

# CLASSIFYING THE SMOOTHNESS OF IMAGES: THEORY AND APPLICATIONS TO WAVELET IMAGE PROCESSING\*

RONALD A. DEVORE<sup>1</sup>, University of South Carolina, AND BRADLEY J. LUCIER<sup>2</sup>, Purdue University

## Abstract

Devore, Jawerth, and Lucier have previously introduced a definition of the smoothness of images that is directly related to the performance of wavelet compression schemes. In this paper we survey previous results on the equivalence between smoothness, rate of decay of the wavelet coefficients, and efficiency of wavelet compression techniques applied to images. We report on other applications including deciding how many pixel quantization intervals are needed to preserve smoothness, and the fast solution of variational problems that arise naturally in several areas of image processing.

## 1. Introduction

The authors of [3] ask the following natural question: “How can one classify or determine which images can be compressed well by various methods of image compression, and specifically by wavelet compression methods?” After this question is made precise, the answer is surprisingly direct—an image can be compressed well if and only if it is contained in certain smoothness spaces called Besov spaces. Whether an image is contained in a smoothness space depends on (a) the rate of decay of a quantity called the modulus of smoothness of the image, or (b) the rate of decay of the coefficients of a wavelet expansion of the image. Because good compression rates are equivalent to membership in these smoothness spaces, images in these smoothness spaces have the *minimal* smoothness necessary to be approximated well by wavelet compression methods. Furthermore, it has been shown [8] that for images in these smoothness classes, no “stable” method of image compression can achieve a higher rate of compression than wavelet-based methods.

Using information theory one can define certain classes of images and determine the *average* performance of certain compression schemes over all images in a given class. The performance of a compression algorithm on a *particular* image in a given class may differ widely from the average. In our framework, the performance of wavelet compression schemes on a particular image is determined by the smoothness of that image in a particular smoothness space.

\*Appeared in ICIP-94, Proceedings of the 1994 IEEE Conference on Image Processing, Nov. 13–16, 1994, Austin, TX, IEEE Press, Los Alamitos, CA, Vol. II, pages 6–10.

<sup>1</sup>Department of Mathematics, University of South Carolina, Columbia, SC 29208. devore@math.scarcolina.edu

<sup>2</sup>Department of Mathematics, Purdue University, West Lafayette, IN 47907. lucier@math.purdue.edu

Once one takes the point of view that we should characterize images by membership in these minimal smoothness spaces, several applications arise. For example, one can determine how to increase the number of pixel quantization intervals as the spatial resolution increases so that the digitized image retains the smoothness of the original intensity field. One can devise wavelet compression methods using scalar quantization that achieve the optimal rate of compression. One can determine how quantization strategies relate to error metrics. And one can solve many variational problems related to image reconstruction and noise removal directly and simply using wavelets [6].<sup>3</sup> In this paper we give an overview of these and other applications of our view of the smoothness of images.

## 2. The Size and Smoothness of Images

We begin with a real-valued intensity field  $F$  defined, for simplicity, on the unit square  $I = [0, 1]^2$ . We use the  $L_p(I)$  spaces,  $0 < p \leq \infty$ , to measure the size of an image. A function  $f$  is in  $L_p(I)$  whenever the  $L_p(I)$  norm of  $f$ ,

$$\|f\|_{L_p(I)} := \begin{cases} \left( \int_I |f(x)|^p dx \right)^{1/p}, & 0 < p < \infty, \\ \sup_{x \in I} |f(x)|, & p = \infty, \end{cases}$$

is finite. (The notation  $:=$  is to be read “defined as.”) For example,  $p = 2$  is the root-mean-square norm,  $p = 1$  is the mean-absolute norm, and  $p = \infty$  is the maximum norm. These spaces are nested—if  $p > p'$  and  $f \in L_p(I)$ , then  $f \in L_{p'}(I)$ , and, by Hölder’s inequality,  $\|f\|_{L_{p'}(I)} \leq \|f\|_{L_p(I)}$ .

Whenever  $p < 1$ ,  $\|\cdot\|_{L_p(I)}$  does not satisfy the triangle inequality, so it is not, strictly speaking, a norm, but a quasi-norm, for which there exists a constant  $C > 1$  such that for all  $f$  and  $g$  in  $L_p(I)$   $\|f + g\|_{L_p(I)} \leq C(\|f\|_{L_p(I)} + \|g\|_{L_p(I)})$ ; in our case,  $C = 2^{\frac{1}{p}-1}$ . We shall not distinguish further between norms and quasi-norms in this paper.

We measure the smoothness of images in smoothness spaces called Besov spaces. Here we begin with  $r$ th differences of the function  $f$ . The  $r$ th difference of a function  $f$  at a point  $x = (x_1, x_2)$  in the direction of a vector  $h = (h_1, h_2)$  is defined recursively as  $\Delta_h^0(f, x) := f(x)$  and

$$\Delta_h^r(f, x) := \Delta_h^{r-1}(f, x+h) - \Delta_h^{r-1}(f, x), \quad r > 0.$$

<sup>3</sup>Note added after publication: See also the paper by S. Osher and L. Rudin in this Proceedings, Vol. I, pages 31–35.

When applied to functions  $f$  in  $\mathcal{P}_r$ , the polynomials in two variables of total degree less than  $r$  (e.g.,  $\mathcal{P}_3$  consists of linear combinations of  $1, x_1, x_2, x_1^2, x_1x_2,$  and  $x_2^2$ ), we have  $\Delta_h^r(f, x) = 0$ ; derivatives have the same property. Heuristically, if  $f$  is “smooth” then  $f$  is close to a polynomial of degree  $< r$ , and both the  $r$ th derivatives of  $f$  and the  $r$ th differences of  $f$  should be “small”. Besov spaces are one way to make this idea precise.

We use the  $L_p(I)$  spaces to measure the size of the  $r$ th differences of  $f$ ; the resulting functions

$$\omega_r(f, t)_p := \sup_{|h|<t} \|\Delta_h^r(f, \cdot)\|_{L_p(I_{rh})}$$

are known as the *moduli of smoothness* of  $f$ . (The interval  $I_{rh}$  consists of all  $x \in I$  for which  $x + rh$  is in  $I$ , i.e., the set of  $x$  for which  $\Delta_h^r(f, x)$  is properly defined.) The supremum is over all vectors  $h$  of size less than  $t$  with any direction.

As  $t$  gets small,  $\omega_r(f, t)_p$  tends to zero. We measure the smoothness of  $f$  by measuring the rate at which  $\omega_r(f, t)_p$  tends to zero. For example, a function  $f$  is in the space  $B_\infty^\alpha(L_p(I))$ ,  $0 < \alpha < r$ , whenever  $\omega_r(f, t)_p = \mathcal{O}(t^\alpha)$ , and we can define the  $B_\infty^\alpha(L_p(I))$  semi-norm of  $f$  to be

$$|f|_{B_\infty^\alpha(L_p(I))} := \sup_{t>0} t^{-\alpha} \omega_r(f, t)_p.$$

(This is a semi-norm because  $|f|_{B_\infty^\alpha(L_p(I))} = 0$  if, e.g.,  $f$  is a constant.) In applications, one needs a more diverse family of spaces that measure more carefully the decay of  $t^{-\alpha} \omega_r(f, t)_p$  as  $t \rightarrow 0$ . The general Besov space  $B_q^\alpha(L_p(I))$ ,  $0 < \alpha < r$ ,  $0 < p \leq \infty$ , and  $0 < q \leq \infty$  is defined as the set of  $f$  for which

$$(1) \quad |f|_{B_q^\alpha(L_p(I))} := \left( \int_0^\infty [t^{-\alpha} \omega_r(f, t)_p]^q \frac{dt}{t} \right)^{1/q},$$

$0 < q < \infty$ , is finite; the semi-norm for  $q = \infty$  is defined above. One can show that the set of functions in  $B_q^\alpha(L_p(I))$  does not depend on  $r$  as long as  $r > \alpha$ . The  $B_q^\alpha(L_p(I))$  norm is given by

$$\|f\|_{B_q^\alpha(L_p(I))} := |f|_{B_q^\alpha(L_p(I))} + \|f\|_{L_p(I)}.$$

An equivalent semi-norm for  $B_q^\alpha(L_p(I))$  is

$$(2) \quad |f|_{B_q^\alpha(L_p(I))} \approx \left( \sum_{k=0}^\infty [2^{\alpha k} \omega_r(f, 2^{-k})_p]^q \right)^{1/q},$$

i.e., the  $\ell_q$  norm of the sequence  $\{2^{\alpha k} \omega_r(f, 2^{-k})_p\}_{k=0}^\infty$ , with the usual change to a supremum when  $q = \infty$ .

Simple properties of  $L_p(I)$  and  $\ell_q$  give properties about  $B_q^\alpha(L_p(I))$ . For example, since  $L_p(I) \subset L_{p'}(I)$  if  $p > p'$ , one has  $B_q^\alpha(L_p(I)) \subset B_q^\alpha(L_{p'}(I))$  if  $p > p'$ . Similarly, since  $\ell_q \supset \ell_{q'}$  if  $q > q'$ , one has  $B_q^\alpha(L_p(I)) \supset B_{q'}^\alpha(L_p(I))$  if  $q > q'$ . Since it is easily seen from (2)

that  $B_\infty^\alpha(L_p(I)) \subset B_q^{\alpha'}(L_p(I))$  if  $\alpha > \alpha'$ , one has immediately that  $B_q^\alpha(L_p(I)) \subset B_{q'}^{\alpha'}(L_p(I))$  if  $\alpha > \alpha'$  or  $\alpha = \alpha'$  and  $q < q'$ .

### 3. Wavelets

There are by now many introductions to wavelets; we recommend the book by Daubechies [2], or, for an introduction motivated by approximation theory, our survey article [7]. In this space we consider a particular family of biorthogonal wavelets discovered by Cohen-Daubechies-Faveau and Herley-Vetterli that are described in [2].<sup>4</sup>

For the present, we restrict our attention to one dimension with  $I := [0, 1]$ . The construction of the simplest wavelet, the Haar wavelet, begins with the *scaling* function  $\phi = \chi_I$ , which satisfies the rewrite rule

$$(3) \quad \phi(x) = \phi(2x) + \phi(2x - 1),$$

i.e., the characteristic function of  $[0, 1]$  is the sum of the characteristic functions of  $[0, 1/2]$  and  $[1/2, 1]$ . The associated wavelet,  $\psi$  is given by

$$\psi(x) = \phi(2x) - \phi(2x - 1).$$

Because we want to think about average intensity of images over square pixels, we introduce  $\tilde{\phi} = \phi$  and  $\tilde{\psi} = \psi$  and the two different scalings

$$(4) \quad \tilde{\phi}_{j,k}(x) := 2^k \phi(2^k x - j), \quad \phi_{j,k}(x) := \phi(2^k x - j)$$

for  $k \geq 0$  and  $0 \leq j < 2^k$ ; similarly, we define

$$(5) \quad \tilde{\psi}_{j,k}(x) := 2^k \psi(2^k x - j), \quad \psi_{j,k}(x) := \psi(2^k x - j).$$

These are the (scaled) dyadic dilates (by  $2^k$ ) and translates (by  $j/2^k$ ) of the scaling function  $\phi_2$ , its “dual” (see below)  $\tilde{\phi}$ , the wavelet  $\psi$ , and its dual  $\tilde{\psi}$ .

We can define piecewise constant approximations  $P_k f$  to any locally integrable function  $f$  on the intervals  $I_{j,k} := [j/2^k, (j+1)/2^k)$ ,  $0 \leq j < 2^k$ , by

$$P_k f := \sum_{0 \leq j < 2^k} \langle f, \tilde{\phi}_{j,k} \rangle \phi_{j,k}.$$

Here,  $\langle f, \tilde{\phi}_{j,k} \rangle := \int_I f \tilde{\phi}_{j,k}$  is the average of  $f$  on  $I_{j,k}$  and  $\phi_{j,k}$  is the characteristic function of  $I_{j,k}$ . If  $f \in L_2(I)$  then  $f = \lim_{k \rightarrow \infty} P_k f$ , so

$$f = P_0 f + \sum_{k \geq 0} (P_{k+1} f - P_k f).$$

Because

$$P_{k+1} f - P_k f = \sum_{0 \leq j < 2^k} \langle f, \tilde{\psi}_{j,k} \rangle \psi_{j,k},$$

<sup>4</sup>Note added after publication: The derivation of biorthogonal wavelets in this section owes much to algorithms like ENO in computational fluid dynamics that incorporate in an essential way reconstruction from cell averages. Ami Harten seems to have been the first to make this connection, and we regret not having cited this previous work.

the set  $\{\phi, \psi_{j,k} \mid k \geq 0, 0 \leq j < 2^k\}$  forms a complete orthogonal basis for  $L_2(I)$ , and, because of our scalings, we have for any  $f \in L_2(I)$ ,

$$(6) \quad f = \langle f, \tilde{\phi} \rangle \tilde{\phi} + \sum_{0 \leq k} \sum_{0 \leq j < 2^k} \langle f, \tilde{\psi}_{j,k} \rangle \psi_{j,k}.$$

To describe the Haar wavelet as one of a family of biorthogonal wavelets, we ask the following question. “If  $f$  is constant on  $[0, 1)$  and we know the average of  $f$  on  $[0, 1)$ , can we calculate the average of  $f$  on  $[0, 1/2)$  and  $[1/2, 1)$ ?” The answer, of course, is yes, the average of  $f$  on each subinterval is the same as the average of  $f$  on the big interval. This implies (3).

A more interesting question is the following: “If  $f = ax^2 + bx + c$  on  $[-1, 2)$ , and we know the average of  $f$  on  $[-1, 0)$ ,  $[0, 1)$ , and  $[1, 2)$ , can we determine the average of  $f$  on  $[0, 1/2)$  and  $[1/2, 1)$ ?” A moment’s reflection shows that we can determine  $a$ ,  $b$ , and  $c$  from the data, and then we can easily calculate the averages on  $[0, 1/2)$  and  $[1/2, 1)$ . One can continue with this procedure, calculating averages on  $[0, 1/4)$ ,  $[1/4, 1/2)$ , etc., and it has been shown that this procedure converges; in the limit one obtains a function  $\phi$  that satisfies  $\phi(x) = -\frac{1}{8}\phi(2x+2) + \frac{1}{8}\phi(2x+1) + \phi(2x) + \phi(2x-1) + \frac{1}{8}\phi(2x-2) - \frac{1}{8}\phi(2x-3)$ . If we now define  $\phi_{j,k}(x) = \phi(2^k x - j)$  and keep the same  $\tilde{\phi}$  and  $\tilde{\psi}_{j,k}$  as for the Haar wavelet, we have

$$P_k f = \sum_j \langle f, \tilde{\phi}_{j,k} \rangle \phi_{j,k}.$$

This formula does not hold near the boundary, and one must construct special boundary wavelets and dual functions, as in [1] and [11]. We ignore such technicalities here.

We have constructed  $\phi$  and  $\tilde{\phi}$  that satisfy

$$(7) \quad \phi(x) = \sum_j a_j \phi(2x-j) \text{ and } \tilde{\phi}(x) = \sum_j b_j \tilde{\phi}(2x-j)$$

for certain finite sequences  $(a_j)$  and  $(b_j)$ . The formulas for the scaling function and its dual lead to formulas for the associated wavelet  $\psi$  and *its* dual  $\tilde{\psi}$ :

$$(8) \quad \begin{aligned} \psi(x) &= \sum_j (-1)^j b_{1-j} \phi(2x-j) \text{ and} \\ \tilde{\psi}(x) &= \sum_j (-1)^j a_{1-j} \tilde{\phi}(2x-j). \end{aligned}$$

One can show that for any  $f$  in  $L_p(I)$ ,  $1 < p < \infty$ , formula (6) holds for these functions  $\phi$ ,  $\tilde{\phi}$ ,  $\psi$ , and  $\tilde{\psi}$ . (Again, we ignore technicalities at the boundary of  $[0, 1)$ .) This formula uses only function averages on the intervals  $I_{j,k}$  for its data, and it reproduces quadratic functions exactly. There is a family of such approximations, which reproduce polynomials of higher and higher (even) degree; details can be found in [2].

In two dimensions, we take  $\phi(x_1)\phi(x_2)$  for the scaling function,  $\tilde{\phi}(x_1)\tilde{\phi}(x_2)$  for its dual, and

$$\Psi = \{\psi(x_1)\phi(x_2), \phi(x_1)\psi(x_2), \psi(x_1)\psi(x_2)\}$$

for the set of wavelets, with the obvious duals. For each  $k \geq 0$  and  $j \in \mathbb{Z}_k^2 := \{(j_1, j_2) \in \mathbb{Z}^k \mid 0 \leq j_1 < 2^k, 0 \leq j_2 < 2^k\}$  we take

$$\phi_{j,k}(x) = \phi(2^k x - j) \text{ and } \tilde{\phi}_{j,k}(x) = 2^{2k} \tilde{\phi}(2^k x - j);$$

similarly for  $\psi_{j,k}$  and  $\tilde{\psi}_{j,k}$ . The biorthogonal wavelet decomposition for  $f$  in two dimensions is then

$$(9) \quad \langle f, \tilde{\phi} \rangle \tilde{\phi} + \sum_{k \geq 0} \sum_{j \in \mathbb{Z}_k^2} \sum_{\psi \in \Psi} c_{j,k,\psi} \psi_{j,k}, \quad c_{j,k,\psi} = \langle f, \tilde{\psi}_{j,k} \rangle.$$

How does this relate to the original intensity field  $F$  defined on  $[0, 1]^2$ ? First, since the duals  $\tilde{\psi}_{j,k}$  are piecewise constant, all the coefficients can be calculated *exactly* from piecewise constant pixel data. Furthermore, given values for  $2^K \times 2^K$  pixels, there are fast wavelet transforms, based on (7) and (8) that generate (9) with  $k < K$ . Finally, this approximation to  $F$  has convergence properties similar to a piecewise quadratic approximation; i.e., we obtain a third-order approximation to  $F$  from piecewise constant data.

#### 4. Wavelets and Smoothness Spaces

One can determine whether an image  $f$  defined on the square  $I = [0, 1]^2$  is in a Besov space  $B_q^\alpha(L_p(I))$  for many values of the parameters  $\alpha$ ,  $q$ , and  $p$  by examining the coefficients in the wavelet expansion of  $f$ . This is discussed in [12] and [10] on all of  $\mathbb{R}^2$  and in [11] for the framework in this paper. In particular, given (9), one has

$$(10) \quad \begin{aligned} \|f\|_{B_q^\alpha(L_p(I))} &\approx \|\langle f, \tilde{\phi} \rangle \tilde{\phi}\|_{L_p(I)} \\ &+ \left( \sum_{k \geq 0} \sum_{j \in \mathbb{Z}_k^2} \sum_{\psi \in \Psi} [2^{\alpha k} \|\langle f, \tilde{\psi}_{j,k} \rangle \psi_{j,k}\|_{L_p(I)}]^q \right)^{1/q} \end{aligned}$$

where  $A(f) \approx B(f)$  means that there are two positive constants  $C_1$  and  $C_2$  such that for all relevant  $f$ ,  $C_1 \leq A(f)/B(f) \leq C_2$ . This equivalence holds for large ranges of  $p$  and  $q$  whenever the scaling function  $\phi$  is in a slightly smoother space than  $B_q^\alpha(L_p(I))$ . This relationship between function smoothness and wavelet coefficients is most important.

If we consider the space  $B_q^\alpha(L_q(I))$  with the special value of the parameters

$$(11) \quad \frac{1}{q} = \frac{\alpha}{2} + \frac{1}{p}, \quad \alpha > 0, \quad 1 < p < \infty,$$

then we can transform (10) into

$$(12) \quad \begin{aligned} \|f\|_{B_q^\alpha(L_q(I))} &\approx \|\langle f, \tilde{\phi} \rangle \tilde{\phi}\|_{L_q(I)} \\ &+ \left( \sum_{k \geq 0} \sum_{j \in \mathbb{Z}_k^2} \sum_{\psi \in \Psi} \|\langle f, \tilde{\psi}_{j,k} \rangle \psi_{j,k}\|_{L_p(I)}^q \right)^{1/q} \end{aligned}$$

by using the scaling (5). Any function  $f$  in  $B_q^\alpha(L_q(I))$  satisfying (11) is also in  $L_p(I)$ ; if  $q$  is reduced or  $\alpha$  is reduced, then this is not necessarily the case. In this sense, the spaces  $B_q^\alpha(L_q(I))$  satisfying (11) have minimal smoothness to be embedded in  $L_p(I)$ .

## 5. Wavelets and Image Compression

How can one characterize images by how well they can be compressed using wavelets? The mathematical theory in [5] and [3] poses this question in the following way.

We consider only lossy compression schemes, so we must make precise how we measure error and compression. For simplicity, we measure the error between the original image and the compressed image in the  $L_p(I)$  spaces. We consider approximations  $\tilde{f}$  to  $f$  with wavelet expansions

$$\tilde{f}_N = \tilde{c}_{0,0}\phi_{0,0} + \sum_{k \geq 0} \sum_{j \in \mathbb{Z}_k^2} \sum_{\psi \in \Psi} \tilde{c}_{j,k,\psi} \psi_{j,k}$$

with at most  $N$  nonzero coefficients  $\tilde{c}_{j,k,\psi}$ ; we use  $N$  as the measure of the size of the compressed image  $\tilde{f}_N$ . Leaving the question of algorithms aside for a moment, we consider the error of *best* approximation

$$E_N(f)_p := \inf_{\tilde{f}_N} \|f - \tilde{f}_N\|_{L_p(I)}$$

and ask how that least error decreases as  $N$  increases. It may happen for a particular image that, as  $N$  increases, the error decreases as a negative power of  $N$ :

$$(13) \quad E_N(f)_p \leq CN^{-\alpha/2}.$$

This is obviously equivalent to

$$N^{\alpha/2} E_N(f)_p \leq C \text{ or } \{N^{\alpha/2} E_N(f)_p\}_{N=0}^\infty \in \ell_\infty.$$

One cannot yet characterize functions  $f$  for which (13) holds, but in [5] it is shown for a large variety of wavelets and for many values of  $\alpha$ ,  $p$ , and  $q$  such that (11) holds,

$$(14) \quad \{2^{\alpha k/2} E_{2^k}(f)_p\}_{k=0}^\infty \in \ell_q \iff f \in B_q^\alpha(L_q(I)).$$

The left-hand side of (14) is equivalent to

$$\sum_k [2^{\alpha k/2} E_{2^k}(f)_p]^q < \infty.$$

Thus, there is a precise characterization of images for which one can achieve certain rates of compression using wavelets.

It is pointed out in [3] that if  $f$  is in  $B_q^\alpha(L_q(I))$  such that (11) holds, large classes of algorithms allow one to calculate for any  $f \in B_q^\alpha(L_q(I))$  approximations  $\tilde{f}_N$  that achieve

$$(15) \quad \|f - \tilde{f}_N\|_{L_p(I)} \leq CN^{-\alpha/2} \|f\|_{B_q^\alpha(L_q(I))}.$$

In fact, any algorithm that chooses compressed coefficients  $\tilde{c}_{j,k,\psi}$  that satisfy

$$\|(c_{j,k,\psi} - \tilde{c}_{j,k,\psi})\psi_{j,k}\|_{L_p(I)} \leq \epsilon$$

for some parameter  $\epsilon$ , with a guarantee that  $\tilde{c}_{j,k,\psi} = 0$  if  $c_{j,k,\psi} \leq \epsilon$ , will achieve (15). Such algorithms include progressive transmission of coefficients (in an order that depends on  $p$ ), threshold coding, and scalar or vector quantization. Since  $\|\psi_{j,k}\|_{L_p(I)} \approx 2^{-2k/p}$ , we see that  $\tilde{c}_{j,k,\psi} = 0$  if

$$|c_{j,k,\psi}| 2^{-2k/p} \leq \epsilon \text{ or } |c_{j,k,\psi}| \leq 2^{2k/p} \epsilon.$$

Thus, the threshold depends on the dyadic dilation  $2^k$  and the space  $L_p(I)$  with which one measures the error, but not on the particular space  $B_q^\alpha(L_q(I))$ —it doesn't matter *what* smoothness space the image belongs to, these algorithms give near-optimal rates of approximation.

Because (14) is an *equivalence*, the spaces  $B_q^\alpha(L_q(I))$  *must* arise as soon as one is interested in rates of approximation like (13). For the same reason, images in  $B_q^\alpha(L_q(I))$  have the *least* smoothness to be approximated to order (13) by wavelets.

We now investigate properties of images in  $B_q^\alpha(L_q(I))$  and their applications.

## 6. Pixel Quantization and Image Smoothness

Perhaps the first question that one might ask is how the smoothness of an intensity field  $F$  is reflected and preserved in the pixel data, which we assume to be the average intensity of  $F$  on each of  $2^K \times 2^K$  square pixels in  $I$ .

We can answer this question readily for the special family of biorthogonal wavelets discussed in §3. Since the wavelet duals  $\tilde{\psi}_{j,k}$  are piecewise constant on dyadic subsquares of  $I$  with  $2^{K-k}$  pixels on a side, we can calculate *exactly* the wavelet coefficients  $c_{j,k,\psi}$  of  $F$  for  $k < K$ ; thus, we have the finite-frequency approximation

$$\langle F, \tilde{\phi} \rangle \phi + \sum_{0 \leq k < K} \sum_{j \in \mathbb{Z}_k^2} \sum_{\psi \in \Psi} \langle F, \tilde{\psi}_{j,k} \rangle \psi_{j,k}$$

caused by spatial averaging of the intensity data over pixels. Because of (12), this approximation is at least as smooth as  $F$  in  $B_q^\alpha(L_q(I))$ .

But the intensity is not only averaged over pixels; an extra error is introduced when these averages are rounded to one of  $2^m$  integer values (typically 256 or 4096) that record the *quantized* pixel values. If we assume that the maximum intensity is normalized to be 1, then one can show in the same way as in [3] for the Haar wavelets that the error introduced by this pixel quantization has smoothness in  $B_q^\alpha(L_q(I))$  bounded by  $C2^{\alpha K} 2^{-m}$ . Thus, the smoothness of the quantized approximation to  $F$  is bounded by

$$C(\|F\|_{B_q^\alpha(L_q(I))} + 2^{\alpha K} 2^{-m}).$$

This shows that if one wants to preserve the smoothness of a quantized image, one needs to increase the number of grey scales,  $2^m$ , in a way that depends both on the presumed smoothness  $\alpha$  and the number of pixels in a row,  $2^K$ . This makes perfectly good sense, since one can imagine that increasing the spatial resolution without increasing the number of grey scales will lead to contouring in the image. Typically,  $\alpha$  ranges between .3 and .7 for natural images [3].

## 7. Image Processing and Variational Problems

Image processing algorithms are often couched in terms like the following: One is given some data derived from an image  $f$  (either pixel values, noisy pixel values, integrals along certain lines in tomography, etc.), and one wants to construct an approximation to the image (a compressed image, a noise-reduced image, a reconstructed image, etc.). Often there is not enough data to reconstruct the image exactly (in tomography) or the data is corrupted (by Gaussian noise, say), or one does not want to calculate the image exactly but a good approximation to it that takes fewer parameters, as in image compression. Some approaches to each of these problems can be posed in the following way: Find a smooth image  $g$  whose data approximates the data for  $f$ . Using wavelets and smoothness spaces one can make precise the words “smooth” and “approximates” and the balance between them. We give an overview here; the details can be found in [6].

We assume we have an approximation space  $X$  (e.g. an  $L_p(I)$  space) and a smoothness space  $Y$  (like the Besov spaces  $B_q^\alpha(L_q(I))$  or the Sobolev spaces  $W^m(L_p(I))$ ) and we look for a function  $g$  that approximately minimizes

$$(16) \quad \|f - g\|_X + \lambda \|g\|_Y$$

for a positive constant  $\lambda$ . The first term in this expression measures the approximation error between  $f$  and  $g$ , while the second term measures the smoothness of  $g$ . The parameter  $\lambda$  determines the relative importance of error and smoothness.

In [6] we show that if one takes  $X = L_2(I)$  and  $Y$  to be one of the Sobolev spaces  $W^\alpha(L_2(I))$  of functions with  $\alpha$  derivatives in  $L_2(I)$ , then one can use the equivalence between smoothness norms and sequence norms (10) to recast (16) in terms of sequences of wavelet coefficients, that, in fact completely decouples: one solves a scalar problem for each wavelet coefficient of  $g$  with the corresponding wavelet coefficient of  $f$  as data. In particular, one can approximately minimize (16) by taking a wavelet projection that corresponds to a low-pass filter, with a frequency limit that depends on  $\lambda$ . This acts both to compress an image or to smooth a noisy image.

If one instead takes for  $Y$  the minimally smooth spaces  $B_q^\alpha(L_q(I))$  satisfying (11) with  $p = 2$ , then (16) again decouples, only now the solution is to keep for

$g$  only those wavelet coefficients of  $f$  that are larger in absolute value than a certain threshold  $\epsilon$  that again depends on  $\lambda$ . This is associated with threshold coding in image compression and with “wavelet shrinkage” technique in noise removal [9].

## REFERENCES

1. A. Cohen, I. Daubechies, and P. Vial, *Wavelets on the interval and fast wavelet transforms.*, Appl. Comput. Harmonic Analysis **1** (1993), 54–81.
2. I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Reg. Conf. Series in Appl. Math., Vol. 61, SIAM, Philadelphia, 1992.
3. R. DeVore, B. Jawerth, and B. Lucier, *Image compression through wavelet transform coding*, IEEE Transactions on Information Theory **38** (1992), 719–746.
4. R. DeVore, B. Jawerth, and B. Lucier, *Data compression using wavelets: Error, smoothness, and quantization*, DCC-91, Data Compression Conference (J. A. Storer and J. H. Reif, eds.), IEEE Computer Society Press, Los Alamitos, CA, 1991, pp. 186–195.
5. R. DeVore, B. Jawerth, and V. Popov, *Compression of wavelet decompositions*, Amer. J. Math. **114** (1992), 737–785.
6. R. A. DeVore and B. J. Lucier, *Fast wavelet techniques for near-optimal image processing*, 1992 IEEE Military Communications Conference (October 11–14, 1992, San Diego California), IEEE Communications Society, New York, NY, pp. 1129–1135.
7. R. DeVore and B. J. Lucier, *Wavelets*, Acta Numerica **1** (1992), 1–56.
8. R. DeVore and X. Yu, *Nonlinear  $n$ -widths in Besov spaces*, Approximation Theory VI: Vol. 1, C. K. Chui, L. L. Schumaker, and J. D. Ward, eds., Academic Press, New York, 1989, pp. 203–206.
9. D. Donoho, *Denoising by soft thresholding*, IEEE Trans. Information Theory (to appear).
10. M. Frazier and B. Jawerth, *A discrete transform and decompositions of distribution spaces*, J. of Functional Analysis **93** (1990), 34–170.
11. D.-G. Kim, *Wavelet decompositions and function spaces on the unit cube*, Ph.D. Thesis, Purdue University, August 1994.
12. Y. Meyer, *Ondelettes et Opérateurs*, I. *Ondelettes*, Herman, Paris, 1990. English translation is published by Cambridge Univ. Press.