Wavelet Statistical Models and Besov Spaces

Hyeokho Choi and Richard G. Baraniuk¹

1.1 Introduction

1.1.1 Natural images models

Natural image models provide a foundation for framing numerous problems encountered in image processing, from compressing images to detecting tumors in medical scans. A good model must capture the key properties of the images of interest. A typical photograph of a natural scene consists of piecewise smooth or textured regions separated by step edge discontinuities along contours. Modeling both the smoothness and edge structure is essential for maximum processing performance.

Research to date in image modeling has been split into two fairly distinct camps, with one group focusing on *deterministic* models and the other pursuing *statistical* approaches. Deterministic models define a set or vector space that contains the images of interest and a metric or norm typically based on smoothness. Statistical models place a probability measure on images, making natural images more likely than unnatural ones.

1.1.2 Deterministic image modeling

Working under the assumption that natural images consist of smooth or textured regions delineated by step edges, we can assess images in terms of the number of derivatives we can compute. The *Sobolev space*² $W^\alpha(L_2)$ contains all images having α derivatives of finite energy. However, since

¹Richard G. Baraniuk is Professor, Electrical and Computer Engineering, Rice University. Hyeokho Choi is Research Professor/Lecturer, Electrical and Computer Engineering, Rice University. This work was supported by NSF, ONR, AFOSR, DARPA, and Texas Instruments. The authors thank the anonymous reviewer for many useful comments.

²Sobolev and Besov spaces will be formally defined in Section 1.3.

natural images contain many step edges, they do not in general belong to any Sobolev space.

To overcome this difficulty, researchers have turned to *Besov spaces*, which can be tuned to measure the smoothness “between” the step edges. Roughly speaking, the Besov space $B_q^\alpha(L_p)$ contains signals having α derivatives between edges.

Besov norms are naturally computed in the wavelet domain, since the wavelet transform computes derivatives of the image at multiple scales and orientations. Besov space concepts have been applied to assess the performance of image estimation [1] and compression [2] algorithms. Indeed, estimation by wavelet thresholding [1] has been proven near optimal for representing and removing noise from Besov space images.

1.1.3 Statistical image modeling

Statistical models characterize natural images by assigning a probability measure to the set of images. Each image is considered as a random realization from the resulting “natural image random process.” The smoothly varying spatial properties of images have motivated considerable research on Markov random field (MRF) models to characterize the statistical dependencies between image pixels. While conceptually appealing, MRFs have severe computational limitations.

As an alternative, statistical models that capture the statistics of the wavelet coefficients have been deployed in applications such as Bayesian estimation [3, 4, 5, 6], detection/classification [4], and segmentation [7]. We consider two wavelet-domain statistical models here: one that models the wavelet coefficients as independent generalized Gaussian random variables, and one that models the wavelet coefficients as an “infinite mixture” of Gaussian random variables correlated in variance.

1.1.4 Links between two paradigms

To date, the deterministic Besov space and wavelet statistical model frameworks have been developed by essentially two distinct communities, and few connections have been made between the two approaches. We summarize the state of affairs in Table 1.1. Only recently, Abramovich et al. showed that their independent wavelet-domain model is related to Besov space [5] (see also [8]). However, the general relationships between Besov spaces and wavelet-domain statistical models have never been clear.

In this chapter, we uncover several surprising relationships between these two seemingly different modeling frameworks that open up generalizations of the Besov space concept both for more accurate modeling and for practical applications. We take an information-theoretic view of the two frameworks and show that a Besov space deterministic model is in a sense equivalent to an independent generalized Gaussian statistical model.

Table 1.1. Comparison of deterministic vs. statistical image models.

deterministic model	statistical model
function space	probability density function
norm $\ \mathbf{x}\ $	likelihood $f(\mathbf{x})$

This connection allows us to see for the first time the equivalence between many processing algorithms proposed in one or the other framework. We also point out some of the shortcomings of Besov spaces as natural image models and use the statistical framework to suggest new, more realistic image models that generalize Besov spaces.

This chapter is organized as follows. In Section 1.2, we review the basic theory of the wavelet transform and the properties of natural images in the wavelet domain. Section 1.3 introduces the Besov space as a deterministic image model, and Section 1.4 presents two simple wavelet statistical image models. Section 1.5 then forges the links between Besov space and the statistical models. Finally, we propose some extensions to the Besov model inspired by advanced statistical models and conclude in Section 1.6.

1.2 Wavelet Transforms and Natural Images

1.2.1 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 $\psi(t)$ [9]. 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 [9]

$$z = \sum_k u_{j_0,k} \phi_{j_0,k} + \sum_{j=j_0}^{\infty} \sum_k w_{j,k} \psi_{j,k} \quad (1.1)$$

with $u_{j,k} := \int z(t) \phi_{j,k}(t) dt$ and $w_{j,k} := \int z(t) \psi_{j,k}(t) dt$.

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 (1.1) employs scaling coefficients only at scale j_0 ; wavelet coefficients at scales $j \geq j_0$ add higher resolution details to the signal.

We can construct 2-D wavelets from the 1-D ψ and ϕ by setting for $\mathbf{x} := (x, y) \in \mathbb{R}^2$, $\psi^{\text{HL}}(x, y) := \psi(x)\phi(y)$, $\psi^{\text{LH}}(x, y) := \phi(x)\psi(y)$, $\psi^{\text{HH}}(x, y) := \psi(x)\psi(y)$, and $\phi(x, y) := \phi(x)\phi(y)$. If we let $\Psi := \{\psi^{\text{HL}}, \psi^{\text{LH}}, \psi^{\text{HH}}\}$,

then the set of functions $\{\psi_{j,k} = 2^j \psi(2^j \mathbf{x} - \mathbf{k})\}_{\psi \in \Psi, j \in \mathbb{Z}, \mathbf{k} \in \mathbb{Z}^2}$ and $\{\phi_{j,k} = 2^j \phi(2^j \mathbf{x} - \mathbf{k})\}_{j \in \mathbb{Z}, \mathbf{k} \in \mathbb{Z}^2}$ forms an orthonormal basis for $L_2(\mathbb{R}^2)$. That is, for every $z \in L_2(\mathbb{R}^2)$, we have

$$z = \sum_{\mathbf{k} \in \mathbb{Z}^2} u_{j_0, \mathbf{k}} \phi_{j_0, \mathbf{k}} + \sum_{j > j_0, \mathbf{k} \in \mathbb{Z}^2, \psi \in \Psi} w_{j, \mathbf{k}, \psi} \psi_{j, \mathbf{k}} \quad (1.2)$$

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

Given discrete samples of the continuous image $z(\mathbf{x})$ and proper prefiltering, we can approximate the discrete samples by 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

$$\tilde{z} = \sum_{\mathbf{k} \in \mathbb{Z}^2} u_{J, \mathbf{k}} \phi_{J, \mathbf{k}} \quad (1.3)$$

or, using the wavelet coefficients,

$$\tilde{z} = \sum_{\mathbf{k} \in \mathbb{Z}^2} u_{j_0, \mathbf{k}} \phi_{j_0, \mathbf{k}} + \sum_{j_0 < j < J, \mathbf{k} \in \mathbb{Z}^2, \psi \in \Psi} w_{j, \mathbf{k}, \psi} \psi_{j, \mathbf{k}}. \quad (1.4)$$

The coefficients $u_{j_0, \mathbf{k}}$ and $w_{j, \mathbf{k}, \psi}$ are closely approximated using 2-D discrete-time wavelet filters and decimators operating on the samples $z(\mathbf{k})$ [9].

1.2.2 Wavelet-domain properties of natural images

Completely specifying the form of the set of all natural images is an impossible task. To get an idea of the difficulty, note that an $N \times N$ natural image z is a vector in the space $\mathbb{R}^{N^2}$, an extremely large space for any reasonable N .³ Two things are clear. First, the set of natural images is extremely small, since it is very unlikely that an arbitrary vector z' picked from $\mathbb{R}^{N^2}$ will resemble a natural image. Second, the tiny set of natural images is extremely complicated. It has been conjectured that natural images lie near a nonlinear manifold in $\mathbb{R}^{N^2}$ [11, 12]. If this is true, modeling natural images can be interpreted as trying to understand the properties of this manifold.

Since studying the natural image set directly is an ill-posed problem, image modeling researchers have typically studied the more gross, or aggregated, properties of images, such as their local smoothness, self-similarity, and so on. The wavelet transform has proven a useful tool in this regard.

The wavelet transform can be interpreted as a multiscale differentiator or edge detector that represents the singularity content of an image at

³1000 $\times$ 1000 digital photographs lie in $\mathbb{R}^{10^6}$! Moreover, the intuition that we humans build up from inhabiting $\mathbb{R}^3$ easily leads us astray in higher-dimensional spaces [10].

multiple scales and three different orientations — horizontal, vertical, and diagonal. Roughly speaking, each image singularity is represented by a cascade of large wavelet coefficients across scale [13]. If the singularity is within the support of a wavelet basis function, then 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.

Several features of wavelet transforms make the wavelet domain well suited to constructing deterministic and statistical image models [13, 14]:

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

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

W3. Energy Compaction: Wavelet transforms of natural images tend to be sparse. A wavelet coefficient is large only if singularities are present within the support of the corresponding wavelet basis element.

W4. Decorrelation: The wavelet coefficients of natural images tend to be approximately statistically 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 small portion of a typical image contains edges. The Compaction and Decorrelation properties (**W3**, **W4**) simplify the statistical modeling of natural images in the wavelet domain as compared with a direct spatial-domain modeling.

The Compaction (**W3**) of signal energy in the wavelet domain results in a nonGaussian marginal probability density of the wavelet coefficients as shown in Figure 1.1(a):

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

As we will see, the Decorrelation property (**W4**) inspires simple spatially localized modeling of the wavelet coefficients. Approaches include modeling each wavelet coefficient independently with a nonGaussian marginal distribution, such as a generalized Gaussian distribution [3, 15] or Gaussian mixture [4, 5].

The self-similarity of natural images [16] translates into a $1/f^\gamma$ Fourier spectrum behavior (see Figure 1.1(b)). This behavior translates into the wavelet domain as (see Figure 1.1(c)):

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

The decay rate is directly related to the image smoothness [13].

Figure 1.1. (a) Peaky and heavy-tailed marginal histogram of the finest scale wavelet coefficients of the Cameraman image. (b) Log-log magnitude plot of the Fourier spectrum of a 1-D slice of the Cameraman image. The spectral decay is close to linear, which implies the $1/f^\gamma$ property. (c) The $1/f^\gamma$ spectral decay induces an exponential decay of wavelet transform variance with scale, plotted here for the Lena, Cameraman, Boats, and Bridge test images.

1.3 Deterministic Image Models

In this section, we introduce Besov spaces as deterministic models for natural images. Since the Besov norm measures the smoothness of an image, it is no wonder that this norm can be related to the local derivatives computed by the wavelet transform. Indeed, wavelets provide the natural (unconditional) bases⁴ for these spaces [9].

1.3.1 Besov function spaces

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

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

$$\Delta_h^{(r)} z(t) := \sum_{k=0}^r \binom{r}{k} (-1)^k z(t + kh) \quad (1.5)$$

for $t \in I_h^r := \{t \in I \mid t + rh \in I\}$. The $L_p(I)$ -modulus of smoothness for $0 < p \leq \infty$ is defined by

$$\omega_r(z, t)_p := \sup_{|h| \leq t} \|\Delta_h^{(r)} z\|_{L_p(I_h^r)}. \quad (1.6)$$

⁴A basis $\{e_n\}$ for a set E is called *unconditional* if (i) for $\sum_n \mu_n e_n \in E$, $\sum_n |\mu_n| e_n \in E$, and (ii) for $\sum_n \mu_n e_n \in E$ and $\epsilon_n = \pm 1$ randomly chosen for every n , $\sum_n \epsilon_n \mu_n e_n \in E$.

The *Besov seminorm* of index (α, p, q) is defined for $r > \alpha$, $0 < p, q \leq \infty$ by

$$|z|_{B_q^\alpha(L_p(I))} := \left\{ \int_0^\infty \left(\frac{\omega_r(z, t)_p}{t^\alpha} \right)^q \frac{dt}{t} \right\}^{1/q}, \text{ if } 0 < q < \infty \quad (1.7)$$

and by

$$|z|_{B_\infty^\alpha(L_p(I))} := \sup_{0 < t < \infty} \left\{ \frac{\omega_r(z, t)_p}{t^\alpha} \right\}. \quad (1.8)$$

The *Besov norm* is then given by

$$\|z\|_{B_q^\alpha(L_p(I))} = \|z\|_{L_p(I)} + |z|_{B_q^\alpha(L_p(I))}. \quad (1.9)$$

The Besov space $B_q^\alpha(L_p(I))$ is the class of functions $z : I \rightarrow \mathbb{R}$ 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.

1.3.2 Wavelets and Besov space

Wavelets provide a simple characterization for the Besov spaces $B_q^\alpha(L_p(I))$. For analyzing ϕ and ψ possessing more than α vanishing moments [13], the Besov norm $\|z\|_{B_q^\alpha(L_p(I))}$ is equivalent⁵ to a sequence norm of the wavelet coefficients [17]

$$\|z\|_{B_q^\alpha(L_p(I))} \asymp \left(\|u_{j_0}\|_p + \left(\sum_{j \geq j_0} 2^{jq(\alpha+1-2/p)} \left(\sum_{\mathbf{k}, \psi \in \Psi} |w_{j,\mathbf{k},\psi}|^p \right)^{q/p} \right)^{1/q} \right). \quad (1.10)$$

The three hyper-parameters have natural interpretations: we take the ℓ_p norm of the wavelet coefficients within each scale (fixed j), weight these norms by an exponential factor of the smoothness parameter α , and then take the ℓ_q norm across scale j . We will take (1.10) as the definition of the Besov norm in the following.

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

$$\|z\|_{B_p^\alpha(L_p(I))} \asymp \|u_{j_0}\|_p + \left(\sum_{j \geq j_0, \mathbf{k}, \psi \in \Psi} 2^{j(\alpha p + p - 2)} |w_{j,\mathbf{k},\psi}|^p \right)^{1/p}, \quad (1.11)$$

which is a weighted ℓ_p norm of the wavelet coefficients. The $B_1^1(L_1)$ and $B_\infty^1(L_1)$ spaces are interesting, because $B_1^1(L_1(I)) \subset \text{BV}(I) \subset B_\infty^1(L_1(I))$,

⁵Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ are *equivalent* and denoted $\|\cdot\|_a \asymp \|\cdot\|_b$ if and only if there exist positive constants m and M such that $m\|x\|_a \leq \|x\|_b \leq M\|x\|_a$ for all x .

with the set of bounded variation images $BV(I)$ [18, 19] a popular image model in its own right. These Besov norms have a particularly simple structure:

$$\|z\|_{B_1^1(L_1(I))} \asymp \|u_{j_0}\|_1 + \sum_{j \geq j_0, \mathbf{k}, \psi \in \Psi} |w_{j,\mathbf{k},\psi}| \quad (1.12)$$

and

$$\|z\|_{B_\infty^1(L_1(I))} \asymp \|u_{j_0}\|_1 + \sup_{j \geq j_0} \sum_{\mathbf{k}, \psi \in \Psi} |w_{j,\mathbf{k},\psi}|. \quad (1.13)$$

1.3.3 Limitations of Besov space

While Besov spaces have proved enormously useful in a range of image processing applications, they are not without their shortcomings.

First is the proper interpretation of Besov norm. While it is clear that within the same Besov space $B_q^\alpha(L_p(I))$ images with small Besov norm are smoother, it is not immediate to interpret the Besov norm as an indicator of how much a given image “looks like” a natural image.

Second is the fact that Besov spaces contain continuous-variable functions, but, in practice, images are discretely sampled versions whose fine-scale information is truncated. Since the Besov norm depends greatly on the fine-scale wavelet coefficients, it is not straightforward to verify the smoothness of the original image given only its samples. As one extreme example, if we assume that all of the unavailable fine-scale wavelet coefficients are zero, then the continuous-space image corresponding to the sampled image is infinitely smooth in the sense of Besov smoothness and resides in every Besov space.

Third, as we readily see from the definition in (1.11), the Besov norm is invariant to coefficient *shuffles* (permutations) within scale j . Because the shuffling of wavelet coefficients destroys the edge structure of the image, a wavelet-shuffled image does not resemble a natural image (see Figure 1.2(b)). However, the wavelet-shuffled image belongs to the same Besov space (with identical Besov norm) as the original image. This implies that there exist many functions in a given Besov space that do not resemble natural images. This is equivalent to saying that the Besov norm lacks spatial localization of image smoothness. For more accurate image modeling, we must develop models that adapt to the local image characteristics.

Fourth, also from (1.11), we see that the Besov norm is invariant to *sign changes* of the wavelet coefficients. (Indeed, this follows directly from the fact that wavelets form an unconditional basis for Besov spaces [9].) As observed in Figure 1.2(c), the image resulting from random wavelet coefficient sign flips also does not resemble a natural image, while it belongs to the same Besov space as the original image with identical Besov norm.

Fifth, the set of natural images is not a linear space. Images are not formed by additive superpositions but rather by complicated nonlinear in-

Figure 1.2. Limitations of Besov space and norm. (a) Image and inverse wavelet transforms of its wavelet coefficients after (b) random shuffles (permutations) within each scale band, and (c) random sign flips. All three images have identical Besov norm, indicating that the Besov norm is blind to much image structure.

Figure 1.3. Natural images do not form a linear space. For example, we would not form a new “natural image” by adding the cameraman and fruit images.

interactions like occlusions. See Figure 1.3, for example. The Besov spaces $B_q^\alpha(L_p(I))$ with $p < 1$ are non-convex, however, and so perhaps all is not lost.

For an extended discussion of additional Besov space shortcomings, including the difficulties encountered in estimating the Besov parameters p, q, α , see [20].

To overcome these limitations of Besov space as a characterization of natural images, it is clear that we must construct new spaces or sets based on more accurate image models. While little progress has been made to date in the deterministic setting, considerable progress has been made in the statistical setting, to which we now turn.

1.4 Statistical Image Models

Statistical image models specify the distribution of the random process generating the images of interest. In this section, we consider two common wavelet-domain statistical image models. We will discuss only models for the wavelet coefficients. The extension to the scaling coefficients is straightforward.

Figure 1.4. Fit (dashed) to the log of the wavelet coefficient histogram of Figure 1.1(a) (solid) by a generalized Gaussian distribution with parameter $\nu = 0.7$.

1.4.1 Independent generalized Gaussian model

The simplest wavelet-domain statistical models assume that the wavelet coefficients are statistically independent (extrapolating from the approximate decorrelation property **W4**). Under this assumption, modeling reduces to simply specifying the marginal distribution of each wavelet coefficient.

We begin with a model that aims to match the peaky, heavy-tailed marginal and exponential variance decay we saw in Figure 1.1. The ubiquitous Gaussian distribution is not a good choice here. A better choice is the zero-mean *generalized Gaussian distribution* (GGD) $\text{GGD}(\sigma^2; \nu)$ with variance σ^2 and shape parameter ν [15]. This probability density function (pdf) is defined as

$$f(x) := \frac{\nu \eta(\nu)}{2\Gamma(1/\nu)} \frac{1}{\sigma} \exp \left\{ - \left(\frac{\eta(\nu)|x|}{\sigma} \right)^\nu \right\}, \quad (1.14)$$

with $\eta(\nu) = \sqrt{\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}}$. The GGD model contains the Gaussian and Laplacian distributions 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 tractability, all wavelet coefficients at scale j are assumed to be independent and identically distributed (iid) with the same variance σ_j^2 and shape ν_j . Under this *iid in scale* model, we set $w_{j,k,\psi} \stackrel{\text{iid}}{\sim} \text{GGD}(\sigma_j^2; \nu_j)$. Thus, σ_j^2 and ν_j do not depend on the spatial location $\mathbf{k}$.

We choose the shape parameter ν_j to match the peakiness and heavy tail of the wavelet coefficient pdf. A large class of natural images share similar shape parameters [3]. In particular, the single fixed value $\nu_j \approx p = 0.7$ across all scales appears quite accurate for many natural images (see Figure 1.4).

The variance σ_j^2 represents the wavelet coefficient energy at scale j . It can be empirically estimated based on the given data [3], or it can be specified to decay exponentially (property **W6**, recall Figure 1.1(c)).

Figure 1.5. (a) Wavelet subband of a *cerulean warbler* bird song $w_{j_0,k}$, (b) its standard deviation $\hat{\sigma}_{j_0,k}$ computed in a local window, and (c) the rescaled subband $r_{j_0,k} = w_{j_0,k}/\hat{\sigma}_{j_0,k}$. Note the dramatic decrease in dynamic range. (d) Log histograms of $w_{j_0,k}$ (inner) and $r_{j_0,k}$ (outer). Note how the rescaling “Gaussianizes” the pdf (makes the log histogram closer to quadratic).

With fixed shape parameter p and fixed variance decay parameter β , the iid-in-scale GGD model Θ_p^β takes the form

$$\Theta_p^\beta : w_{j,k,\psi} \stackrel{\text{iid}}{\sim} \text{GGD}(\sigma_j^2; p) \text{ with } \sigma_j = 2^{-j\beta} \sigma_0. \quad (1.15)$$

The success of this simple GGD wavelet model in image denoising and texture classification problems [3, 6] is testimony to its accuracy for natural images.

1.4.2 Local Gaussian model

This model is inspired by how humans cope with the huge dynamic range in natural imagery and sounds — up to $1 : 10^5$ — that, if left untreated, would saturate the sensors in our eyes and ears. Humans depend on a form of *automatic gain control* encoded in the neural circuits of their sensory systems. The basic idea in the visual system is that the sensitivity of the eye decreases in bright regions and increases in dark regions. A roughly equivalent operation on an image is to renormalize each image pixel by some function of the energy of the pixels around it. Moreover, we can perform the same operation in the wavelet domain. Figure 1.5 illustrates the idea for a 1-D bird song signal with high dynamic range.

Local energy normalization has a calming effect on the marginal distribution of the wavelet coefficients, pulling up the values of the small coefficients and squashing down the large values. We see from Figures 1.5 and 1.6 that this is indeed the case. To compute the histogram in Figure 1.6, we renor-

Figure 1.6. Histogram of wavelet coefficients in a subband of the Lena image after local variance normalization. The variance was estimated as the average energy of the wavelet coefficients in a 3×3 window around each coefficient. The distribution is much closer to a Gaussian than the unnormalized histogram of Figure 1.1(a).

malized each wavelet coefficient $w_{j,k}$ in a subband of the Lena image by dividing by the standard deviation of its eight surrounding neighbors in a 3×3 window. The resulting normalized coefficients are given by

$$\tilde{w}_{j,k} := \frac{w_{j,k}}{\hat{\sigma}_{j,k}} \quad (1.16)$$

with the local standard deviation computed from

$$\hat{\sigma}_{j,k}^2 := \frac{1}{L-1} \sum_{i \in \mathcal{N}_L(k)} w_{j,i}^2 \quad (1.17)$$

and $\mathcal{N}_L(k)$ an L -point neighborhood around the point k . (In our example, $L = 8$.)

Figures 1.5 and 1.6 suggest a simple *locally Gaussian model* only slightly more complex than the iid GGD model from Section 1.4.1. We model the wavelet coefficients as independent Gaussian random variables but with a slowly varying variance field $\hat{\sigma}_{j,k}^2$ computed from (1.17)

$$\Omega_2 : w_{j,k} \sim \mathcal{N}(0, \hat{\sigma}_{j,k}^2). \quad (1.18)$$

Although this model assumes that each coefficient is independent, the local variance estimation procedure implicitly captures dependencies between the sizes of neighboring coefficients. Thus, the model describes the wavelet coefficient joint statistics more accurately than the iid GGD model. Since each wavelet coefficient is allowed to have a different variance, the overall statistics of an entire subband of coefficients will be a Gaussian mixture distribution that approximates a GGD.

This locally Gaussian model is precisely the model behind the high-performance Estimation-Quantization (EQ) image coding algorithm [21,

[22]. More general distributions than the Gaussian can also be chosen.⁶ In the EQ coder, the window takes the form of a causal neighborhood following the standard scanning order of the wavelet coefficients both within and across scale. State-of-the-art compression performance and simplicity of implementation (no zero trees or other fancy appendages) results. An image denoising algorithm using similar methods also performs very well [23]. The size of the variance estimation window in [23] is determined by a bootstrap.

1.5 Links between Besov Space and Statistical Models

While seemingly quite disparate frameworks, there are deep commonalities between the above deterministic and statistical image models. In this section, we uncover a strong link between the Besov space model of Section 1.3 and the wavelet statistical image models of Section 1.4.

1.5.1 GGD realizations live in Besov space

The connection between wavelet-domain statistical models and Besov spaces was perhaps first noticed by Abramovich et al. [5] in their variation of the wavelet Gaussian mixture model. The decay of wavelet coefficients across scale determines the smoothness of the corresponding image; hence realizations of a statistical model with exponentially decaying variance will belong to certain Besov spaces. Interpreted in terms of the independent GGD model of (1.15), we have the following theorem.⁷

Theorem 1. *Suppose each wavelet coefficient $w_{j,k,\psi}$ is distributed according to the independent GGD model Θ_p^β from (1.15) with $\beta > 0$ and $\sigma_0 > 0$. Then, for $0 < p, q < \infty$, the image realizations from this model are almost surely in $B_q^\alpha(L_p(I))$ if and only if $\beta > \alpha + 1$.*

A very similar result applies for a large class of wavelet coefficient marginal pdf models, including the Gaussian mixture [5, 24] and even finite support pdf's [8]. It is also possible to impose further dependencies in the Gaussian mixture model to capture additional wavelet properties such as magnitude persistence across the scales of the wavelet tree [25].

⁶LoPresto et al. [22] argue that a GGD is more appropriate than a Gaussian distribution in the EQ algorithm only because in the compression case the variance must be estimated from *quantized* coefficients. If we are allowed to estimate the variance from the original coefficients, then the Gaussian becomes a better choice.

⁷The proof parallels the proof of Theorem 1 in [5].

Theorem 1 tells us that the class of realizations from the independent GGD statistical model is essentially a Besov space. But what of the Besov norm, which measures distance in this space?

1.5.2 Besov norm as a normalized likelihood

Consider normalizing the likelihood of a realization of a statistical model in terms of the maximum likelihood achievable. That is, define the *normalized likelihood function* by

$$f^N(x|\Theta) := \frac{f(x|\Theta)}{\sup_x f(x|\Theta)}, \quad (1.19)$$

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

For an infinite sequence of random variables, we can define the normalized likelihood as the limit (if it exists) of the normalized likelihood of its truncated subsequences. For the independent wavelet-domain models considered in Section 1.3, we can compute the limit as we move to finer scales, defining the normalized likelihood

$$f^N(\mathbf{w}) = \lim_{J \rightarrow \infty} \frac{\prod_{j=0}^J \prod_{\mathbf{k}, \psi} f_{j, \mathbf{k}, \psi}(w_{j, \mathbf{k}, \psi})}{\sup \prod_{j=0}^J \prod_{\mathbf{k}, \psi} f_{j, \mathbf{k}, \psi}(w_{j, \mathbf{k}, \psi})} \quad (1.20)$$

when $w_{j, \mathbf{k}, \psi} \sim f_{j, \mathbf{k}, \psi}(w_{j, \mathbf{k}, \psi})$. For the independent GGD model with exponentially decaying variance Θ_p^β from (1.15), the normalized likelihood is well defined, since the supremum is finite and the limit exists.

For the model Θ_p^β , the normalized likelihood computed using the coefficients between scales 0 and J equals

$$\begin{aligned} f^N(\mathbf{w}_J) &= \prod_{j=0}^J \prod_{\mathbf{k}, \psi} \exp \left\{ - \left(\frac{\eta(p)|w_{j, \mathbf{k}, \psi}|}{\sigma_j} \right)^p \right\} \\ &= \exp \left\{ \sum_{0 \leq j \leq J, \mathbf{k}, \psi} - \left(\frac{\eta(p)|w_{j, \mathbf{k}, \psi}|}{\sigma_j} \right)^p \right\}. \end{aligned} \quad (1.21)$$

Taking the negative log of the normalized likelihood function, we obtain

$$\begin{aligned} -\log f^N(\mathbf{w}) &= \sum_{0 \leq j, \mathbf{k}, \psi} \left(\frac{\eta(p)|w_{j, \mathbf{k}, \psi}|}{\sigma_j} \right)^p \\ &= \eta(p)^p \sum_{0 \leq j, \mathbf{k}, \psi} (2^{j\beta} |w_{j, \mathbf{k}, \psi}|)^p, \end{aligned} \quad (1.22)$$

which is equivalent to $\|z\|_{B_p^\alpha(L_p)}^p$, the Besov norm of the function z , when $\beta = \alpha + 2 - 2/p$.

With $p = 2$, we obtain an iid Gaussian model for the wavelet coefficients, and the corresponding normalized negative log likelihood is equivalent to the Sobolev norm of z . With $p = 0.7$ — corresponding to the $\nu = 0.7$ that closely matches the generalized Gaussian exponent of many real images (recall Figure 1.4) — we obtain a non-convex set $B_p^\alpha(L_p)$ (recall the discussion of Section 1.3.2).

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

$$\{w : f^N(w|\Theta_p^\beta) \neq 0\} \text{ with } \beta = \alpha + 2 - 2/p. \quad (1.23)$$

Thus, the signals in the Besov space $B_p^\alpha(L_p)$ are the “likely” signals under the statistical model Θ_p^β .

This link between Besov space and the GGD statistical model immediately unites many seemingly disparate algorithms. For instance, the deterministic variational problem of Chambolle et al. [26] corresponds to a Bayesian wavelet-based statistical estimator regularized by the likelihood of the approximant. The deterministic interpolation algorithm proposed by Choi et al. [27] becomes a maximum likelihood interpolator under the wavelet-domain GGD model.

1.5.3 Ties to information theory and coding

Much progress has been made on modeling image pdf’s in the context of *image coding*. A typical wavelet-domain image compressor consists of two stages. First, the pdf of each wavelet coefficient is estimated as accurately as possible. (This step corresponds to confining the given image to a Besov-like set.) Then, the coefficient is quantized according to that pdf. (This corresponds to specifying the location of the image in the set.) For efficient compression, we must make the confining set as small as possible to reduce the number of code bits needed to specify the location within the set.

When images are described by a statistical model, the optimal compression performance is achieved by the Shannon source code for the model [28]. State-of-the-art image compression algorithms feature underlying statistical image models either implicitly [29] or explicitly [21]. Images conforming to the pdf model are well compressed, and thus the compressed codelength of an image in some sense measures how much it can be considered a “natural image” (at least in the eyes of the compression algorithm). The Shannon codelength enables an information-theoretic characterization of the natural image set.

All leading wavelet-domain image compression algorithms employ some type of spatially adaptive pdf estimation procedure, which is often hidden in the algorithm. To minimize the number of bits to describe the pdf prediction itself, a very simple parametric pdf form, such as a zero-mean generalized Gaussian or Laplacian, is typically assumed.

First, consider a simple compression algorithm based on the independent GGD model Θ_p^β from (1.15). Let w_J be the set of all the wavelet coefficients up to scale J . Under the Θ_p^β model, the pdf of w_J is given by

$$\begin{aligned} f(w_J) &= C_J \prod_{j=0}^J \prod_{\mathbf{k}, \psi} \exp \left\{ - \left(\frac{\eta(p) |w_{j,\mathbf{k},\psi}|}{\sigma_j} \right)^p \right\} \\ &= C_J \exp \left\{ \sum_{0 \leq j \leq J, \mathbf{k}, \psi} - \left(\frac{\eta(p) |w_{j,\mathbf{k},\psi}|}{\sigma_j} \right)^p \right\}, \end{aligned} \quad (1.24)$$

with

$$C_J := \prod_{j=0}^J \prod_{\mathbf{k}, \psi} \frac{\nu \eta(p)}{2\Gamma(1/p)} \frac{1}{\sigma_j}. \quad (1.25)$$

Because the exponential term equals 1 when all the wavelet coefficients are zero, C_J corresponds to the likelihood of an image having all zero wavelet coefficients, that is, a constant image. Using optimal source coding for this model, the Shannon codelength L_J of the wavelet coefficients up to scale J is given by

$$L_J := -\log f(w_J) = -\log C_J - \sum_{0 \leq j \leq J, \mathbf{k}, \psi} \left(\frac{\eta(p) |w_{j,\mathbf{k},\psi}|}{\sigma_j} \right)^p. \quad (1.26)$$

Recalling that $-\log C_J$ is the Shannon codelength of a constant image, we can write

$$\tilde{L}_J := L_J - (-\log C_J) = \sum_{0 \leq j \leq J, \mathbf{k}, \psi} \left(\frac{\eta(p) |w_{j,\mathbf{k},\psi}|}{\sigma_j} \right)^p. \quad (1.27)$$

If we set $\beta = \alpha + 2 - 2/p$ and compute the limit of $\tilde{L}_J$ as $J \rightarrow \infty$, we obtain the relation

$$\lim_{J \rightarrow \infty} \tilde{L}_J = \eta(p)^p \|z\|_{B_p^\alpha(L_p)}^p, \quad (1.28)$$

where $\alpha = \beta - 2 + 2/p$. Because the constant image is the simplest image to code, we see that the Besov norm (ignoring the constant $\eta(p)^p$) measures how much more codelength is required to code the image z beyond that required for the constant image. In this sense, the Besov norm measures the coding complexity of the image under the iid GGD model.

The interpretation of the Besov norm as the Shannon codelength indicates that the Besov norm measures the coding complexity of images (see also [30]). Thus, the Besov regularization of Chambolle et al. [26] is equivalent to the minimum description length (MDL) denoising algorithm [31] under the independent GGD image prior. And the maximum smoothness interpolation algorithm of [27] simply finds the least complex signal among all signals satisfying the sampling constraints.

Unfortunately, the close equivalence between Besov spaces and the iid GGD model implies that the iid GGD model suffers from all of the same problems as Besov spaces, including shuffle and sign invariance (recall Figure 1.2).

1.5.4 Local Gaussian model = local Sobolev space

Under the locally Gaussian model Ω_2 from (1.18), the likelihood computation and definition of the normalized likelihood are straightforward. The likelihood of the entire set of wavelet coefficients is given by the product

$$f(\mathbf{w}) = \prod_{j,k} g(w_{j,k}), \quad (1.29)$$

where $g(w) := (2\pi\hat{\sigma}_{j,k}^2)^{-1/2} \exp(-\frac{w^2}{2\hat{\sigma}_{j,k}^2})$ is the zero-mean Gaussian density function with variance $\sigma_{j,k}^2$.

From (1.29), the negative normalized log likelihood becomes

$$-\log f^N(\mathbf{w}) = \sum_{j,k} \frac{w_{j,k}^2}{2\hat{\sigma}_{j,k}^2}, \quad (1.30)$$

a weighted ℓ_2 “norm” of the wavelet coefficients that resembles a locally adapted Sobolev norm.⁸ As a notation for the “norm” defined using this model, we use $\|\cdot\|_{IG}$. That is,

$$\|z\|_{IG}^2 := \sum_{j,k} \frac{w_{j,k}^2}{2\hat{\sigma}_{j,k}^2}. \quad (1.31)$$

The measure $\|z\|_{IG}^2$ is *not* shuffle invariant. However, since this spatially adapted model also uses zero-mean pdfs, the measure remains invariant to sign flips of the wavelet coefficients.

1.6 Conclusions and Beyond Besov Spaces

We have discovered several relationships between deterministic Besov spaces and wavelet-domain statistical image models. In particular, we have shown that the Besov norm is equivalent to a log likelihood or Shannon codelength under an independent GGD wavelet model.

While much progress has been made in both deterministic and statistical image modeling, we are still far from pinning down the set of natural images in the wavelet (or any other) domain. The most recent progress has been

⁸Note that (1.30) is not a valid norm, since it will not converge in general.

made in the statistical context, attacking the shuffle invariance and sign flip invariance of the Besov norm and the independent GGD wavelet model.

Jaffard's *oscillation spaces* are not invariant to wavelet-coefficient shuffling [8]. Likewise for the statistical *hidden Markov tree* (HMT) model, which fuses a Gaussian mixture marginal pdf with Markov dependencies through scale to capture the persistence of large and small wavelet coefficient values [4]. Encouragingly, the deterministic image model induced by the HMT is not a convex space. The promising complex wavelet HMT model [25] is neither shuffle invariant (thanks to an HMT model on the complex wavelet magnitude values) nor sign flip invariant (thanks to a model on the complex wavelet phase).

Future progress will require explicit attention to the fact that image edges lie along regular 1-D contours. The complex HMT model mentioned above has the capability to capture this property. Alternatively, we can move from the wavelet transform to other transforms — *curvelets* [32, 33] or *bandelelets* [34], for example — that directly restructure the edge contours into a form that is easier to model.

Clearly the development of appropriate transforms and modeling frameworks for the transform coefficients of natural images will remain a challenging area of research for some time to come.

References

- [1] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, pp. 1200–1224, Dec. 1995.
- [2] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. on Information Theory*, vol. 38, no. 2, pp. 719–746, March 1992.
- [3] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian priors," in *Proc. IEEE-SP Int. Symp. Time-Freq. and Time-Scale Anal.*, Pittsburgh, PA, Oct. 6–9 1998, pp. 633–636.
- [4] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.*, vol. 46, no. 4, pp. 886–902, April 1998.
- [5] F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *J. Roy Stat. Soc. Ser. B*, vol. 60, pp. 725–749, 1998.

- [6] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE Int. Conf. on Image Proc. — ICIP 1996*, Lausanne, Switzerland, Sept. 1996.
- [7] H. Choi and R. G. Baraniuk, "Image segmentation using wavelet-domain classification," in *Proc. SPIE Conf. Math. Modeling, Bayesian Estimation, and Inverse Problems*, Denver, CO, July 1999, vol. 3816, pp. 306–320.
- [8] S. Jaffard, "Beyond Besov spaces," *Preprint*, 2001.
- [9] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, New York, 1992.
- [10] E. A. Abbot and A. Lightman, *Flatland: A Romance of many dimensions*, Penguin Classics, New York: NY, 1998.
- [11] A. B. Lee, K. S. Pederson, and D. Mumford, "The complex statistics of high-contrast patches in natural images," in *Proc. 2nd International IEEE Workshop on Statistical and Computational Theories of Vision*, Vancouver, July 2001.
- [12] M. T. Orchard, *Personal Communication*, 2000.
- [13] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1998.
- [14] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice Hall, Englewood Cliffs, NJ, 1995.
- [15] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.
- [16] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Phys. Rev. Lett.*, vol. 73, no. 6, pp. 814–817, 1994.
- [17] Y. Meyer, *Wavelets and Operators*, Cambridge University Press, Cambridge, 1992.
- [18] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton, FL, 1992.
- [19] S. Osher L. I. Rudin and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.
- [20] C. S. Güntürk, "Harmonic analysis of two problems in signal quantization and compression," *Ph.D Thesis, Princeton University*, 2000.
- [21] S. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Data Compression Conference*, Snowbird, Utah, 1997, pp. 221–230.

- [22] S. LoPresto, K. Ramchandran, and M. T. Orchard, "Wavelet image coding via rate-distortion optimized adaptive classification," in *Proc. of NJIT Symposium on Wavelet, Subband and Block Transforms in Communications*, New Jersey Institute of Technology, 1997.
- [23] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Proc. — ICASSP '99*, Phoenix, AZ, March 1999.
- [24] H. Choi and R. Baraniuk, "Wavelet-domain statistical models and Besov spaces," in *Proc. of SPIE Conf. Wavelet Applications in Signal Proc. VII*, Denver, July 1999, vol. 3813, pp. 489–501.
- [25] J. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian wavelet domain image modeling using hidden Markov models," *IEEE Trans. on Image Proc.*, vol. 10, pp. 1056–1068, July 2001.
- [26] 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.*, vol. 7, pp. 319–355, March 1998.
- [27] H. Choi and R. Baraniuk, "Interpolation and denoising of nonuniformly sampled data using wavelet domain processing," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Proc. — ICASSP '99*, Phoenix, AZ, March 1999.
- [28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 1991.
- [29] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.
- [30] A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard, "On the importance of combining wavelet-based non-linear approximation with coding strategies," *IEEE Trans. on Information Theory*, Submitted.
- [31] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the MDL criterion," in *Wavelets in Geophysics*, 1994, pp. 299–324, New York: Academic Press, Editors: E. Foufoula-Georgiou and P. Kumar.
- [32] E. Candès and D. Donoho, "Ridgelets: The key to high-dimensional intermittency?," *Phil. Trans. R. Soc. Lond. A.*, vol. 357, pp. 2495–2509, 1999.

- [33] E. Candès and D. Donoho, "Curvelets: A surprisingly effective non-adaptive representation of objects with edges," in *Curves and Surface Fitting: Saint-Malo 1999*. 2000, Vanderbilt University Press, Nashville, TN, Editors A. Cohen, C. Rabut and L. L. Schumaker.
- [34] E. L. Pennec and S. Mallat, "Image compression with geometrical wavelets," in *IEEE Int. Conf. on Image Proc. — ICIP '01*, Thessaloniki, Greece, Oct. 7–10 2001.