wavelet-domain statistical image processing attractive:[1,2]

W1. Locality: Each wavelet coefficient represents the image content localized in spatial location and frequency.

W2. Multiresolution: The wavelet transform analyzes the image at a nested set of scales.

W3. Energy Compaction: The wavelet transforms of real-world images tend to be sparse. A wavelet coefficient is large only if singularities are present within the support of the wavelet basis.

W4. Decorrelation: The wavelet coefficients of real-world images tend to be approximately decorrelated.

The Locality and Multiresolution properties (**W1**, **W2**) enable the wavelet transform to efficiently represent the local edge content of images with large coefficients, resulting in the Compaction property (**W3**), because only a very small portion of a typical image contains edges. The Compaction and Decorrelation properties (**W3**, **W4**) simplify the statistical modeling of real-world images in the wavelet domain as compared with a direct spatial-domain modeling. Because most of the wavelet coefficients tend to be small, we need model only a small number of coefficients accurately. The Compaction (**W3**) of signal energy in wavelet domain can be restated in terms of a distinctive marginal distribution of the wavelet coefficients:

W5. NonGaussianity: The wavelet coefficients have peaky, heavy-tailed, nonGaussian marginal statistics.

The Decorrelation property (**W4**) inspires simple spatially localized modeling of the wavelet coefficients. There have been several approaches to model each wavelet coefficient independently with a nonGaussian marginal distribution, such as the generalized Gaussian distribution (GGD)[4,5] or the Gaussian mixture (GM) distribution.[6,7]

The parameters of the model can be determined by an empirical estimation based on the observed data,[5,6] enabling subsequent *empirical Bayesian* processing. For true Bayesian processing, we can specify the model parameters based on the general properties of the images. For example, Silverman et al.[7] impose a certain structure on the wavelet coefficient energy decay across scale. Their model is based on the Energy Compaction property (**W3**) and the $1/f$-type spectral behavior of the real-world images, stated as:

W6. Exponential decay across scale: The magnitudes of the wavelet coefficients of real-world images tend to decay exponentially across scale.

Although independent models are simple and easy to handle, modeling the residual dependencies between the wavelet coefficients improves the modeling of real-world images considerably.[6,8] The relation between the singularities and the behavior of the wavelet coefficients across scale indicates the following strong dependencies between the wavelet coefficients:

W7. Persistence and Clustering: Large/small values of the wavelet coefficients tend to propagate across scales.[9]

This property tells us that the wavelet transforms of the real-world images have a local dependency structure that should not be ignored. Modeling this property lead to a major breakthrough in wavelet-domain image compression.[10]

The *wavelet-domain hidden Markov tree (HMT)* model captures the persistence of the wavelet coefficient magnitudes across scale.[6] By modeling each nonGaussian marginal pdf as a 2-density Gaussian mixture with the hidden state signifying large/small magnitude of the coefficient, the persistence of the wavelet coefficient magnitudes across scale is modeled by linking the hidden states across scale in a Markov-1 tree and specifying the state transition probability matrix for each link to statistically quantify the degree of persistence of the large/small coefficients.

Although the parameters of HMT model can be trained to match a set of training data, the *universal HMT (uHMT) model* of Romberg et al.[11] enables truly Bayesian processing based on the HMT model. The uHMT specifies the model parameters *a priori* based on the general properties of the images of interest, providing a general model for a broad class of imagery.

1.3. Besov spaces vs. wavelet statistical models

We can differentiate between two fundamentally different ways of modeling the images. One constructs a set (or space) containing the images of interest along with a function quantifying how much a given image is like a real image. Besov space (with the Besov norm) is one of these *deterministic* modeling methods. The other involves constructing a *statistical* model for images. If the model characterizes the properties of typical real-world images, then the likelihood function specified by the model tells us how much a given image is like a real image.

The decay of the wavelet coefficient magnitude across scale is related to the local smoothness of the signal.[1] Thanks to the ability of wavelet transforms to represent signal smoothness, wavelets provide unconditional bases for the Besov spaces that characterize the signal smoothness deterministically.[12]

Because many images encountered in the real world consist of textured regions separated by edges, they possess particular smoothness characteristics, which are also exhibited by the exponential decay of coefficient magnitudes in the wavelet domain (**W6**). Besov spaces contain the real-world images through a deterministic characterization of the smoothness properties.

Besov spaces have been utilized mainly to assess the performance of signal estimation[13] or compression[14] algorithms. For example, signal estimation by wavelet thresholding[13] was proved optimal for signals in Besov spaces. For image compression, it was shown that the asymptotic decay of the approximation error is related to the smoothness parameters of the Besov space to which the image belongs.

Wavelet-domain statistical models have been used in various contexts of statistical signal and image processing. In particular, with the model specifying the joint pdfs, they have been used for the signal estimation under Bayesian settings,[5–7,15] signal detection/classification,[6] and image segmentation.[16]

Because both Besov spaces and wavelet-domain statistical models characterize typical real-world images, certain relationships between them exist. In Silverman et al.,[7] the realization of their statistical model was shown to belong to certain scales of Besov spaces with probability 1. This is because their model specifies an exponential energy decay across scale in the wavelet domain that is directly related to the smoothness of the signal, and thus to Besov spaces.

Although both Besov spaces and the wavelet statistical models characterize the real-world signals and images, there has been very little research on the relationship between them. Silverman et al.[7] showed that their independent wavelet-domain model is related to the Besov spaces. However, the general relationships between Besov spaces and the wavelet-domain models were never clear. In this paper, we uncover new relationship between these two seemingly different modeling frameworks that opens up generalizations of the Besov space concept both for more accurate modeling and for practical application.

As suggested in the model of Silverman et al.,[7] Besov spaces are related to *independent* statistical models of the wavelet coefficients. Because independent models ignore the key dependencies between the wavelet coefficients, Besov spaces leave much to be desired as models for real-world images. As we improve the wavelet-domain independent models by modeling dependencies, we can constrain Besov spaces more tightly to contain only the images of interest.

Besov spaces have other limitations, as well. Because Besov spaces are defined in terms of functions of continuous variables, they are not directly applicable to sampled, discrete signals and images. The original definition of Besov

spaces was concerned with the asymptotic behavior of the signal at infinitely fine scales. However, in practical image processing, the very fine scale information is lost when the image is discretized through sampling, and the asymptotic behavior is impossible to characterize. To deal with discretized images, we need to adapt the theory of Besov spaces to finite-dimensional, sampled data.

Another important point to note in Besov space theory is the interpretation of the Besov norm. The definition of Besov spaces is concerned with whether the Besov norm (of a continuous-time function) is finite or not. Although the Besov norm is actually a norm, it is not so clear what the Besov norm indicates. Some implicitly assume that the Besov norm measures the degree how much an image is like a real-world image, but this interpretation has never been made lucid.

1.4. Paper organization

In this paper, we investigate these unclear issues regarding Besov spaces and Besov norms by considering the relationship between Besov spaces and wavelet-domain statistical models. We establish a connection between the Besov norm and the likelihood function under the commonly used independent statistical model in the wavelet domain. This connection enables us to extend the concept of Besov spaces to finite, sampled data and also to endow a useful meaning to the Besov norm. Further generalization is possible by considering more powerful wavelet-domain statistical models and utilizing the same interpretation of the likelihood functions.

In Section 2, we review the theory of wavelets and Besov spaces. Section 3 considers two independent wavelet models and their connection with the Besov space. Section 4 establishes a new relationship between the Besov norm and one of the independent wavelet models, enabling us to generalize the concept of Besov space to finite, sampled data. Section 5 demonstrates an application of the generalized Besov theory to image denoising applications using projections. We discuss further generalizations of the concepts using other wavelet-domain models in Section 6. Section 7 summarizes the paper and discusses future research directions in image modeling.

2. WAVELETS AND BESOV SPACE

2.1. Wavelet transform

The discrete wavelet transform (DWT) represents a one-dimensional (1-D) signal $z(t)$ in terms of shifted versions of a lowpass scaling function $\phi(t)$ and shifted and dilated versions of a prototype bandpass wavelet function $\psi(t)$.[12] For special choices of $\phi(t)$ and $\psi(t)$, the functions $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$ and $\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k)$ with $j, k \in \mathbb{Z}$ form an orthonormal basis, and we have the representation[12]

$$z \;=\; \sum_k u_{j_0,k}\,\phi_{j_0,k} \;+\; \sum_{j=j_0}^{\infty}\sum_k w_{j,k}\,\psi_{j,k}, \tag{1}$$

with $u_{j,k} \;\equiv\; \int z(t)\,\phi^*_{j,k}(t)\,dt$ and $w_{j,k} \;\equiv\; \int z(t)\,\psi^*_{j,k}(t)\,dt$.

The *wavelet coefficient* $w_{j,k}$ measures the signal content around time $2^{-j}k$ and frequency $2^j f_0$. The scaling coefficient $u_{j,k}$ measures the local mean around time $2^{-j}k$. The DWT (1) employs scaling coefficients only at scale j_0; wavelet coefficients at scales $j > j_0$ add higher resolution details to the signal.

We can easily construct two-dimensional (2-D) wavelets from the 1-D ψ and ϕ by setting for $\mathbf{x} \equiv (x, y) \in \mathbb{R}^2$, $\psi^{\mathrm{HL}}(x,y) = \psi(x)\phi(y), \psi^{\mathrm{LH}}(x,y) = \phi(x)\psi(y), \psi^{\mathrm{HH}}(x,y) = \psi(x)\psi(y)$, and $\phi(x,y) = \phi(x)\phi(y)$. If we let $\Psi \equiv \{\psi^{\mathrm{HL}}, \psi^{\mathrm{LH}}, \psi^{\mathrm{HH}}\}$, then the set of functions $\{\psi_{j,\mathbf{k}} \equiv 2^j\psi(2^j\mathbf{x} - \mathbf{k})\}_{\psi\in\Psi, j\in\mathbf{Z}, \mathbf{k}\in\mathbf{Z}^2}$ and $\{\phi_{j,\mathbf{k}} \equiv 2^j\phi(2^j\mathbf{x} - \mathbf{k})\}_{j\in\mathbf{Z}, \mathbf{k}\in\mathbf{Z}^2}$ forms an orthonormal basis for $L_2(\mathbb{R}^2)$. That is, for every $z \in L_2(\mathbb{R}^2)$, we have

$$z = \sum_{j>j_0, \mathbf{k}\in\mathbf{Z}^2, \psi\in\Psi} w_{j,\mathbf{k},\psi}\psi_{j,\mathbf{k}} + \sum_{\mathbf{k}\in\mathbf{Z}^2} u_{j_0,\mathbf{k}}\phi_{j_0,\mathbf{k}}, \tag{2}$$

with $w_{j,\mathbf{k},\psi} \equiv \int_{\mathbb{R}^2} z(\mathbf{x})\psi_{j,\mathbf{k}}(\mathbf{x})d\mathbf{x}$ and $u_{j_0,\mathbf{k}} = \int_{\mathbb{R}^2} z(\mathbf{x})\phi_{j_0,\mathbf{k}}(\mathbf{x})d\mathbf{x}$.

For the discrete processing of images, the original image is given as the discrete samples of the continuous image $z(\mathbf{x})$. With proper prefiltering, we can approximate the discrete samples as the scaling coefficients of $z(\mathbf{x})$ at a certain

scale J, that is, the sampled image $z(\mathbf{k}) = u_{J,\mathbf{k}}$. Equivalently, we can build a continuous-time image corresponding to $z(\mathbf{k})$ as

$$\widetilde{z} = \sum_{k \in \mathbf{Z}^2} u_{J,\mathbf{k}} \phi_{J,\mathbf{k}}, \tag{3}$$

or, using the wavelet coefficients

$$\widetilde{z} = \sum_{j_0 < j < J, \mathbf{k} \in \mathbf{Z}^2, \psi \in \Psi} w_{j,\mathbf{k},\psi} \psi_{j,\mathbf{k}} + \sum_{k \in \mathbf{Z}^2} u_{j_0,\mathbf{k}} \phi_{j_0,\mathbf{k}}. \tag{4}$$

The coefficients $w_{j,\mathbf{k},\psi}$ and $u_{j_0,\mathbf{k}}$ are easily computed using the 2-D discrete-time wavelet filters and decimators operating on the samples $z(\mathbf{k})$.[12]

2.2. Besov function spaces

The theory of smoothness function spaces plays an ever more important role in signal and image processing. We shall consider the family of *Besov spaces* $B_q^\alpha(L_p(I))$ over a finite domain I, e.g., the square $[0,1]^2$, for $0 < \alpha < \infty$, $0 < p \leq \infty$, and $0 < q \leq \infty$. These spaces have, roughly speaking, "α derivatives" in $L_p(I)$; the parameter q allows us to make finer distinctions in smoothness.[14]

For $r > 0$ and $h \in \mathbb{R}^2$, define the r-th difference of a function z as

$$\Delta_h^{(r)} z(t) \equiv \sum_{k=0}^{r} \binom{r}{k} (-1)^k z(t + kh) \tag{5}$$

for $t \in I_h^r \equiv \{t \in I | t + rh \in I\}$. The $L_p(I)$-*modulus of smoothness* for $0 < p \leq \infty$ is defined as

$$\omega_r(z,t)_p \equiv \sup_{|h| \leq t} \|\Delta_h^{(r)} z\|_{L_p(I_h^r)}. \tag{6}$$

The Besov seminorm of index (α, p, q) is defined for $r > \alpha$, where $1 \leq p, q \leq \infty$, by

$$|z|_{B_q^\alpha(L_p(I))} \equiv \left\{ \int_0^\infty \left(\frac{\omega_r(z,t)_p}{t^\alpha} \right)^q \frac{dt}{t} \right\}^{1/q}, \text{ if } 0 \leq q < \infty \tag{7}$$

and by

$$|z|_{B_\infty^\alpha(L_p(I))} \equiv \sup_{0 < t < \infty} \left\{ \frac{\omega_r(z,t)_p}{t^\alpha} \right\}. \tag{8}$$

The Besov space norm is given by

$$\|z\|_{B_q^\alpha(L_p(I))} = |z|_{B_q^\alpha(L_p(I))} + \|z\|_{L_p(I)}. \tag{9}$$

The Besov space $B_q^\alpha(L_p(I))$ is then the class of functions $z : I \longrightarrow \mathbb{R}$ satisfying $z \in L_p(I)$ and $|z|_{B_q^\alpha(L_p(I))} < \infty$. Various settings of the parameters yield more familiar spaces. For example, when $p = q = 2$, $B_2^\alpha(L_2(I))$ is the Sobolev space $W^\alpha(L_2(I))$, and when $\alpha < 1$, $1 \leq p \leq \infty$, $q = \infty$, $B_\infty^\alpha(L_p(I))$ is the Lipschitz space.

Wavelets provide a simple characterization for the Besov spaces $B_q^\alpha(L_p(I))$, $0 < \alpha < \infty$, $0 < p \leq \infty$, $0 < q \leq \infty$. For analyzing ϕ and ψ possessing $r > \alpha$ vanishing moments,[1] the Besov norm $\|z\|_{B_q^\alpha(L_p(I))}$ is equivalent to the sequence norm

$$\|z\|_{B_q^\alpha(L_p(I))} \asymp |u_{j_0,\mathbf{k}}| + \left(\sum_{j \geq j_0} \left(\sum_{\mathbf{k}, \psi \in \Psi} 2^{\alpha j p} 2^{j(p-2)} |w_{j,\mathbf{k},\psi}|^p \right)^{q/p} \right)^{1/q}. \tag{10}$$

The three hyperparameters have natural interpretations: a p-norm of the wavelet coefficients is taken within each scale j, a q-norm is taken across scale, and the smoothness parameter α controls the rate of decay of the $w_{j,\mathbf{k},\psi}$ across scale (frequency). We will take (10) as the definition of the Besov norm in the following.

In signal and image processing applications, we have two particular cases of interest. First, when $p = q$, the Besov norm reduces to

$$\|z\|_{B_p^\alpha(L_p(I))} \asymp |u_{j_0,\mathbf{k}}| + \left(\sum_{j \geq j_0, \mathbf{k}, \psi \in \Psi} 2^{\alpha j p} 2^{j(p-2)} |w_{j,\mathbf{k},\psi}|^p \right)^{1/p}, \tag{11}$$

which is a weighted l_p norm of the wavelet coefficients. The second space of interest in image processing is $B_\infty^1(L_1)$, because we have $B_1^1(L_1(I)) \subset \mathrm{BV}(I) \subset B_\infty^1(L_1(I))$, and typical real-world images belong to $\mathrm{BV}(I)$. The corresponding Besov norm is given by

$$\|z\|_{B_\infty^1(L_1(I))} \asymp |u_{j_0,\mathbf{k}}| + \sup_{j \geq j_0} \sum_{\mathbf{k}, \psi \in \Psi} |w_{j,\mathbf{k},\psi}|. \tag{12}$$

3. INDEPENDENT WAVELET-DOMAIN STATISTICAL MODELS

The simplest wavelet-domain statistical models are obtained by assuming that the coefficients are independent of each other (based on the approximate decorrelation property (**W4**)). Under the independence assumption, modeling reduces to simply specifying the marginal distribution of each wavelet coefficient.

The popular Gaussian distribution is not accurate enough to model the peaky, heavy-tailed marginal distribution (**W5**) of wavelet coefficients. After Mallat et al.[4] discovered that the statistics of wavelet coefficients of the real-world images are well approximated by the *generalized Gaussian distribution* (GGD), there have been attempts to model the wavelet coefficients statistically using the GGD priors in the context of the Bayesian inference.[5,15] Another distribution commonly used to approximate nonGaussian density is the mixture of multiple Gaussian distributions. The 2-density Gaussian mixture (GM) model has been successfully used to model the statistics of wavelet coefficients (see Crouse et al.[6] and the references therein).

Because typical real-world images have a $1/f$-type spectral behavior across scales (**W6**), for an independent wavelet model of real-world images, the variance of the marginal distribution tends to decay exponentially across scale. The asymptotic decay of the wavelet coefficient magnitudes is directly related to the local signal smoothness, and thus the independent wavelet model with the exponentially decaying variance is closely related to the theory of Besov spaces. The connection between wavelet-domain statistical models and Besov spaces was first noticed by Silverman et al.[7] in their variation of the Gaussian mixture model.

Observation of the relationship between the wavelet-domain models and Besov spaces is important for resolving the practical problems of Besov space theory pointed out in Section 1.3. With the interpretation of the Besov space and Besov norm in terms of the wavelet-domain image models, the generalization of the Besov theory for practical image processing is at hand.

Before we establish the strong and useful relationship between the Besov space and the wavelet-domain statistical models through likelihood functions, we generalize the results of Silverman et al.[7] to more general cases of the independent GGD and GM wavelet-domain models.

3.1. Independent Generalized Gaussian model

The zero-mean generalized Gaussian density $\mathrm{GGD}_\nu(0, \sigma^2)$ with shape parameter ν and variance σ^2 is defined as

$$f(x) = \frac{\nu \eta(\nu)}{2\Gamma(1/\nu)} \frac{1}{\sigma} \exp\left\{ -[\eta(\nu)|x|/\sigma]^\nu \right\}, \tag{13}$$

where $\eta(\nu) = \sqrt{\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}}$. The GGD model contains the Gaussian and Laplacian distribution as special cases, using $\nu = 2$ and $\nu = 1$, respectively.

In an independent GGD wavelet model, each wavelet coefficient is generated independently according to a zero-mean GGD. For the tractability of the model, all wavelet coefficients at each scale are assumed to be independent and identically distributed (iid). We refer the model with this assumption to be *iid in scale*. That is, under the independent GGD model, we have $w_{j,\mathbf{k},\psi} \overset{\text{iid}}{\sim} \mathrm{GGD}_{\nu_j}(0, \sigma_j^2)$. Due to the iid-in-scale assumption the shape parameter ν_j and the variance σ_j^2 do not depend on the spatial location $\mathbf{k}$.

The shape parameter ν_j of the GGD function at scale j is determined to characterize the peakiness and heavy tail of the actual distribution of the wavelet coefficients. We may specify the shape parameters without reference to the given data, because most real-world images tend to have similar energy compaction in the wavelet domain.[5] For practical applications, further simplification of the model comes from assuming that the shape parameter ν_j is the same over all scales.

The variance σ_j^2 represents the signal energy at scale j. It can be empirically estimated based on the given data by estimating the variance of wavelet coefficients at scale j,[5] or it can be specified to decay exponentially (property **W6**).

The success of the independent GGD wavelet model in image denoising problems[5] demonstrates that the model is useful for real-world images. With the exponential decay of wavelet coefficients of real-world images (**W6**), the GGD model of interest takes the form:

$$w_{j,\mathbf{k},\psi} \overset{\text{iid}}{\sim} \text{GGD}_\nu(0, \sigma_j^2) \text{ with } \sigma_j = 2^{-j\beta}\sigma_0. \tag{14}$$

Because the decay of wavelet coefficients across scale determines the smoothness of the underlying signal, the realizations of the independent GGD model in (14) with exponentially decaying variance can be shown to belong to certain scales of Besov space. More specifically, we have the following theorem:[†]

THEOREM 3.1. *Suppose each wavelet coefficient $w_{j,\mathbf{k},\psi}$ is distributed as in (14) with $\beta > 0$ and $\sigma_0 > 0$. Then, for $0 < p, q < \infty$, the realizations of the independent GGD model are almost surely in $B_q^\alpha(L_p(I))$ if and only if $\beta > \alpha + 1$.*

3.2. Independent Gaussian mixture model

A zero-mean GM distribution with M mixture densities is defined as $f(x) = \sum_{m=1}^M p_m f_m(x)$, where $f_m(x)$ is a zero-mean Gaussian density with variance σ_m^2, and the p_m's are mixture probabilities such that $\sum_{m=1}^M p_m = 1$. While we can increase the number of mixture densities M to approximate the distribution of each wavelet coefficient with arbitrary fidelity, even with only 2 mixture densities, a GM models the key properties of the wavelet transform.[7,6] For independent GM models, we assume the models are iid in scale.

With the iid-in-scale assumption, a GM model with 2 mixture densities is specified by 4 parameters at each scale, the variances $\sigma_{j,m}^2$ and the mixture probabilities $p_{j,m}$, for $m = 1, 2$. The variances characterize the decay of the density function as we move away from zero, and the mixture probabilities represent the degree of energy compaction in the wavelet domain.

To model the exponential decay of the wavelet coefficient magnitudes across scale, we assume that $\sigma_{j,m} = \sigma_{0,m} 2^{-j\beta_m}$ for some $\beta_m > 0$. Furthermore, we assume that the mixture probabilities converge at fine scales, that is $p_{j,m} \longrightarrow p_m$, as $j \to \infty$. Then, we can prove the following relationship with the Besov space:[‡]

THEOREM 3.2. *Suppose we have an independent Gaussian mixture wavelet model such that each $w_{j,\mathbf{k},\psi}$ is distributed according to the GM density $\sum_{m=1}^M p_{j,m} f_{j,m}(w_{j,\mathbf{k},\psi})$ with $\sigma_{j,m} = \sigma_{0,m} 2^{-j\beta_m}$ for $\beta_m > 0$ and $p_{j,m} \to p_m$ exponentially as $j \to \infty$. Suppose that the mixture densities are ordered such that β_m does not decrease as m increases and that $p_1 > 0$. Then, the realizations of the model are almost surely in $B_q^\alpha(L_p)$ for $p, q < \infty$ if and only if $\beta_1 > \alpha + 1$.*

3.3. Besov space as a model of real-world images

In view of Theorems 3.1 and 3.2, certain Besov spaces contain the images that are modeled by the independent GGD and GM models. These results agree with the fact that typical real-world images possess smoothness properties, which are captured by the Besov spaces.

However, we should not be misled by the above two theorems to conclude that all functions in a Besov space look like real-world images. The theorems merely tell us that the images generated by the above wavelet-domain models are contained in the Besov spaces almost surely. While most real-world images may belong to a Besov space with very high probability, there exist other elements in the same space that are far from resembling real-world images. In words, Besov spaces contain many signals of less interest to us.

[†]The proof of the theorem parallels the proof of Theorem 1 in Silverman et al.[7]

[‡]Proof is similar to that of Theorem 3.1.

Another point to note is that, as illustrated by Theorems 3.1 and 3.2, a Besov space is closely related to *independent* wavelet-domain models. Because the residual key properties of wavelet coefficients are missed by these independent models (e.g., the persistence property **W7**), the categorization of images using Besov spaces ignores these unmodeled properties. We will revisit these points in Section 5.

4. STATISTICAL INTERPRETATION OF BESOV SPACE

Although typical real-world images belong to the Besov space with proper smoothness parameters, we have several problems when we apply the theory to the practical image processing.

The first problem is a proper interpretation of Besov norm. The definition of Besov space is only related to whether the Besov norm is finite. If the Besov norm is finite, then the signal is in the corresponding Besov space. However, it is not clear how we can compare the (smoothness) properties of the signal with other signals in the space. Intuitively, the size of the Besov norm is related to the smoothness of the signal, meaning small Besov norms correspond to smooth signals and vice versa. However, it is not easy to interpret the Besov norm as an indicator of how much a given image "looks like" a real-world image.

The second problem, as mentioned in Section 1.3, comes from the fact that Besov spaces are spaces of continuous-time functions. However, in practice, only a discrete sampled version of the underlying continuous-time image is available. Given a sampled image, the natural question would be whether the given image belongs to a certain Besov space. Because the Besov space contains continuous-time images, the direct answer to this question would come from seeing if the original continuous-time image (before sampling) belongs to that space. However, because we lose the fine-scale information when we sample the image and the Besov spaces are concerned with the asymptotic behavior of the image at infinitely fine scales, it is not straightforward to check the smoothness of the original image.

Other related problems involve computing the Besov norm from the sampled data and its interpretation. Having the wavelet coefficients available only down to a certain fine scale, the Besov norm completely depends on what we assume for the unknown coefficients. An obvious way of computing the Besov norm of a sampled image would be assuming all the unknown coefficients to be zero, yielding the *truncated Besov norm* obtained by truncating the summation in (11) at a certain scale. However, it is not justified to define this truncated Besov norm as the Besov norm of the original sampled data.

To overcome these difficulties, we establish a new interpretation of the Besov norm in terms of the likelihood function under a statistical image model. We show that the likelihood with proper normalization under the independent GGD model is equivalent to the Besov norm for certain Besov spaces. This naturally endows a meaning to the magnitude of the Besov norm as well as provides a natural framework for developing a Besov-like theory for finite sampled images.

We first define the normalized likelihood function under the wavelet-domain statistical models. Then, we show that the normalized likelihood is equivalent to the Besov norm for $B_p^\alpha(L_p)$. In Section 4.3, a generalization of the Besov space to finite sampled data is made.

4.1. Normalized likelihood function

Given a pdf $f(x|\mathbf{\Theta})$ of the random variable $\mathbf{x}$ under the model $\mathbf{\Theta}$ and the realizations x_1 and x_2, we can compute the likelihoods of the data $f(x_1|\mathbf{\Theta})$ and $f(x_2|\mathbf{\Theta})$. Because the likelihoods indicate how "likely" the given data are under the given pdf, $f(x_1|\mathbf{\Theta}) > f(x_2|\mathbf{\Theta})$ implies x_1 is more likely than x_2 under the model.

However, the likelihood $f(x_1|\mathbf{\Theta})$ is meaningful only when it is compared with another likelihood such as $f(x_2|\mathbf{\Theta})$, and we cannot tell how likely x_1 is merely based on the value $f(x_1|\mathbf{\Theta})$. To be able to tell how likely an observation is under a given pdf model, we must normalize the likelihood appropriately. A natural way of accomplishing the normalization is to compare the likelihood with the maximum likelihood achievable. We define the *normalized likelihood function* $f^N(x|\mathbf{\Theta})$ by

$$f^N(x|\mathbf{\Theta}) = \frac{f(x|\mathbf{\Theta})}{\sup_x f(x|\mathbf{\Theta})}, \tag{15}$$

with the assumption that $0 < \sup_x f(x|\mathbf{\Theta}) < \infty$. Then, $f^N(x|\mathbf{\Theta}) \in [0,1]$ and we can tell that an observation x is "likely" if $f^N(x|\mathbf{\Theta})$ is close to 1 and "not likely" if it is close to 0. We can easily generalize the concept of the normalized likelihood to finite random vectors using their joint pdfs.

For discrete random processes, the joint pdf of an infinite random vector is not defined, and we cannot define the normalized likelihood as in (15). However, we can generalize the normalized likelihood using the limit of the normalized likelihood when the limit exists. Let $X_1, X_2, \ldots$ be an infinite sequence of random variables. Then for the vector of the first n random variables $\boldsymbol{x}_n = \{X_k\}_{k=1}^n$, we define the normalized likelihood $f_n^N(\boldsymbol{x}_n)$ using (15). Then, if $\lim_{n\to\infty} f_n^N$ exists, we define the normalized log likelihood of the infinite sequence to be the limit.

For the independent wavelet-domain models considered in Section 3, we can take the limit as we move to the fine scales, defining the normalized likelihood

$$f^N(\boldsymbol{w}) = \lim_{J\to\infty} \frac{\prod_{j=0}^J \prod_{\mathbf{k},\psi} f_{j,\mathbf{k},\psi}(w_{j,\mathbf{k},\psi})}{\sup \prod_{j=0}^J \prod_{\mathbf{k},\psi} f_{j,\mathbf{k},\psi}(w_{j,\mathbf{k},\psi})}, \tag{16}$$

when $w_{j,\mathbf{k},\psi} \sim f_{j,\mathbf{k},\psi}(w_{j,\mathbf{k},\psi})$. For the independent GGD and GM models with exponentially decaying variance as defined in Section 3, the normalized likelihood is well defined because the supremum is finite and the limit exists.

4.2. Besov spaces and the GGD model

Consider an independent GGD wavelet model with exponentially decaying variance across scale:

$$\boldsymbol{\Theta}_p^\alpha \ : \ w_{j,\mathbf{k},\psi} \overset{\text{iid}}{\sim} \text{GGD}_\nu(0, \sigma_j^2) \text{ with } \sigma_j = 2^{-j(\alpha+2-2/p)}\sigma_0. \tag{17}$$

The normalized likelihood computed using the coefficients between scales 0 and J is

$$\begin{aligned} f_J^N(\text{w}) &= \prod_{j=0}^J \prod_{\mathbf{k},\psi} \exp\{-[\eta(\nu)|w_{j,\mathbf{k},\psi}|/\sigma_j]^\nu\} \\ &= \exp\left\{ \sum_{0\le j\le J,\mathbf{k},\psi} -[\eta(\nu)|w_{j,\mathbf{k},\psi}|/\sigma_j]^\nu \right\}. \end{aligned} \tag{18}$$

Computing the negative log of the normalized likelihood function, we obtain the negative log normalized likelihood function:

$$\begin{aligned} -\log f^N(\boldsymbol{w}) &= \sum_{0\le j,\mathbf{k},\psi} \eta(\nu)|w_{j,\mathbf{k},\psi}|/\sigma_j]^\nu \\ &= \eta(\nu) \sum_{0\le j,\mathbf{k},\psi} (2^{j(\alpha+2-2/p)}|w_{j,\mathbf{k},\psi}|)^\nu, \end{aligned} \tag{19}$$

which is equivalent to $\|z\|^p_{B^\alpha_p(L_p)}$, the homogeneous Besov norm of the function z. When $p = 2$, we obtain the iid Gaussian model for wavelet coefficients, and the corresponding normalized negative log likelihood is equivalent to the Sobolev norm of the function.

In terms of the normalized likelihood function for the iid GGD model $\boldsymbol{\Theta}_p^\alpha$, the Besov space $B_p^\alpha(L_p)$ can be equivalently defined as the set

$$\{\boldsymbol{w} : f^N(\boldsymbol{w}|\boldsymbol{\Theta}_p^\alpha) \neq 0\}. \tag{20}$$

Thus, the signals in the Besov space $B_p^\alpha(L_p)$ are the "likely" signals under the statistical model $\boldsymbol{\Theta}_p^\alpha$.

The characterization of Besov space in terms of the likelihood function can be generalized to other scales of Besov space that are of interest in image processing, such as $B_\infty^1(L_1)$. For $B_\infty^1(L_1)$, we are interested in the scale-by-scale normalized likelihood under the GGD model with $\nu = 1$ and $\beta = 0$. It is easily shown that

$$B_\infty^1(L_1) = \{\boldsymbol{w} \ : \ \inf_j g_j^N(\boldsymbol{w}) > 0\}, \tag{21}$$

where $g_j^N(\boldsymbol{w}) = \prod_{\mathbf{k},\psi} \exp(|w_{j,\mathbf{k},\psi}|/\sigma_0)$ is the normalized log likelihood of the wavelet coefficients at scale j. That is, rather than the joint likelihood of all the wavelet coefficients, $B_\infty^1(L_1)$ is concerned with the scale of least likelihood.

Because the Besov norm is the negative log likelihood, reducing the Besov norm corresponds to increasing the likelihood. In view of this relationship, the variational problem of Chambolle et al.[17] is an estimator regularized by the likelihood. The interpolation algorithm proposed by Choi et al.[18] is a maximum likelihood interpolator under the wavelet-domain GGD model.

4.3. Besov balls for sampled data

In practice, the available data are given as finite samples of the image, and we have no access to the wavelet coefficients beyond a certain finest scale. Suppose we have wavelet coefficients up to scale J and the coefficients with scale $j > J$ are unknown. Let $\boldsymbol{w}_J$ be the vector of available wavelet coefficients. Because the definition of Besov spaces is concerned with the asymptotic decay of the signal energy across scale, it is not clear how the Besov space theory can be applied to practical signal/image processing. Through the generalization of the theory for finite data based on the statistical connection in (20), we can modify the Besov space definition for finite data $\boldsymbol{w}_J$.

The set of "likely" signals can be modified as

$$\{\boldsymbol{w}_J : f_J^N(\boldsymbol{w}_J|\boldsymbol{\Theta}_p^\alpha) \geq \epsilon\}, \tag{22}$$

where $f_J^N(\boldsymbol{w}_J|\boldsymbol{\Theta}_p^\alpha)$ is the normalized pdf of the (finite) vector $\boldsymbol{w}_J$, and ϵ is a positive constant controlling the "likelihood" of signals. We can choose ϵ so that the set contains the signals of interest. This is the generalization of the Besov space for finite data. This generalization of the Besov space for finite data is meaningful in its own right, and we can extend the theory of Besov spaces to the problems with finite data.

Under the independent GGD model, the negative log likelihood function is the truncated form of the Besov norm defined in (11), and the set of signals is a "ball" in Besov space defined as $\{x : \|x\|_B \leq R\}$, where R is the radius of the ball and $\|\cdot\|_B$ is the truncated Besov norm for finite samples. Although we considered Besov spaces $B_p^\alpha(L_p)$ above, the Besov balls can be defined for other scales of Besov spaces in the same way. In particular, the space $B_\infty^1(L_1)$ is desirable for modeling images.[17] If $p, q \geq 1$, the Besov balls are convex sets in L_p.

5. SIGNAL ESTIMATION USING PROJECTIONS ONTO BESOV BALLS

The notion of a Besov ball for finite sampled data leads us to an intuitively appealing signal denoising algorithm. Given a noisy signal, we can project it onto a Besov ball defined as above to obtain an estimate of the original signal that is close to the give data and is "likely" in terms of the underlying wavelet statistical model. Given an image z, we model the noisy observation as $z + n$, where n is an additive Gaussian white noise with variance σ^2 corrupting the observation. The estimation problem is to restore the original image z from the noisy observation $z + n$.

Suppose that we have M different wavelet bases, $W_1, \ldots, W_M$, and that the number of vanishing moments of each wavelet basis is greater than the Besov smoothness parameter α of the Besov space under consideration. For each wavelet basis W_i, we define the Besov $B_i(r_i)$ with radius r_i as $B_i(r_i) = \{z : \|x\|_B \leq r_i\}$ where the norm is the Besov norm defined using wavelet basis W_i. With properly chosen r_i, the original signal z is in $B_i(r_i)$ for all i, and hence we have $z \in \cap_i B_i(r_i)$. To obtain an estimate of z, we project the noisy signal $x + n$ onto $\cap_i B_i(r_i)$ in the l_2 sense. We can find the projection using the method of projection onto convex sets (POCS)[19] starting at $z + n$. For an $N \times N$ image, each Besov ball $B_i(r_i)$ is a convex set in $\mathbb{R}^{N^2}$.

As we change the radius r_i, the size of the Besov ball changes. In $\mathbb{R}^{N^2}$, the change of basis from a orthonormal wavelet domain to another orthonormal wavelet domain corresponds to a rotation of the coordinate axes, and hence each Besov ball $B_i(r_i)$ is a rotation of other Besov balls defined using other wavelet transforms. Multiple wavelet bases define Besov balls with different rotations and possibly with different sizes, but all contain the original signal z. As we increase the number of wavelet domains, the size of the intersection $\cap_i B_i(r_i)$ never increases, and it defines a smaller set of feasible signals. By projecting the noisy signal onto this small convex set, we can improvement wavelet denoising over single basis algorithms.

The proper choice of the radius of each Besov ball is essential for the success of the algorithm. This is analogous to choosing the threshold in wavelet thresholding. We should choose r_i so that the Besov ball $B_i(r_i)$ contains the true noise-free signal z. At the same time, r_i should be chosen as small as possible to make the ball small for accurate denoising. Ideally, r_i should be chosen as $r_i = \|z\|_B$, where the norm is the truncated Besov norm defined using basis W_i. However, because z is unknown, we must estimate $\|z\|_B$. A method of estimating the Besov norm of z is to compute the Besov norm of an estimate of z obtained using some other method. One natural choice is the Besov norm of the denoised signal using Donoho and Johnstone's universal thresholding.

However, because the signal is over-smoothed by this universal thresholding, we need to choose somewhat larger r_i than the Besov norm of the thresholded signal. For real images, we found out that 1.5 to 2 times the Besov norm of the SureShrink signal estimate provides a good estimate of the actual Besov norm.

Figure 1. Image denoising examples using projections onto Besov Balls; (a) 512×512 original image with 256 gray levels, (b) Noisy image ($\sigma = 35$, PSNR = 17.2dB), (c) MATLAB's spatially adaptive Wiener filter (3×3 window, PSNR = 27.0dB), (d) Hard thresholding in D_{12} domain using "best" threshold (PSNR = 26.1dB); the threshold was chosen to maximize PSNR, (e) Besov ball projection in $B^1_\infty(L_1)$ using D_{12} wavelet (PSNR = 27.2dB), and (f) Besov ball projection in $B^1_\infty(L_1)$ using 5 wavelet bases, D_6, D_8, D_{10}, D_{12} and D_{14} (PSNR = 28.3dB).

The choice of underlying Besov spaces $B^\alpha_q(L_p)$ depends on the property of the signal to be denoised. For real world images, we found that $B^1_\infty(L_1)$ yields good performance as well as simple numerical implementation.

Figure 1 compares the PSNR ($10 \log_{10}(\frac{255^2}{\text{MSE}})$) of various denoising methods for a real image. As the wavelet bases, we used Daubechies orthonormal wavelet filters. Here, D_n refers to the Daubechies length n filter. We observe that the proposed denoising algorithm outperforms other existing methods both visually and in MSE. In the denoising simulations involving many different images and different noise variances, the Besov ball algorithm demonstrated excellent performance, especially when the SNR of the noisy observation was low.

6. TOWARD GENERALIZATIONS OF BESOV SPACE

6.1. Limitations of Besov space

Because the Besov norm is closely related to zero-mean wavelet-domain statistical models that are iid in scale, it has several limitations.

As is easily observed from the definition of the Besov norm in (11), the Besov norm is invariant to coefficient *shuffling* (permuting) within scale. Because the shuffling of wavelet coefficients destroys the edge structure of the image, the image resulting after the coefficient shuffling does not resemble a real-world image (see Fig. 2(b)). However, the wavelet-shuffled image in Fig. 2(b) belongs to the same Besov space (with identical Besov norm) as the original image in Fig. 2(a). This implies that not all elements of the Besov space look like real-world images. Shuffle invariance is related to the relationship of the Besov space with the *independent* GGD model. Because independent models ignore the persistence of wavelet coefficients (**W7**) related to the edges in the image, the Besov space suffers from the same problem.

Another problem of the Besov norm is that it is invariant to *sign changes* of the wavelet coefficients. Because the wavelet basis is an unconditional basis for Besov spaces, it is invariant to sign changes. As observed in Fig. 2(c), the

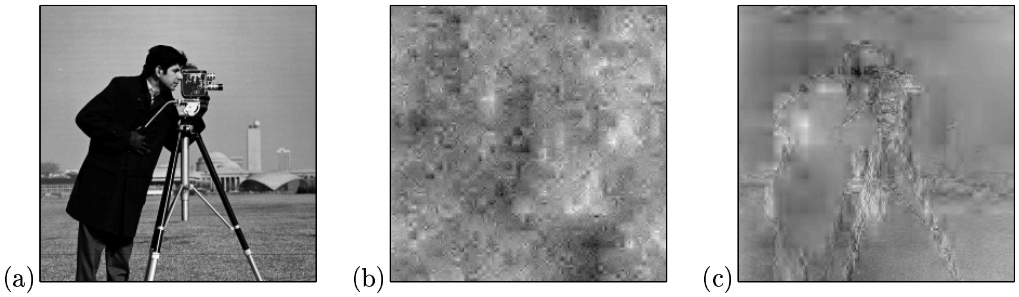

Figure 2. Limitations of Besov space and norm. (a) Image and inverse wavelet transforms of its wavelet coefficients after (b) random shuffles (permutations) and (c) random sign changes within each scale band. All three images have identical Besov norm, indicating that the Besov norm is blind to much image structure.

image resulting from random sign flips of wavelet coefficients does not look like a real image at all, while it belongs to the same Besov space as the original image with identical Besov norm. This invariance to the sign change is related to the relationship of the Besov space with the *zero-mean* statistical model in the wavelet domain, because a zero-mean distribution symmetric around zero is insensitive to sign changes.

To overcome these limitations of the Besov space as a characterization of the real-world images, we need to construct new spaces or sets based on more accurate image models.

6.2. Possible generalizations

The normalized likelihood function in (20) and the space or set defined in (20) and (22) are very general, and they can be applied to other statistical image models without modification. By using more accurate models than the independent models, the correspondingly defined space or set would be more tightly constrain the real-world images.

One possible generalization of Besov space is using the wavelet-domain HMT model[6] that captures the persistence property of the wavelet coefficients for the real-world images. The universal HMT (uHMT)[11] is a general image model that captures the interscale dependencies of the wavelet coefficients. The normalized likelihood function for the uHMT model is shuffle variant because it captures edge information. However, the set of nonzero normalized likelihood defined using the wavelet-domain HMT model using (20) is no longer a linear space, and the normalized likelihood function does not correspond to a norm. Further complication results for practical applications, because the normalized likelihood function has many local maxima.

The sign invariance problem of the Besov norm can be cured by developing a model that captures the sign of wavelet coefficients. For this purpose, the wavelet-domain HMT model with more than 2 mixture densities looks promising. For example, we can use 3-density GM distribution with the 3 Gaussian densities modeling small, large positive, and large negative wavelet coefficients, respectively. The correlations of the wavelet coefficient signs are modeled through the hidden state transition probabilities between the 3 densities across scale.

For more accurate modeling of the images, the model should capture the edge structure or the geometry information contained in images. In general, as we make the model more accurate to incorporate more detailed properties of real-world images, the set of real-world images constructed based on the model will become more complicated and harder to manipulate.

7. CONCLUSIONS

In this paper, we have developed relations between Besov spaces and wavelet-domain statistical models. We have shown that the Besov norm is equivalent to the normalized log likelihood function under an independent GGD model in the wavelet domain. This relationship enabled us to generalize the Besov space theory to the finite sampled data which is applicable in practical image processing. Based on the generalization of the Besov theory to discrete sampled data, we defined the Besov ball containing the real-world images in the sampled, discrete domain. The Besov ball concept inspired a new image denoising algorithm that uses multiple wavelet bases. This is in contrast to the shift invariant wavelet denoising that uses a tight frame consisting of different shifts of a single wavelet basis.

Besov spaces have several limitations as a set of real-world images, because they contain many non-realistic images. The relationship between Besov space and the wavelet-domain statistical model developed in this paper can be generalized to more general image models. The development of other general models and the corresponding theory of Besov-like sets remain future research topics.

REFERENCES

1. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1998.
2. M. Vetterli and J. Kovačević, *Wavlets and Subband Coding*, Prentice Hall, Englewood Cliffs:NJ, 1995.
3. W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *J. Appl. Comp. Harm. Anal.* **3**(2), pp. 186–200, 1996.
4. S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, pp. 674–693, July 1989.
5. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian priors," in *Proc. IEEE TFTS Symposium*, pp. 633–636, (Pittsburgh, PA), Oct. 6-9 1998.
6. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.* **46**, pp. 886–902, April 1998.
7. F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *J. Roy Statist. Soc. Ser. B* **60**, pp. 725–749, 1998.
8. E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *31st Asilomar Conf on Signals, Systems and Computers*, (Pacific Grove, CA), November 1997.
9. S. Mallat and W. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. on Info. Theory* **38**(2), pp. 617–643, 1992.
10. J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Proc.* **41**, pp. 3445–3462, Dec. 1993.
11. J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," in *Proceedings of SPIE*, (Denver, CO), July 1999.
12. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, New York, 1992.
13. D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.* **90**, pp. 1200–1224, Dec. 1995.
14. R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. on Information Theory* **38**, pp. 719–746, March 1992.
15. E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *IEEE Int. Conf. on Image Proc. — ICIP 1996*, (Lausanne, Switzerland), Sept. 1996.
16. H. Choi and R. G. Baraniuk, "Image segmentation using wavelet-domain classification," in *Proceedings of SPIE*, (Denver, CO), July 1999.
17. A. Chambolle, R. A. DeVore, N. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. on Image Proc.* **7**, pp. 319–355, July 1998.
18. H. Choi and R. Baraniuk, "Interpolation and denoising of nonuniformly sampled data using wavelet domain processing," in *Proc. of ICASSP 99*, (Phoenix, AZ), March 1999.
19. D. C. Youla and H. Webb, "Image restoration by method of convex projections: Part 1-theory," *IEEE Trans. Medical Imaging* **MI-1**, pp. 81–94, October 1982.