DATA COMPRESSION USING WAVELETS: ERROR, SMOOTHNESS, AND QUANTIZATION*

Extended Abstract

RONALD A. DeVORE,[1] BJÖRN JAWERTH,[1] AND BRADLEY J. LUCIER[2]

Abstract. Recently, a theory, developed by DeVore, Jawerth, and Popov, of nonlinear approximation by both orthogonal and nonorthogonal wavelets has been applied to problems in surface and image compression by DeVore, Jawerth, and Lucier. This theory relates precisely the norms in which the error is measured, the rate of decay in that error as the compression decreases, and the smoothness of the data. In addition, one can interpret the error incurred by the quantization of wavelet coefficients in terms of this theory. In this talk we give an overview of the previous results, and expand our argument, made earlier for image compression, that frequency-amplitude response curves that arise quite naturally in problems involving human visual and audio perception should be used to decide the quantization strategy for wavelet coefficients and the norm in which to measure the error in compressed data.

1. INTRODUCTION

In this talk we present an overview of recent theoretical results about methods of data compression based on wavelet decompositions. This theory, developed by DeVore, Jawerth, and Popov [5], relates precisely a one-parameter family of norms used to measure the error between the original and compressed data, the rate of decrease in this error as more coefficients are included in the approximation, and the smoothness of images in certain smoothness classes called Besov spaces. We recall how this theory has been applied to surface compression [4] and image compression [3]. The paper on image compression emphasizes how coefficient quantization strategies can be interpreted in this theory, and, alternately, how the theory can be used to choose a quantization strategy. In addition, [3] introduces the notion that the Contrast Sensitivity Threshold (CST) curve [7], which gives the relationship between contrast and frequency at the limits of human perception, should be used to choose the norm in which to measure the error in compressed images and to choose the quantization strategy.

In this talk we explain further how the functional relationship between intensity (or amplitude) and frequency at the limits of human perception in vision or hearing can guide one to choose norms and coefficient quantization strategies to minimize the human *perception* of error. Furthermore, we emphasize that this theory gives a practical way to measure the *smoothness* of functions representing surfaces, images,

[1] Department of Mathematics, University of South Carolina, Columbia, SC 29208
[2] Department of Mathematics, Purdue University, West Lafayette, IN 47907

or sound data, using a notion of error that is capable of measuring the smoothness even of the discontinuous, almost fractal, functions that arise, e.g., as image intensities. Finally, we give an explicit series of steps that can be followed to apply this theory to any situation where the perception of a signal or function depends in a known (and hopefully well-behaved) way on both intensity and frequency.

2. Mathematics

2.1. Wavelet decompositions. A *wavelet decomposition* of a function f defined on $\mathbb{R}^d$ is (typically) an expression of the form

$$(2.1) \qquad f = \sum_{k \geq 0, \ j=(j_1,\ldots,j_d)\in\mathbb{Z}^d} c_{j,k}\phi_{j,k},$$

where the coefficients $c_{j,k}$ depend on f and the functions $\phi_{j,k}(x) := \phi(2^k(x-j/2^k))$ are the dyadic dilates (by 2^k) and translates (by $j/2^k$) of a single function ϕ called a *wavelet*. (Orthogonal expansions in $\mathbb{R}^d$ usually use 2^d-1 functions ϕ.) We consider a wavelet transform to consist of (1) the functions $\phi_{j,k}$ and (2) a particular method of choosing the coefficients $c_{j,k}$. (If the functions $\phi_{j,k}$ are redundant, or linearly dependent, then there may be more than one way to calculate the coefficients $c_{j,k}$.)

When wavelet decompositions are used as a basis for compressing digitized image or sound data, or surface data that is sampled from a model, we must associate a function f defined on $\mathbb{R}^d$ to discrete data that is defined only at the sample points. While one could associate to each data set a function f that is independent of the wavelet transform being applied, it is more amenable to our analysis to allow the representation to depend on the transform; often a discrete transform is applied to the discrete data to get coefficients $c_{j,k}$, and we *define* f by (2.1). We then view the data compression problem as one of approximating f by a second (compressed) function $\tilde{f}$. For a lossless algorithm, the original and compressed functions will be the same and the error between them will be zero. We will generally consider algorithms that introduce differences between the original and compressed data in order to achieve higher compression levels.

Given the wavelet decomposition, which we consider to be determined by the functions $\phi_{j,k}$ and the method of choosing the coefficients $c_{j,k}$, the algorithm calculates quantized coefficients $\tilde{c}_{j,k}$, and the compressed function takes the form

$$(2.2) \qquad \tilde{f} = \sum \tilde{c}_{j,k}\phi_{j,k}.$$

The method of quantizing coefficients involves applying a strategy, which we consider fixed, that depends on one or more parameters, which are allowed to vary. We store or transmit a coded representation of the coefficients $\tilde{c}_{j,k}$, typically through some type of entropy coding.

In the following sections we examine and discuss the mathematical issues treated in the earlier paper [5] together with the applications and extensions to the problems of image and surface compression found in [3], [4].

2.2. Error metrics. Based on the intended application, one must decide how to measure the error between f and $\tilde{f}$, which we assume to be defined on some interval I in $\mathbb{R}^d$. (For example, $d = 1$ for audio data and $d = 2$ for image or surface data.) There are many possible choices of such a metric; we will use (somewhat arbitrarily) the $L^p(I)$ norms with $0 < p \leq \infty$ as error metrics. These norms, defined by

$$\|f - \tilde{f}\|_{L^p(I)} := \left(\int_I |f(x) - \tilde{f}(x)|^p \, dx \right)^{1/p},$$

include as special cases the mean square error

$$\|f - \tilde{f}\|_{L^2(I)} := \left(\int_I |f(x) - \tilde{f}(x)|^2 \, dx \right)^{1/2},$$

the mean absolute error

$$\|f - \tilde{f}\|_{L^1(I)} := \int_I |f(x) - \tilde{f}(x)| \, dx,$$

and the maximum error

$$\|f - \tilde{f}\|_{L^\infty(I)} := \max_{x \in I} |f(x) - \tilde{f}(x)|.$$

The parameter p gives added flexibility, in that the relative sizes of the component functions $c_{j,k}\phi_{j,k}$ with amplitude $c_{j,k}$ and frequency 2^k, given by $\|c_{j,k}\phi_{j,k}\|_{L^p(I)}$, can be changed by varying the parameter p. In other words, varying p allows us to change the relative importance of amplitude and frequency in measuring the size of basic functions.

One error metric will not suffice for all applications. For surface compression it is natural to choose the $L^\infty(I)$ metric, because two parts that are to fit together must be machined to a certain fixed tolerance. For image compression, one hopes to choose an error metric that parallels the human visual system, so that image differences judged to be large by the human eye are mathematically large and image differences which, for whatever reason, are insignificant to the eye will have small size in the error metric. There is a similar design issue for audio data. One important property of both the human visual and auditory systems is that the lower threshold of perceiving a signal depends on both the intensity and the spatial or temporal frequency of that signal. We argue in §3.1 that these intensity-frequency curve at the threshold of human perception can profitably be used to choose an error metric for data compression.

2.3. Algorithm efficiency and smoothness of data. After deciding on a space X whose metric $\| \cdot \|_X$ will be used to measure the error between f and $\tilde{f}$, we address the question of how to measure the efficiency of a given algorithm. Recall that an algorithm depends on three things: the choice of representation functions $\phi_{j,k}$, the method of calculating the coefficients $c_{j,k}$ (which together we call the transform), and the quantization strategy. A given algorithm generates different compressed functions $\tilde{f}$ depending on the parameters of the quantization strategy.

We will compare algorithms on the basis of the error $\|f - \tilde{f}\|_X$ and the number of nonzero quantized coefficients $\tilde{c}_{j,k}$. Suppose that a compression algorithm produces a family $\{\tilde{f}\}$ of approximations corresponding to different parameters in the quantization strategy. We introduce for this algorithm the error function

$$a_N(f)_X := \inf_{\tilde{f} \text{ has } \leq N \text{ coefficients}} \|f - \tilde{f}\|_X.$$

In other words, a_N measures the compression error that is achieved if the number of coefficients in the compressed function does not exceed N.

We ask the following fundamental question: If the $\phi_{j,k}$ are fixed, how smooth are functions that can be approximated to $O(N^{-\alpha/d})$ with $\leq N$ coefficients by some algorithm that uses the functions $\phi_{j,k}$. For the spaces $X = L^p$ and for many classes of representation functions $\phi_{j,k}$, DeVore, Jawerth, and Popov [5] have shown that, roughly speaking,

(2.3)
$$\sigma_N(f)_{L^p(I)} := \inf_{\text{all algorithms using } \phi_{j,k}} a_N(f)_{L^p(I)} = O(N^{-\alpha/d}) \iff f \in B_q^\alpha(L^q(I)),$$

where $q = 1/(\alpha/d + 1/p)$ and the Besov space $B_q^\alpha(L^q(I))$ consists of functions that have α bounded "derivatives" in $L^q(I)$. More precisely,

$$\left(\sum_{N=1}^\infty [N^{\alpha/d} \sigma_N(f)_{L^p(I)}]^q \frac{1}{N} \right)^{1/q} < \infty \iff f \in B_q^\alpha(L^q(I)).$$

In particular, this is true for box splines [1] (with piecewise constant approximations as a special case) when $0 < p < \infty$ and orthogonal wavelets [9], [2] (of which the Haar transform is a special case) when $1 < p < \infty$. (Later, DeVore, Petrushev, and Yu [6] extended the results for box splines to the case $p = \infty$.) So, if we consider N, the number of coefficients in the representation of $\tilde{f}$, to be a measure of the amount of information one must use to represent the compressed data, one can hope to achieve a particular rate of error decay in $L^p(I)$ if and only if f is in a specific Besov smoothness space $B_q^\alpha(L^q(I))$. In proving this theorem, DeVore et al. provide specific algorithms for each set of functions $\phi_{j,k}$ that give the optimal rate of convergence. One should note that for a wide class of representation functions $\phi_{j,k}$, the optimal selection of coefficients for a given α results in the same smoothness classes.

The equivalence (2.3) suggests that membership in Besov spaces $B_q^\alpha(L^q(I))$ is an appropriate way of classifying the smoothness of data, in that we can check the effectiveness of a given compression algorithm by seeing how it performs on functions in $B_q^\alpha(L^q(I))$. However, it is of practical interest to measure smoothness in the spaces $B_q^\alpha(L^q(I))$ only if common surfaces, images, or signals are in these spaces. In [3] and [4] evidence is presented that this is indeed the case for surfaces with various singularities and with images.

2.4. Algorithms for data compression. To say that a function $f \in B_q^\alpha(L^q(I))$ has enough smoothness to be approximated to $O(N^{-\alpha/d})$ in $L^p(I)$ by

functions of the form $\tilde{f}$, does not, in and of itself, explain how to find an algorithm to calculate a particular set of approximations $\tilde{f}$. We briefly describe here a particular algorithm for orthogonal wavelets and indicate how similar algorithms can be constructed for box splines and other wavelets.

We restrict our attention for the moment to one dimension. Given any positive value of r, Daubechies [2] and Mallat [8] have shown how to construct a C^r function ϕ such that $\phi_{j,k}(x) := \phi(2^k x - j)$, the dyadic dilates and translates of ϕ, form a complete orthogonal set in $L^2(I)$. (In $\mathbb{R}^d$, there is a set of $2^d - 1$ such functions that together satisfy this property.) Thus, any function $f \in L^2(I)$ can be written as

$$(2.4) \qquad f = \sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \frac{< f, \phi_{j,k} >}{< \phi_{j,k}, \phi_{j,k} >} \phi_{j,k} = \sum c_{j,k} \phi_{j,k}.$$

It is clear from the orthogonality of $\{\phi_{j,k}\}$ that when approximating f in $L^2(I)$ by $\tilde{f} = \sum \tilde{c}_{j,k} \phi_{j,k}$ with at most $\mathcal{N}$ nonzero terms, then the best choice consists of the $\mathcal{N}$ largest values of $\|c_{j,k} \phi_{j,k}\|_{L^2(I)}$, i.e.,

$$\tilde{c}_{j,k} = \begin{cases} c_{j,k}, & \text{for the } \mathcal{N} \text{ largest values of } \|c_{j,k}\phi_{j,k}\|_{L^2(I)}, \\ 0, & \text{otherwise.} \end{cases}$$

In this way we minimize

$$\|f - \tilde{f}\|_{L^2(I)}^2 = \sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \|(c_{j,k} - \tilde{c}_{j,k})\phi_{j,k}\|_{L^2(I)}^2.$$

This suggests that after fixing a positive ϵ, one can choose *any* numbers $\tilde{c}_{j,k}$ that satisfy

$$(2.5) \qquad \|(c_{j,k} - \tilde{c}_{j,k})\phi_{j,k}\|_{L^2(I)} \leq \epsilon$$

and we will obtain, if not the best approximation according to our criterion of minimizing the error with a fixed number of nonzero $\tilde{c}_{j,k}$, then at least a *good* approximation. In fact, from a practical point of view, this flexibility is an asset, for it allows us to choose values of $\tilde{c}_{j,k}$ that have a small number of significant bits.

The condition (2.5) can be generalized to approximation in $L^p(I)$ in $\mathbb{R}^d$. Under some technical conditions on the wavelets ϕ and the on the range of α, the following theorem can be proved; see [5], [3].

THEOREM 2.1. *Choose a positive integer N and numbers $1 < p < \infty$ and $0 < \alpha$. Let q satisfy $q^{-1} = p^{-1} + \alpha/d$. Write f as (2.4). Choose quantized coefficients $\tilde{c}_{j,k}$ that satisfy*

$$(2.6) \qquad \|(c_{j,k} - \tilde{c}_{j,k})\phi_{j,k}\|_{L^p(I)}^q \leq \frac{1}{N}.$$

(We assume that $\|c_{j,k}\phi_{j,k}\|_{L^p(I)}^q \leq 1/N$ implies $\tilde{c}_{j,k} = 0$.) Our compressed function is

$$\tilde{f} := \sum_k \sum_j \tilde{c}_{j,k} \phi_{j,k}.$$

Then for each $0 < \alpha$ and $0 < p < \infty$ there exist constants C_1 and C_2 such that for all $f \in B_q^\alpha(L^q(I))$:

(1) The number, $\mathcal{N}$, of nonzero coefficients $\tilde{c}_{j,k}$ satisfies

$$(2.7) \qquad \mathcal{N} \leq C_1 N \|f\|_{B_q^\alpha(L^q(I))}^q.$$

(2) The error $f - \tilde{f}$ satisfies

$$(2.8) \qquad \|f - \tilde{f}\|_{L^p(I)} \leq C_2 N^{-\alpha/d} \|f\|_{B_q^\alpha(L^q(I))}^{q/p}$$

and

$$(2.9) \qquad \|f - \tilde{f}\|_{L^p(I)} \leq C_1^{\alpha/d} C_2 \mathcal{N}^{-\alpha/d} \|f\|_{B_q^\alpha(L^q(I))}.$$

The paper by DeVore, Jawerth, and Popov [5] contains extensions of the above theorem to $0 < p < \infty$ in the case of box splines. DeVore, Jawerth, and Lucier [3] examine in some detail integer transforms related to the Haar transform for image compression in $L^p(I)$. DeVore, Petrushev, and Yu [6] give a different way to choose the coefficients $\tilde{c}_{j,k}$ for box splines so that (2.7), (2.8), and (2.9) hold in $L^\infty(I)$. A simpler algorithm for $L^\infty(I)$, but which requires greater than optimal smoothness, is contained in [4]. This short list is meant to emphasize that for any p there is an algorithm that will give near-optimal approximation in $L^p(I)$ using wavelets of some kind; the choice of p, which is investigated in §3, depends solely on the application.

2.5. How to choose quantization levels. To transmit images or other signal data along a communications channel, one sends integer codes that represent the quantized coefficients of the transformed data. Here we discuss how to choose quantization levels based on the criterion introduced in Theorem 2.1; this material was also presented in [3].

Specifically, for some $N > 0$, and a given representation

$$f = \sum c_{j,k} \phi_{j,k},$$

one chooses quantized coefficients $\tilde{c}_{j,k}$ such that

$$\|(c_{j,k} - \tilde{c}_{j,k}) \phi_{j,k}\|_{L^p(I)}^q \leq \frac{1}{N},$$

or, equivalently,

$$(2.10) \qquad \|(c_{j,k} - \tilde{c}_{j,k}) \phi_{j,k}\|_{L^p(I)} \leq \frac{1}{N^{1/q}}.$$

Because $\|\phi_{j,k}\|_{L^p(I)} = 2^{-dk/p} \|\phi\|_{L^p(I)}$ in $\mathbb{R}^d$, (2.10) says that one should require

$$(2.11) \qquad \frac{N^{1/q} \|\phi\|_{L^p(I)}}{2^{1+dk/p}} |c_{j,k} - \tilde{c}_{j,k}| \leq \frac{1}{2}.$$

Inequality (2.11) implies that when approximating f in $L^p(I)$, one should choose a quantization separation for $c_{j,k+1}$ that is $2^{d/p}$ times the quantization separation

for $c_{j,k}$. Therefore, we can take for coefficients $\tilde{c}_{j,k}$ and integer codes $\text{code}_{j,k}$ to represent those coefficients,

$$\text{code}_{j,k} := \text{round}\left(\frac{N^{1/q}\|\phi\|_{L^p(I)}}{2^{1+dk/p}}c_{j,k}\right) \quad \text{and} \quad \tilde{c}_{j,k} = \frac{2^{1+dk/p}}{N^{1/q}\|\phi\|_{L^p(I)}}\text{code}_{j,k}.$$

Thus, in two dimensions, as for images, one should reduce by one the number of bits one sends of $\tilde{c}_{j,k}$ for each level k to approximate f in $L^2(I)$. To approximate in $L^1(I)$, coefficients $\tilde{c}_{j,k+1}$ should have two fewer significant bits than $\tilde{c}_{j,k}$. In one dimension, as for audio data, one should keep one fewer bit per increased level for approximation in $L^1(I)$, and keep one-half fewer bit per level for approximation in $L^2(I)$ (or at least one should increase the quantization interval by approximately $\sqrt{2}$ for each level). In practice, we suggest setting $\tilde{c}_{j,k} = c_{j,k}$ for k less than some fixed level K, and then reducing the number of bits in $\tilde{c}_{j,k}$ for higher k according to the formula (2.11).

3. Applications to Data Compression

Sometimes an error criterion is externally and unambiguously applied; for example, we would like parts designed by a CAD system to fit together well, and so a maximum tolerance in the design may be specified. For this application one would use an $L^\infty(I)$ error criterion. Several applications are given in [4] of wavelet-based compression of surfaces while controlling the maximum error.

For other applications, it is the human perception of the compressed data that is most important. DeVore, Jawerth, and Lucier [3] have suggested how to use certain information about the sensitivity of people to visual stimuli at various frequencies and intensities to choose an error metric for image compression that would approximately model this response. Here we expand on this approach and suggest how it can be applied to other situations, such as compressing audio data.

3.1. How human responses to stimuli can lead to the choice of error metric. In §2.4 we gave a general framework for approximating functions f by finite linear combinations $\tilde{f} = \sum \tilde{c}_{j,k}\phi_{j,k}$. This approximation process is nonlinear because the choice of which $\phi_{j,k}$ to use depends on the function f that will be approximated. The only free parameter in the theory is the parameter p, which determines the error metric. In [3] it is claimed that for image processing one should use the value $p = 1$. That claim, which is made assuming that the middle frequency information of an image will be kept, is based on the precise way that the threshold of human perception of an oscillating pattern depends on intensity and frequency. Thus, we implicitly assume that the purpose of the images is to "look the best" to a human observer. Here we expand upon that specific claim and show how to choose an error metric in any situation where perception depends in a well-behaved way on both intensity and frequency of sensation.

Let us stick with images for a while; at the end we will speculate somewhat on compressing audio signals.

The problem of approximating an image $f = \sum c_{j,k}\phi_{j,k}$ by wavelets can be

thought of as choosing from among the set of "features" $\{c_{j,k}\phi_{j,k}\}$ a finite number $\{\tilde{c}_{j,k}\phi_{j,k}\}$, which, in the case of image compression, should most accurately reflect the visual perception of the image. One quantization strategy that satisfies our theory is simply to order the set $\{\|c_{j,k}\phi_{j,k}\|_{L^p(I)}\}$ and choose the $\mathcal{N}$ features with the largest norm, so that if

$$\|c_{j,k}\phi_{j,k}\|_{L^p(I)} \geq \|c_{l,m}\phi_{l,m}\|_{L^p(I)}$$

then the "feature" $c_{j,k}\phi_{j,k}$ will be chosen before $c_{l,m}\phi_{l,m}$. One would prefer to choose features that the human eye will find most striking, in some sense, and to leave behind those features that the human visual system does not perceive. Thus, one would like to choose $c_{j,k}\phi_{j,k}$ before $c_{l,m}\phi_{l,m}$ if and only if

$$\|c_{j,k}\phi_{j,k}\|_{\text{eye}} \geq \|c_{l,m}\phi_{l,m}\|_{\text{eye}},$$

if such a norm could be found.

Each feature $c_{j,k}\phi_{j,k}$ has an intensity of about $c_{j,k}$ and a characteristic frequency of about 2^k, as can be seen by examining the Fourier transform of $\phi_{j,k}$. It is how people perceive a feature with a certain frequency and contrast that is of primary interest. It is well known that a high frequency pattern superimposed on a grey background will not be discernible from grey, even at high intensity. The threshold of perception of an oscillating pattern of a particular frequency and intensity is known as the Contrast Sensitivity Threshold (CST) curve; see, e.g., [7]. The frequencies to which the human visual system is most sensitive are taken to be the middle frequencies, and people are less able to discern either high frequency or low frequency oscillations from grey. Let us assume that we will keep in our image all mid- and low-frequency information, so our compression problem will be to decide which high-frequency "features" to include in the compressed image. Let us pick a "feature" $c\phi_{j,k}$ with (high) frequency 2^k that is just barely discernible from a grey background. In fact $c\phi_{l,k+1}$ will *not* be discerned from grey; because of the higher frequency, the value of c must be increased to some other value, $\bar{c}$, to be at the threshold of perception. Thus, in our previous imprecise language, we have

$$\|c\phi_{j,k}\|_{\text{eye}} = \|\bar{c}\phi_{l,k+1}\|_{\text{eye}}.$$

The question is, what is the relationship between c and $\bar{c}$, and is there a value of p such that the *same* relationship holds for the $L^p(I)$ norm? In fact, for high frequencies, $\bar{c}/c = 4$, and the only value of p that has the same relationship is $p = 1$. If $\bar{c}/c$ had been 2, then one could have said that the $L^2(I)$ norm would be better for generating compressed images. But that is not the case. In Figure 1 we present two pictures that have identical levels of compression, the same underlying functions $\phi_{j,k}$, the same methods of calculating the coefficients $c_{j,k}$, and that differ only in whether the L^2 or the L^1 criterion was used to choose the compressed coefficients $\tilde{c}_{j,k}$. We feel that this L^1 picture, among many others, looks better than the L^2 picture at the same level of compression.

There is a similar threshold curve for audio signals, although it is complicated by the direction the signal is coming from. For high frequencies there is a similar

FIGURE 1. *The left image is compressed using the $L^2(I)$ quantization strategy; it has 4545 nonzero coefficients $\tilde{c}_{j,k}$ and 3587 bytes in the compressed file, or 0.109 bits per pixel. The right image uses $L^1(I)$ quantization; it has 4514 nonzero coefficients and the same number of bits per pixel. Even though the images have the same compression rate, each strategy chooses different "features" to keep.*

relationship for sounds as for images: if you have a barely perceptible signal with intensity c and frequency 2^k, then one must increase the intensity to another value $\bar{c}$ to hear a signal of frequency 2^{k+1}. The ratio $\bar{c}/c$ will now determine the norm in which to measure the error in our approximation. Because these signals are one-dimensional ($d = 1$), if $\bar{c}/c$ is 2, then we should approximate in $L^1(I)$; if it is about $\sqrt{2}$ then we should approximate in $L^2(I)$.

3.2. The smoothness of data, and the choice of wavelets. After one has decided on a norm $\| \cdot \|_{L^p(I)}$ for measuring the error in compressed data, one should examine "typical" images or sounds to see which of the smoothness classes $B_q^\alpha(L^q(I))$, $1/q = \alpha/d + 1/p$ contain these "typical" signals. Note that q depends not only on α and p, but also on the dimension d of the signal, so that different conclusions are possible for one- and two-dimensional signals. The smoothness of the the data one would like to compress is important, because, typically, one must have the rth derivatives of ϕ bounded for $r = \lfloor \alpha \rfloor$ if Theorem 2.1 is to hold. Thus, the smoothness of the data determines the smoothness needed in the wavelet.

We report in Table 1 the estimated Besov space smoothness of various images compressed using a modified Haar wavelet transform and the $L^1(I)$ coefficient quantization strategy. (See [3] for details.) The images are the green components of color images of "lenna" (in Figure 1), an F-16 flying over mountains, a bridge over a stream in a forest, and an aerial view of an airport and surrounding terrain. If we observe that the error $\|f - \tilde{f}\|_{L^1(I)} \approx C\mathcal{N}^{-\beta}$, then we estimate $\alpha = 2\beta$ and $\|f\|_{B_q^\alpha(L^q(I))} = C$. (The correlation coefficient indicates the goodness of fit on a log-log scale.) The first two images have a Besov space smoothness of $\alpha \approx 0.6$; $\alpha \approx 0.35$

for the other images. Because in all cases $\alpha < 1$, these figures suggest that piecewise constant wavelet approximations achieves the highest *rate* of approximation for image compression in the $L^1(I)$ metric. That the the latter two images have significantly less smoothness than the first two images expresses mathematically what may be concluded on a purely subjective basis simply by looking at them.

TABLE 1. ESTIMATED SMOOTHNESS OF IMAGES.

	Lenna	F-16	Bridge	Airport
Estimated α	0.599	0.597	0.370	0.306
Estimated $\|f\|_{B_q^\alpha(L^q(I))}$	0.677	0.676	0.571	0.516
Correlation Coefficient	-0.999	-0.993	-0.994	-0.992

REFERENCES

[1] C. DE BOOR AND R. A. DEVORE, *Approximation by smooth multivariate splines*, Trans. Amer. Math. Soc., 276 (1983), pp. 775-788.
[2] I. DAUBECHIES, *Orthonormal basis of compactly supported wavelets*, Comm. on Pure and Appl. Math., 41 (1988), pp. 909-996.
[3] R. A. DEVORE, B. JAWERTH, AND B. J. LUCIER, *Image compression through wavelet transform coding*, preprint.
[4] ———, *Surface compression*, preprint.
[5] R. A. DEVORE, B. JAWERTH, AND V. A. POPOV, *Compression of wavelet decompositions*, preprint.
[6] R. A. DEVORE, P. PETRUSHEV, AND X. M. YU, *Nonlinear wavelet approximations in the space $C(\mathbb{R}^n)$*, in Proceedings of the US/USSR Conference on Approximation, Tampa, Springer-Verlag, New York, 1990, to appear.
[7] R. F. HESS AND E. R. HOWELL, *The threshold contrast sensitivity function in strabismic amblyopia: Evidence for a two type classification*, Vision Research, 17 (1977), pp. 1049–1055.
[8] S. MALLAT, *Multiresolution approximation and wavelet orthonormal bases of $L^2(\mathbb{R})$*, Trans. Amer. Math. Soc., 315 (1989), pp. 69–87.
[9] Y. MEYER, *Wavelets and operators*, in Analysis at Urbana I, Proceedings of the Special Year in Modern Analysis at the University of Illinois, 1986–1987, E. R. Berkson, N. T. Peck, and J. Uhl, ed., London Mathematical Society Lecture Notes Series, 137, Cambridge University Press, New York, 1989, pp. 256–365.