Visibility of Wavelet Quantization Noise

Andrew B. Watson, Gloria Y. Yang, Joshua A. Solomon, and John Villasenor, *Member, IEEE*

***Abstract*—The discrete wavelet transform (DWT) decomposes an image into bands that vary in spatial frequency and orientation. It is widely used for image compression. Measures of the visibility of DWT quantization errors are required to achieve optimal compression. Uniform quantization of a single band of coefficients results in an artifact that we call *DWT uniform quantization noise*; it is the sum of a lattice of random amplitude basis functions of the corresponding DWT synthesis filter. We measured visual detection thresholds for samples of DWT uniform quantization noise in Y, Cb, and Cr color channels. The spatial frequency of a wavelet is $r2^{-\lambda}$, where r is display visual resolution in pixels/degree, and λ is the wavelet level. Thresholds increase rapidly with wavelet spatial frequency. Thresholds also increase from Y to Cr to Cb, and with orientation from lowpass to horizontal/vertical to diagonal. We construct a mathematical model for DWT noise detection thresholds that is a function of level, orientation, and display visual resolution. This allows calculation of a "perceptually lossless" quantization matrix for which all errors are in theory below the visual threshold. The model may also be used as the basis for adaptive quantization schemes.**

***Index Terms*—Discrete wavelet transform, image compression, quantization, wavelet.**

I. Introduction

Wavelets form a large class of signal and image transforms, generally characterized by decomposition into a set of self-similar signals that vary in scale and (in two dimensions) orientation [1]. The discrete wavelet transform (DWT) is a particular member of this family that operates on discrete sequences, and which has proven to be an effective tool in image compression [2]–[7]. The DWT is closely related to and in some cases identical to subband codes [8], perfect-reconstruction filterbanks [9], and quadrature mirror filters. In a typical compression application, an image is subjected to a two-dimensional (2-D) DWT whose coefficients are then quantized and entropy coded.

DWT compression is lossy, and depends for its success upon the invisibility of the artifacts. However, in the published literature there are few data [10] and no formulae describing the visibility of DWT artifacts. The purpose of this paper is to provide this information, and to show in a preliminary way how it may be used in the design of wavelet compression systems. In this research we have generally followed earlier work on the discrete cosine transform [11]–[19], with some important differences that will be discussed below.

Manuscript received October 29, 1995; revised October 16, 1996. The work of A. B. Watson was supported by NASA Life and Biomedical Sciences and Applications Division under Grant 199-06-12-39. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Amy R. Reibman.

A. B. Watson and J. A. Solomon are with NASA Ames Research Center, Moffett Field, CA 94035 USA (e-mail: beau@vision.arc.nasa.gov).

G. Y. Yang and J. Villasenor are with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA.

Publisher Item Identifier S 1057-7149(97)03928-6.

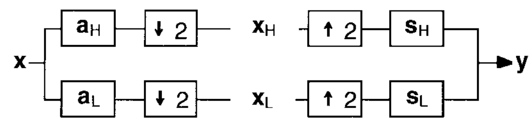


Fig. 1. A two-channel perfect-reconstruction filterbank.

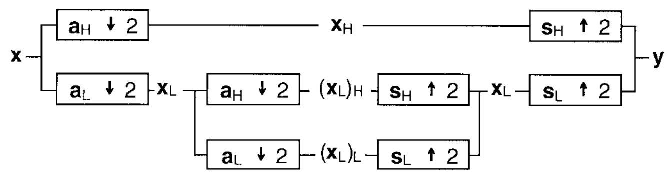


Fig. 2. Two-level 1-D discrete wavelet transform.

II. Background

A. Discrete Wavelet Transform

Fig. 1 illustrates the elements of a one-dimensional, (1-D) two-channel perfect-reconstruction filterbank. The input discrete sequence $\mathbf{x}$ is convolved with highpass and lowpass analysis filters $\mathbf{a}_H$ and $\mathbf{a}_L$, and each result is downsampled by two, yielding the transformed signals $\mathbf{x}_H$ and $\mathbf{x}_L$. The signal is reconstructed through upsampling and convolution with high and low synthesis filters $\mathbf{s}_H$ and $\mathbf{s}_L$. For properly designed filters, the signal $\mathbf{x}$ is reconstructed exactly ($\mathbf{y} = \mathbf{x}$).

A DWT is obtained by further decomposing the lowpass signal $\mathbf{x_L}$ by means of a second identical pair of analysis filters, and, upon reconstruction, synthesis filters, as shown in Fig. 2. This process may be repeated, and the number of such stages defines the *level* of the transform.

With 2-D signals such as images, the DWT is typically applied in a separable fashion to each dimension. This may also be represented as a four-channel perfect reconstruction filterbank, as shown in Fig. 3. Now each filter is 2-D, with the subscript indicating the separable horizontal and vertical components, and the downsampling operation is applied in both dimensions. The resulting four transform components consist of all possible combinations of high- and low-pass filtering in the two dimensions. As in the 1-D case, the process may be repeated a number of times, in each case by applying

TABLE I
COEFFICIENTS OF LINEAR-PHASE 9/7 SYNTHESIS FILTERS $\mathbf{s}_L$ AND $\mathbf{s}_H$ (ORIGIN IS AT INDEX 0, COEFFICIENTS FOR NEGATIVE INDICES FOLLOW BY SYMMETRY).

index	0	1	2	3	4
$\mathbf{s}_L$	0.788486	0.418092	-0.0406894	-0.0645389	
$\mathbf{s}_H$	-0.852699	0.377403	0.110624	-0.0238495	-0.0378285

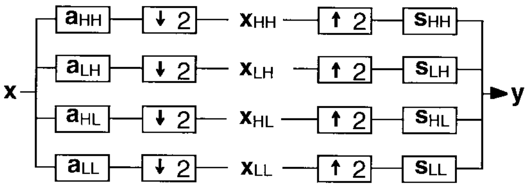


Fig. 3. Four-channel, 2-D perfect-reconstruction filterbank.

the component $\mathbf{x}_{LL}$ as input to a second stage of identical filters.

B. Levels, Orientations, and Bands

Here we adopt the term *level* (λ) to describe the number of 2-D filter stages a component has passed through, and we use the term *orientation* (θ) to identify the four possible combinations of lowpass and highpass filtering the signal has experienced. We index orientations as follows: $\{1, 2, 3, 4\} = \{LL, HL, HH, LH\}$ where low and high are in the order horizontal–vertical. Each combination of level and orientation $\{\lambda, \theta\}$ specifies a single *band*. This terminology is illustrated in Fig. 4.

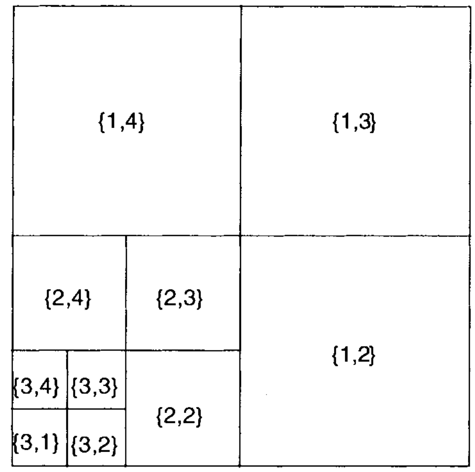


Fig. 4. Indexing of DWT bands. Each band is identified by a level and an orientation $\{\lambda, \theta\}$. This example shows a three level transform.

C. Linear-Phase 9/7 Wavelets

For the purpose of this research, it is necessary to choose a particular DWT, that is, a particular pair of filters $\mathbf{a}_L$ and $\mathbf{a}_H$. We selected the linear-phase 9/7 biorthogonal filters [20]. They were chosen because i) they are symmetrical (linear-phase); ii) they are in wide use [3]; iii) they have been argued to have certain mathematical properties attractive for image compression [5]; and iv) they have been adopted as part of the FBI standard for compression of fingerprint images [2]. Since we shall be dealing primarily with synthesis filters, we give the synthesis coefficients in Table I and show them graphically in Fig. 5. For a perfect-reconstruction filterbank, the synthesis filters may be derived directly from the analysis filters.

D. DWT Quantization Matrix

Compression of the DWT is achieved by quantization and entropy coding of the DWT coefficients. Typically, a uniform quantizer is used, implemented by division by a factor Q and rounding to the nearest integer. The factor Q may differ for different bands. It will be convenient to speak of a *quantization matrix* to refer to a set of quantization factors corresponding to a particular matrix of levels and orientations.

Quantization of a single DWT coefficient in band $\{\lambda, \theta\}$ will generate an artifact in the reconstructed image that is proportional to the impulse response of the corresponding synthesis filter cascade. Examples of impulse responses for two levels and four orientations are shown in Fig. 6. Although they are rendered as images with equal size to emphasize self-similarity, the images in the upper row (level 2) in fact are twice as large (in pixels) in each dimension.

A particular quantization factor Q in one band will result in coefficient errors in that band that are approximately uniformly distributed over the interval $[-Q/2, Q/2]$. The error image will be the sum of a lattice of basis functions with amplitudes proportional to the corresponding coefficient errors. Thus, to predict the visibility of the error due to a particular Q, we must measure the visibility thresholds for individual basis functions and error ensembles.

E. Display Visual Resolution

Visibility of DWT basis functions will depend upon display visual resolution in pixels/degree. Given a viewing distance v in cm and a display resolution d in pixels/cm, the effective display visual resolution (DVR) r in pixels/degree of visual angle is

$$r = dv \tan\left(\frac{\pi}{180}\right) \approx d\frac{v\pi}{180} \approx d\frac{v}{57.3}. \quad (1)$$

A useful mnemonic is that visual resolution is the viewing distance in *pixels (dv)* divided by 57.3. Table II provides some illustrative examples.

F. Wavelet Level, Display Resolution, and Spatial Frequency

We have indexed DWT basis functions by a level λ and an orientation θ. By their nature, wavelet bases of one orientation at different levels are essentially scaled versions of

TABLE II
EXAMPLES OF VISUAL RESOLUTION FOR VARIOUS DISPLAYS. THE HDTV EXAMPLE ASSUMES 1152 ACTIVE LINES AT A VIEWING DISTANCE OF THREE PICTURE HEIGHTS

Display	Resolution (pixels/inch)	Distance (inches)	DVR (pixels/degree)
Computer Display	72	12	15.1
Low quality printing	300	12	62.8
High quality printing	1200	12	251.4
HDTV	48	72	60.3

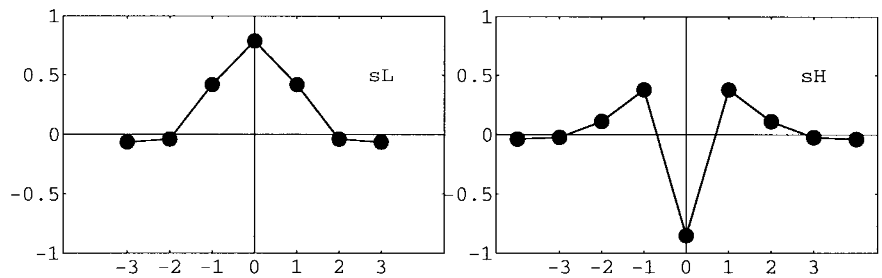


Fig. 5. Linear-phase 9/7 synthesis filters.

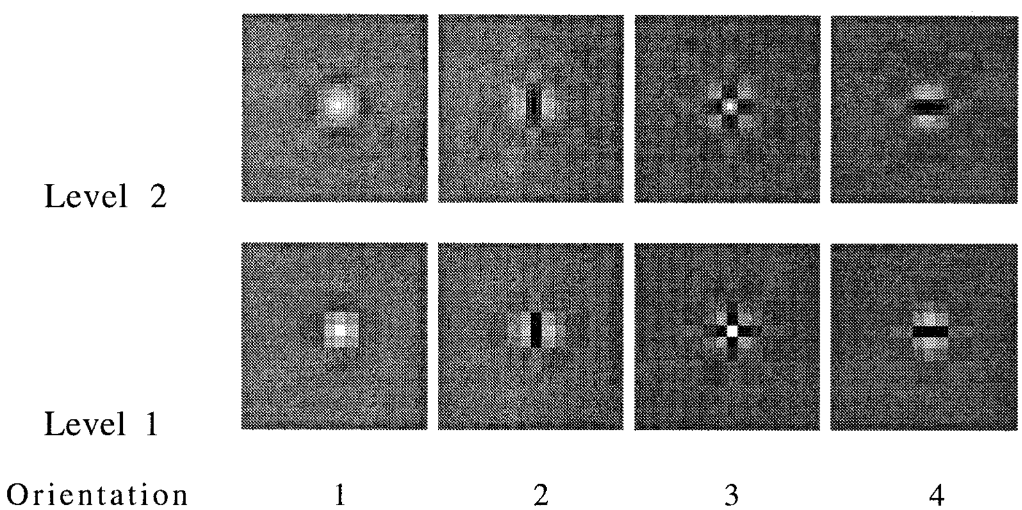


Fig. 6. Linear-phase 9/7 wavelet basis functions at two levels. The images for level 1 are 16 × 16 pixels, those for level 2 are 32 × 32 pixels.

one another (Fig. 6). In terms of the signal that reaches the eye, the magnification of the basis function that results from a move down one level in the transform is equivalent to a decrease by a factor of two in display resolution. A metric that incorporates this equivalence, and that clearly expresses the visual resolution of a given basis function, is spatial frequency expressed in cycles/degree.

A single basis function encompasses a band of spatial frequencies, and at this point it is only necessary that we identify this band in some consistent fashion. The DWT operates essentially by bisecting a frequency band at each level. At the first level of the transform, the selected band extends from the Nyquist frequency, which will be half the display resolution, to half the Nyquist. At the next level, the band will be lower by a factor of two, and so on. Therefore, we will take the Nyquist frequency of the display resolution as the nominal spatial frequency of the first DWT level, and the frequency of each subsequent level will be reduced by a factor of two. Thus for a display resolution of r pixels/degree, the spatial frequency f of level λ will be

$$f = r2^{-\lambda}\,\text{cycles/degree.} \tag{2}$$

G. Gamma Correction

Digital gray-scale images typically contain values that represent so-called gamma-corrected luminance Y' [21]. This is a power function of luminance, with an exponent of around 1/2.3. Likewise, digital color images are typically represented by gamma-corrected R′, G′, and B′, which are similarly power functions of corresponding linear primaries. Color transforms in wide use, such as YC_bC_r, are linear transforms of these nonlinear quantities.

Image compression algorithms typically operate directly on these corrected values, rather than on luminance values themselves. This means that in the particular example of a wavelet transform, the artifact due to quantization of a particular coefficient will be a wavelet basis function in this nonlinear intensity domain. To allow direct predictions, we therefore conducted our experiments in the gamma-corrected domain, using 2.3 as the defining exponent. In our earlier work on the discrete cosine transform (DCT), we chose to estimate visibility of DCT basis functions of luminance, corresponding to a display gamma of one, but that required somewhat indirect predictions of visibility of artifacts in the gamma-corrected domain. Nevertheless for comparison, we also collected one set of thresholds using a display gamma of one. This display gamma was arranged through manipulation of color look-up tables in the computer-display interface [22].

The specific color space we investigate is YCbCr [23], [24]. For simplicity of notation, in the remainder of this paper we will use the Y, C_b, and C_r to designate values in this gamma-corrected color space.

III. Methods

A. Stimuli

Stimuli were modulations of either Y, C_b, or C_r channels of a color image. In each case the two unmodulated channels were set to a constant value of zero. These produce images that are black/white, yellow/purple, and red/green, respectively. All modulations were added to an otherwise uniform ($YC_bC_r = \{128, 0, 0\}$) image of size 1024 × 1024 pixels.

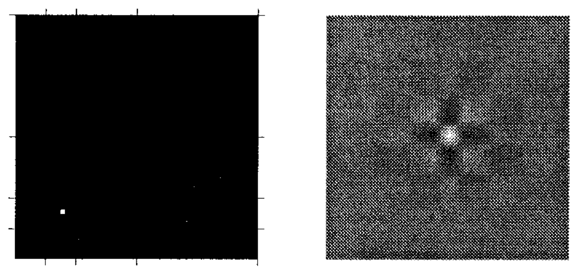
Modulations were either single DWT basis functions or samples of DWT uniform quantization noise. In either case, individual modulation images were scaled to produce amplitudes in the range of [0, 126]. When added to the mean of 128, this yields gray levels ranging from [2–254]. We reserved gray levels 0, 1, and 255 for fixed elements of the display, such as fixation marks. The peak amplitude of the modulated signal is our measure of stimulus intensity. The modulated channel, plus the two remaining unmodulated channels, were then transformed to R′G′B′ using the rule

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.402 \\ 1 & -0.3441 & -0.7141 \\ 1 & 1.772 & 0 \end{bmatrix} \begin{bmatrix} Y \\ C_b - 128 \\ C_r - 128 \end{bmatrix}. \quad (3)$$

Gray-scale stimuli were presented on an Apple 12-in monochrome display (family # M1050, manufactured 3/91) with a resolution of 30.1 pixels/cm, and were viewed from a distance of 121.9 cm, yielding a display visual resolution of 64 pixels/degree. The mean luminance of the display ($Y = 128$) was 14 cd/m^2. The measured gamma was 2.3.

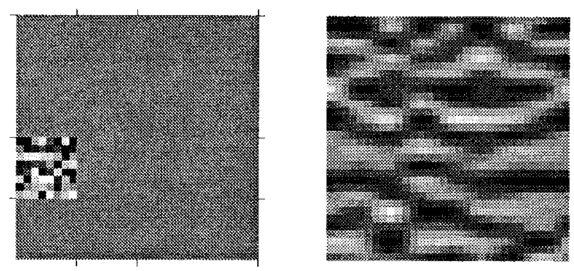
Color stimuli were presented on a Taxan 20-in color monitor (UV 1095, manufactured 2/91) with a resolution of 35.26 pixels/cm, viewed from a distance of 104 cm, for a display visual resolution of 64 pixels/degree. The mean luminance of the display (R′G′B′ = $\{128, 128, 128\}$) was 17.3 cd/m^2. The measured gamma of the monitor was 2.31, and the maximum luminance was 87.5 cd/m^2. The CIE Yxy chromaticities of the three color guns were R = $\{26.8, 0.613, 0.353\}$, G = $\{55.2, 0.281, 0.605\}$, B = $\{6.52, 0.143, 0.058\}$.

To vary the display visual resolution we pixel-replicated the stimuli by factors of one (no replication), two, or four in both dimensions, yielding effective visual resolutions of 64, 32, and 16 pixels/degree. For all stimuli, the duration was 16 frames in duration at a frame rate of 60 Hz, or 267 ms. The time course was a Gaussian $\exp\left[-\pi(1-t/8)^2\right]$ where t is in frames.



(a) (b)

Fig. 7. Construction of DWT basis function stimulus. (a) Three-level DWT, with the band levels separated by tick marks and progressing in order from top to bottom and right to left (see Fig. 4). A single coefficient in band $\{3, 3\}$ is set to one, the rest is set to zero. (b) Inverse DWT of the transform in A. This is a basis function for band $\{3, 3\}$. Image size is 64 × 64.



(a) (b)

Fig. 8. Construction of a DWT noise image. (a) Two-level DWT with band $\{2, 4\}$ filled with noise. (b) Inverse DWT of the transform in A. Image size is 32 × 32.

1) DWT Basis Function Stimuli: We created images of basis functions by setting to the value one a single coefficient in band $\{\lambda, \theta\}$ in an otherwise zero DWT, and computing the inverse DWT. Image width was the smallest power of two large enough to accommodate the support of the basis function, which is equal to $2^{(\lambda+3)}$. An example for band $\{3, 3\}$ is shown in Fig. 7.

2) DWT Uniform Quantization Noise Stimuli: Samples of DWT uniform quantization noise were produced by filling one band of an otherwise zero DWT with samples drawn uniformly from an interval [−1, 1], and inverse transforming the result. The image size was selected for each level in such a way that the size of the filled band was always 8 × 8. For level λ, this means that image width was $2^{(\lambda+3)}$. An example is shown in Fig. 8.

B. Threshold Measurement

In the gray-scale experiments, stimulus amplitude was controlled by means of look-up tables [22], [25]. This allowed display of signals with amplitudes less than one. For color experiments (C_b and C_r) stimuli of various amplitudes were computed in advance as digital movies, and thus limited to integer amplitudes.

To measure detection thresholds for individual stimuli, we used a two-alternative forced-choice (2AFC) procedure. Each trial consisted of two 267-ms time intervals, one containing a uniform gray screen with luminance Y_m, and one containing the stimulus added to the uniform gray screen. A pause of 534 ms separated the two intervals, which were marked by audible warning tones. Following the presentation, the observer selected the interval that appeared to contain the stimulus. From trial to trial, the amplitude of the stimulus was varied adaptively using a Quest staircase [26]. Following 32 trials, a Weibull function was fit to the proportion correct expressed as a function of log amplitude, and threshold was estimated as the amplitude yielding 82% correct [27].

A small gray cross (3×3 pixels, $Y = 96$) at the center of the screen served as a fixation point and aid to accommodation. The cross was extinguished during each stimulus presentation, but remained on between the two intervals of the trial and between trials.

Three observers took part in the experiments. Observer gyy was a 23-year-old female, sfl was a 21-year-old male, and abw was a 43-year-old male. Observers gyy and abw were corrected myopes, sfl was emmetropic. Viewing was binocular with natural pupils in an otherwise darkened room.

IV. Results

We begin our discussion of results with an examination of gray-scale data for two observers (gyy and sfl) for two different stimuli (basis functions and noise patterns) at two different display gammas (1 and 2.3). This will reveal some basic patterns in the data, as well as differences due to stimulus type and display gamma. We then demonstrate from a subset of the data that DWT level has little effect upon visual thresholds, once the effect of spatial frequency has been factored out. We next compare basis function and noise thresholds, and show how one may be predicted systematically from the other.

Following these analyses, we will consider only thresholds for noise patterns collected with a display gamma of 2.3. The gray-scale data are fit with a mathematical model, and thresholds for color wavelets are presented and fit by the same model.

A. Gray-Scale Results

Fig. 9 shows gray-scale thresholds for various DWT signals and observers as a function of spatial frequency and orientation. In these and all subsequent figures, thresholds are expressed as the peak amplitude of the signal, in units of digital levels, with an implicit range of 2–254 between darkest and brightest levels. Because the signals are superimposed on a background of 128, the largest possible amplitude is 126. The first panel shows luminance amplitude thresholds, obtained with a display gamma of one, for single basis functions. They show a rapid ascent at higher frequencies, and also show an effect of orientation: Highest thresholds are for orientation 3 (obliques), lowest are for orientation 1 (lowpass), and intermediate thresholds are obtained for orientations 2 and 4 (horizontal and vertical). The second panel shows comparable data for a display gamma of 2.3. Thresholds are generally lower, but the pattern of results is similar. The third and fourth panels show thresholds for two observers for noise

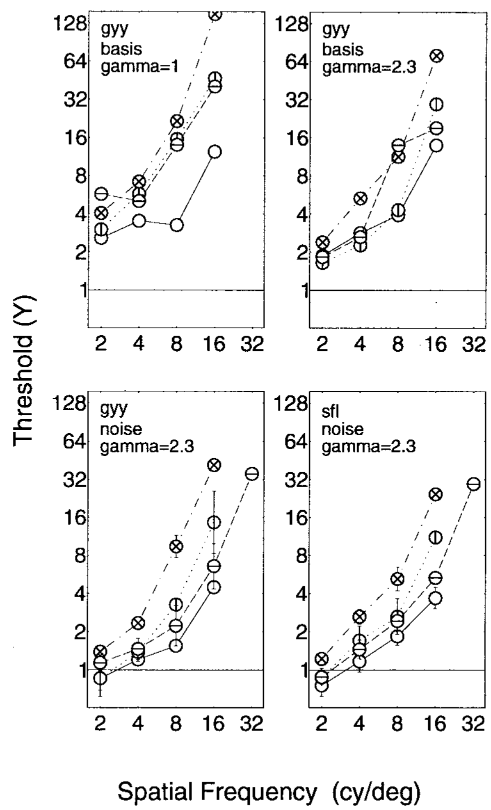


Fig. 9. Thresholds for DWT signals. Orientations are indicated by the line within each symbol. Text in each panel indicates observer, stimulus, and gamma. Error bars of plus and minus one standard deviation are included where repeated measures were taken.

images at a display gamma of 2.3. The pattern is again similar, with a further general reduction in thresholds. For the reasons outlined in the introduction, we have focused upon the case of display gamma = 2.3. Accordingly, all subsequent analyses and discussions refer only to this case.

B. Effect of DWT Level

To verify that DWT level or display visual resolution *per se* have little effect upon visual thresholds when spatial frequency f is held constant, we have collected thresholds for noise images at three display resolutions. Display resolution was varied by pixel replication of 1, 2, or 4 in each dimension from a basic value of 64 pixels/degree, yielding effective visual resolutions of 64, 32, and 16 pixels/degree. Due to nonlinearities between horizontally adjacent pixels in typical monitors [28], we only used an orientation of 4 (vertical modulation) at 64 pixels/degree.

Thresholds for display resolutions of 16, 32, and 64 pixels/degree, all at orientation 4, are shown in Fig. 10. Where multiple measurements have been made, error bars are shown. Though there is some indication of a small systematic difference for gyy between 32 and 64 pixels/degree, in general thresholds are largely unaffected by resolution, once they are expressed as a function of spatial frequency in cycles/degree.

Fig. 11 shows additional data for observer sfl at 16 and 32 pixels/degree. There is again little evidence of any substantial

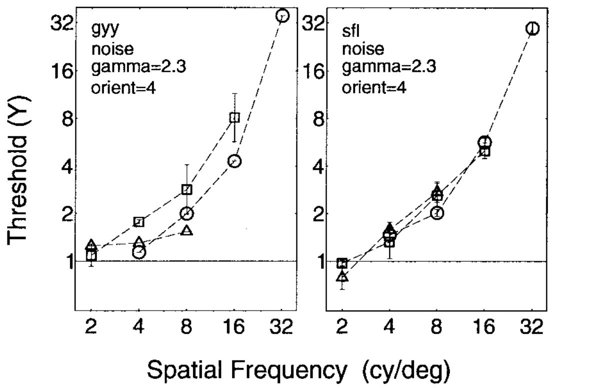


Fig. 10. Thresholds at display resolutions of 16 (triangles), 32 (squares) and 64 pixels/degree (circles), for orientation 4.

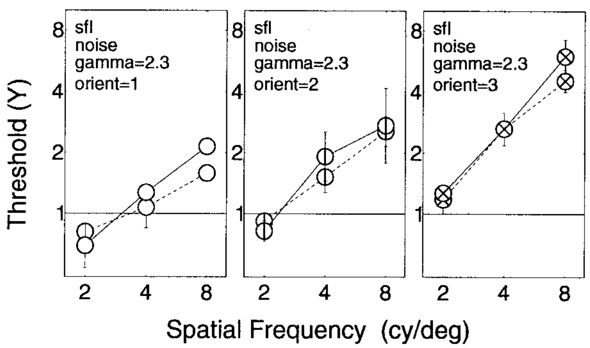


Fig. 11. Thresholds at display resolutions of 16 (dashed line) and 32 pixels/degree (solid line), for orientations 1, 2, and 3.

effect of resolution *per se*, once the thresholds are plotted as a function of spatial frequency in cycles/degree.

C. Single Basis Functions versus Noise Images

Fig. 12 plots the difference between log thresholds for single basis functions and for noise images, taken from the second and third panels of Fig. 9. As expected, basis function thresholds are uniformly higher than noise thresholds.

We considered a simple spatial probability summation model to account quantitatively for the difference between basis function and noise thresholds [27], [29]. In this context, this model asserts that the Minkowski sum over individual basis functions amplitudes is equal for all basis functions ensembles at threshold. In particular, if threshold for a single basis function is Y_{basis}, and if Y_i are the amplitudes of the basis functions that make up the threshold noise stimulus, then

$$Y_{basis} = \left[\sum_{i=1}^{N} |Y_i|^{\beta}\right]^{1/\beta}. \tag{4}$$

For detection of simple contrast stimuli, an exponent β of about 3–4 is typically observed. The threshold contrast Y_{basis} is measured directly, but the set of amplitudes Y_i must be derived from the threshold amplitude of the noise stimulus Y_{noise}. Let A_{basis} be the amplitude of the basis function that results from a unit DWT coefficient $D = 1$. Then in general, $Y = AD$. A particular set of random coefficients $\{D_i\}$ will result in a noise waveform with amplitude A_{noise}. Thus, if the

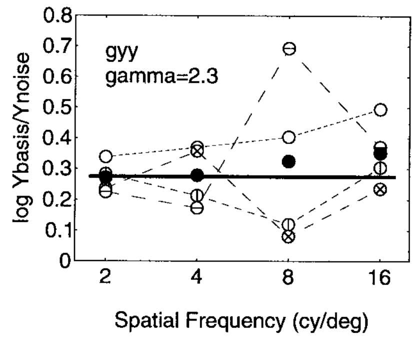


Fig. 12. Difference between log thresholds for DWT noise and basis functions. Open symbols show data for individual orientations, solid symbols are the means. The heavy line is the prediction from probability summation.

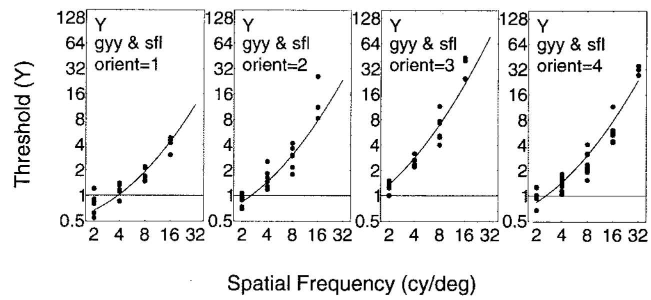


Fig. 13. Fit of the threshold model to grayscale data of observers gyy and sfl.

noise amplitude at threshold is Y_{noise}, the corresponding coefficients are $\{D_i\}Y_{noise}/A_{noise}$. The individual basis function amplitudes are then

$$Y_i = \frac{A_{basis}}{A_{noise}} Y_{noise} D_i. \tag{5}$$

Combining (4) and (5), we find

$$\frac{Y_{basis}}{Y_{noise}} = \frac{A_{basis}}{A_{noise}} \left[\sum_{i=1}^{N} |D_i|^{\beta}\right]^{1/\beta}. \tag{6}$$

The ratio A_{basis}/A_{noise} is close to one for all noise stimuli (almost always slightly less than one), which makes sense since the random numbers were drawn from a uniform distribution over $\{-1, 1\}$. In log units, it averages -0.0205.

To compute the second term in this prediction we first note that exactly the same random samples $\{D_i\}$ were used for each noise stimulus. With $\beta = 4$, this term equals 0.295 log units, for a combined prediction of 0.2745 log units, independent of resolution or orientation. This value is plotted as the horizontal line in Fig. 12. It is clear that probability summation provides an excellent account of the difference between basis and noise thresholds.

This is a useful observation since it provides a way to predict thresholds for individual basis functions from uniform noise thresholds. These may then be used to predict visibility of noise produced by nonuniform quantization [2].

D. Gray-Scale Model

We have experimented with various models to express the threshold for gray-scale DWT noise as a function of spatial

TABLE III
PARAMETERS FOR DWT THRESHOLD MODEL FOR THE Y CHANNEL

Color	Observer	rms	a	k	f_0	g_1	g_3
Y	gyy & sfl	0.134	0.495	0.466	0.401	1.501	0.534

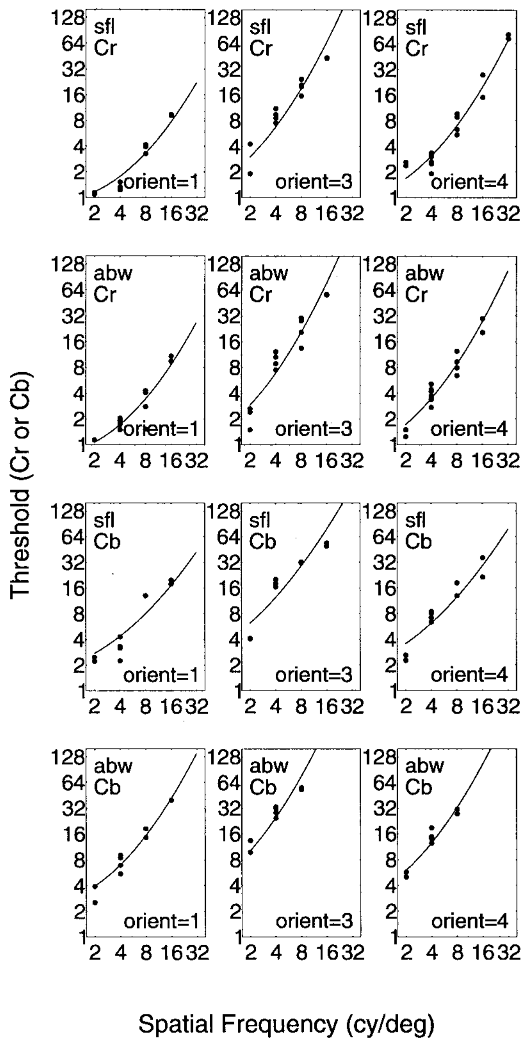


Fig. 14. Thresholds for DWT uniform noise in C_r and C_b channels.

frequency and orientation. Each model was fit to all the gray-scale (Y) noise data for observers gyy and sfl (a total of 103 thresholds). Parameters were optimized with respect to the summed squared error in log Y. One model that provides a reasonable fit is

$$\log Y = \log a + k(\log f - \log g_\theta f_0)^2. \tag{7}$$

This is a parabola in log Y versus log f coordinates, with a minimum at $g_\theta f_0$ and a width of k^{-2} . The term g_θ shifts the minimum by an amount that is a function of orientation, and where $g_2 = g_4 = 1$. The term a defines the minimum threshold. The optimized parameters and rms error (of log Y) are given in Table III. The fit is shown in Fig. 13.

E. Color Results

Fig. 14 shows results for observers sfl and abw at orientations 1, 3, and 4. We did not collect data for orientation 2 (horizontal) because the gray-scale data suggest that it largely duplicates the results from orientation 4 (vertical), and because horizontal modulations are more subject to display limitations. Data at 2, 4, and 8 cycles/degree were collected with a zoom of 4, at 16, zoom = 2, and at 32, zoom = 1. For all measurements, stimulus was a DWT noise pattern, and display gamma = 2.3.

The effects of spatial frequency and orientation are similar to those evident in the gray-scale data. However, there is a general elevation of all thresholds, by about a factor of two for C_r thresholds and about a factor of four for C_b thresholds. Observer abw is also somewhat less sensitive than observer sfl.

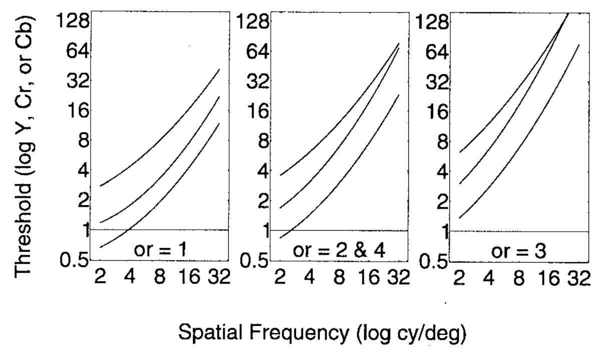


Fig. 15. Model predictions for Y (bottom curve), C_r (middle), and C_b (top) at each orientation for observer sfl.

TABLE IV
PARAMETERS FOR DWT YC_bC_r THRESHOLD MODEL

Color	Observer	rms	a	k	f_0	g_1	g_3
Y	gyy & sfl	0.134	0.495	0.466	0.401	1.501	0.534
Cr	sfl	0.113	0.944	0.521	0.404	1.868	0.516
	abw	0.127	0.803	0.539	0.328	2.017	0.589
Cb	sfl	0.145	1.633	0.353	0.209	1.520	0.502
	abw	0.093	2.432	0.520	0.269	1.706	0.599

F. Color Model

We have applied the same model used for gray-scale thresholds to the color thresholds in Fig. 14. We have fit the data of each color channel separately. Also, because they clearly differed in sensitivity, we have fit separately the data of the two observers. The solid curves in Fig. 14 show the various fits. The parameters are in Table IV, along with the Y parameters from Table III.

To illustrate the differences between the model thresholds for the three color channels, we plot them together in Fig. 15. In this figure we have used C_r and C_b parameters from sfl, who is considerably more sensitive than abw. The Ycurve is generally about a factor of two below the C_r curve, which is in turn about a factor of two below the C_b curve, although this difference declines at higher spatial frequencies, because the C_b curve is somewhat broader than Yor C_r. This broadening is likely due to the intrusion of a luminance detecting channel at high frequencies and high contrasts. The C_b wavelets do

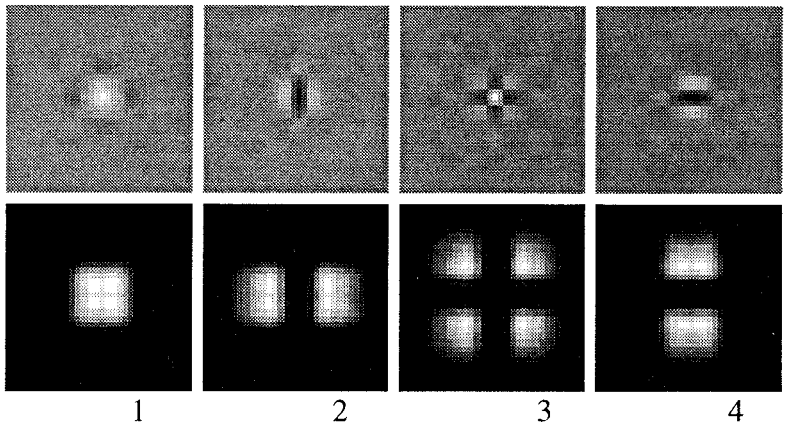


Fig. 16. Wavelets and their Fourier spectra at four orientations.

have a luminance component because the C_b color axis is not orthogonal to the human luminance axis.

G. Theoretical Account of Model Parameters

Although our interest in the threshold model is primarily a practical one, we offer the following general explanation of the estimated model parameters in Table IV. First, we note that although the model is a parabola, all data collected lie to one side of the parabola. This monotonic ascent of thresholds with spatial frequency is consistent with two factors: i) the decline of contrast sensitivity with increasing spatial frequency [30], and ii) the decreasing size of our noise stimuli with increasing spatial frequency. For the Y data, the parabola minimum is at about 0.4 cycles/degree. This is similar to estimates obtained for Gabor functions of fixed log bandwidth, which also decline in size with spatial frequency [31].

The effects of orientation, manifest in the parameters g_1 and g_3, can be understood as follows. Fig. 16 depicts the wavelets and their Fourier spectra at the four orientations. The parameters g_1 and g_3 describe the thresholds for orientations 1 and 3 as frequency shifts relative to threshold for orientations 2 and 4. From the nature of dyadic wavelets, orientation 1 has a spectrum that is approximately a factor of two lower in spatial frequency than orientations 2 or 4. This would suggest a factor of $g_1 = 2$. However, at orientation 1 the signal energy is spread over all orientations, which we know to be less visually efficient than to concentrate them at a narrow range, as in the spectra for orientations 2 and 4. Thus, we expect a slight increase in threshold, which can be mimicked by a slight reduction in f_0. Thus, the final prediction is slightly less than 2, which is what we obtain.

For orientation 3, two similar effects are at work. First, because of the Cartesian splitting of the spectrum, the spatial frequency is about $\sqrt{2}$ above that of orientations 2 and 4, yielding a prediction of $g_3 = 2^{-1/2}$. But here again the spectrum is distributed over two orthogonal orientations (45° and 135°), which should result in a log threshold increase of about $2^{1/4}$ (a shift of $2^{-1/4}$) or a total prediction of $g_3 = 2^{-3/4} = 0.59$, which is indeed just above what is obtained. A third effect, the well-known oblique effect [32], may contribute the final small amount of threshold elevation.

V. Quantization Matrices

We now use the model developed above to compute quantization matrices for the linear-phase 9/7 DWT. The basic idea is to construct a "perceptually lossless" quantization matrix by using a quantization factor for each level and orientation that will result in a quantization error that is just at the threshold of visibility. Although the actual visibility of quantization errors will depend upon the combined visibility of errors across the various bands, this combination is typically quite inefficient, the ensemble threshold is likely to be only slightly lower than for any one band alone (see analogous arguments in Section IV-C with respect to combination of errors across space). For uniform quantization and a given quantization factor Q, the largest possible coefficient error is $Q/2$. The amplitude of the resulting noise is approximately $A_{\lambda,\theta}Q/2$. (The approximation is because the amplitudes of individual noise signals depend upon the particular noise sample, but as noted above they are typically very close to the basis function amplitudes.) Thus, we set

$$Q_{\lambda,\theta} = 2\frac{Y_{\lambda,\theta}}{A_{\lambda,\theta}}. \tag{8}$$

The basis function amplitudes $A_{\lambda,\theta}$ are given for six levels in Table V. Image compression applications do not typically require more than this many levels, but additional amplitudes may be approximated by noting that the ratio of magnitudes of adjacent levels converges to two.

Combining (7), (8), and (2),

$$Q_{\lambda,\theta} = \frac{2}{A_{\lambda,\theta}}\, a 10^{k(\log(2^{\lambda} f_0 g_\theta / r))^2}. \tag{9}$$

Table VI shows example matrices computed from this formula.

Fig. 17 shows an example image uncompressed and compressed using the quantization matrix of Table VI, and twice that matrix. Viewed from the appropriate distance (23 inches, approximately arm's length), the quantization artifacts should be invisible for the left image, and visible for the right (examine the boundaries between each parrot and background). Using typical entropy coding techniques, the resulting bit rates for these two examples are 1.05 and 0.67 b/pixel.

TABLE V
BASIS FUNCTION AMPLITUDES $A_{\lambda,\theta}$ FOR A SIX-LEVEL LINEAR-PHASE 9/7 DWT

Orientation	Level 1	2	3	4	5	6
1	0.62171	0.34537	0.18004	0.091401	0.045943	0.023013
2	0.67234	0.41317	0.22727	0.11792	0.059758	0.030018
3	0.72709	0.49428	0.28688	0.15214	0.077727	0.039156
4	0.67234	0.41317	0.22727	0.11792	0.059758	0.030018

The quantization matrix is inevitably a function of the display visual resolution, as is evident from (9). Fig. 18 shows Y quantization factors for display visual resolutions of 16, 32, and 64 pixels/degree. These figures show that for low visual resolution (16 pixels/degree), the quantization factors are small and almost invariant with level and orientation. At the middle resolution, typical of office viewing of desktop computer images, the function is still a rather flat function of level for all orientations except 3, which shows a large elevation at the lowest level. At the highest visual resolution, oblique, horizontal, and vertical factors are strong functions of level, while the lowpass signal is still nearly invariant with level.

VI. DISCUSSION AND EXTENSIONS

A. Downsampled Chromatic Channels

Because human sensitivity to chromatic variation is lower than that to luminance variation at higher spatial frequencies, it is common in DCT and DWT transform coding to downsample the chromatic channels. This is easily accommodated in the current scheme, provided that the value of r is altered appropriately for the calculation of quantization matrices via (9). For example, if the true display visual resolution is 32 pixels/degree, and chroma is downsampled by two in each dimension, then the corrected value of r is 16 pixels/degree.

B. Light Adaptation

The model developed above is fit to data collected at one mean luminance. The Y, C_b, and C_r thresholds that we have measured and computed are expressed in gray-scale units, analogous to luminance and chromatic contrasts. Contrast thresholds for both luminance and color wavelets are likely to vary little with increasing mean luminance [33]. Thus, matrices computed by the formulae presented here should be valid over a wide range of display luminances, since variation in overall display luminance will alter in proportion both the signal luminance and the mean luminance, thus preserving signal contrast. However, for a fixed display luminance, spatial variations in the local mean luminance of the image will produce local variations in visual thresholds [10]. At photopic levels, thresholds will be roughly proportional to the local mean. These variations can be accommodated by more complex quantization matrix designs [34], and may also drive spatially adaptive quantization schemes [35].

C. Masking and Adaptive Quantization

The thresholds measured above were for signals presented against an otherwise uniform background. It is well known

TABLE VI
QUANTIZATION FACTORS FOR FOUR-LEVEL 9/7 DWT FOR $r = 32$ PIXEL/DEGREE

Color	Orientation	Level 1	2	3	4
Y	1	14.05	11.11	11.36	14.5
	2	23.03	14.68	12.71	14.16
	3	58.76	28.41	19.54	17.86
	4	23.03	14.69	12.71	14.16
Cb	1	55.25	46.56	48.45	59.99
	2	86.79	60.48	54.57	60.48
	3	215.84	117.45	86.74	81.23
	4	86.79	60.48	54.57	60.48
Cr	1	25.04	19.28	19.67	25.6
	2	60.02	34.34	27.28	28.5
	3	184.64	77.57	47.44	39.47
	4	60.02	34.34	27.28	28.5

that thresholds rise when targets are presented against complex backgrounds as a result of visual masking. It is for this reason that wavelet quantization schemes often set quantization factors according to the variance of the coefficients.

A thorough treatment of masking in the context of DWT artifacts is beyond the scope of this paper, but we describe here a simple way in which the threshold model may be used to augment adaptive quantization schemes. One possibility is to compute a measure of variance within a band that is scaled by the visibility of signal within that band

$$\tilde{\sigma}^2_{\lambda,\theta} = D^{-2}_{\lambda,\theta}\sigma^2_{\lambda,\theta} \tag{10}$$

where $D_{\lambda,\theta}$, is the visual threshold for a particular level and orientation, expressed in units of the DWT coefficient. This visually effective band variance might then be used to adjust the band quantization factors, for example

$$\begin{aligned}\tilde{Q}_{\lambda,\theta} &= 2D_{\lambda,\theta}\sqrt{1+\tilde{\sigma}^2_{\lambda,\theta}}\\ &= 2\sqrt{D^2_{\lambda,\theta}+\sigma^2_{\lambda,\theta}}\,.\end{aligned} \tag{11}$$

Recent models of visual masking suggest that the visually effective variance should be computed over a broad range (perhaps all) orientations, but over only a limited range of space and spatial frequency [36]–[38]. The expressions above are easily altered to accommodate this idea, but herein we have not contemplated quantization factors that differ over space. While this might be valuable, it presents additional problems in conveying the side information necessary to define the various matrices, and to associate the various matrices with regions of the image.

Another possible use of the coefficient thresholds is in the context of a highly adaptive scheme such as that designed for the DCT by Watson [18], [34], [39]. In that method, the visibility of the total ensemble of actual quantization errors is computed, based on a mathematical model of DCT uniform quantization noise thresholds, and the quantization matrix is

Fig. 17. Original image (top) and compressed with perceptually lossless DWT quantization matrix (bottom left) and twice that matrix (bottom right). Image dimensions are 256 × 256 pixels. Quantization matrix is designed for a viewing distance of 23 in.

optimized to produce minimum perceptual error for a given bit rate.

D. Other Wavelets

It is desirable to extend our model to thresholds for other wavelets. This requires either empirical thresholds for the wavelet in question, or a more general model of human visual sensitivity. We and others are making efforts in the latter direction [40], [41].

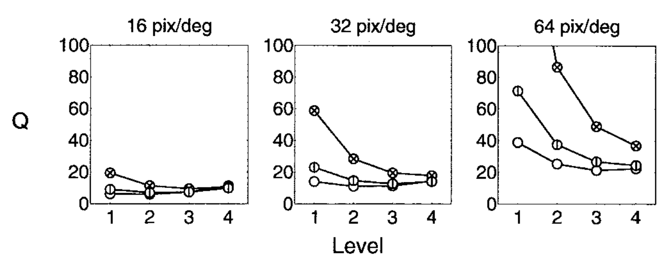


Fig. 18. Quantization matrices for three display visual resolutions plotted as functions of level, with orientation indicated by symbol markings.

VII. Conclusions

We have measured visual thresholds for samples of uniform quantization noise of a DWT based on the linear-phase 9/7 wavelet. Thresholds were collected for gamma-corrected signals in the three channels of the YC_bC_r color space. We have constructed a mathematical model for the thresholds, which may be used to design a simple "perceptually lossless" quantization matrix, or which may be used to weight quantization errors or masking backgrounds in more elaborate adaptive quantization schemes. These perceptual data, models, and methods may enhance the performance of wavelet compression schemes.

ACKNOWLEDGMENT

The authors thank A. J. Ahumada and H. A. Peterson for useful discussions.

REFERENCES

[1] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 2091–2110, 1989.

[2] T. Hopper, C. Brislawn, and J. Bradley, "WSQ grey-scale fingerprint image compression specification, Version 2.0," Criminal Justice Inform. Services, Fed. Bureau Investigation, Washington, DC, Feb. 1993.

[3] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using the wavelet transform," *IEEE Trans. Image Processing,* vol. 1, pp. 205–220, 1992.

[4] J. M. Shapiro, "An embedded wavelet hierarchical image coder," in *Proc. ICASSP,* 1992.

[5] J. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for efficient image compression," *IEEE Trans. Image Processing,* vol. 4, pp. 1053–1060, 1995.

[6] A. B. Watson, "Efficiency of an image code based on human vision," *J. Opt. Soc. Amer. A,* vol. 4, pp. 2401–2417, 1987.

[7] A. B. Watson and A. J. Ahumada, Jr., "A hexagonal orthogonal oriented pyramid as a model of image representation in visual cortex," *IEEE Trans. Biomed. Eng.,* vol. 36, pp. 97–106, 1989.

[8] J. W. Woods, *Subband Image Coding.* Norwell, MA: Kluwer, 1991.

[9] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Trans. Signal Processing,* vol. 40, pp. 2207–2232, 1992.

[10] R. J. Safranek and J. D. Johnston, "A perceptually tuned subband image coder with image-dependent quantization and post-quantization data-compression," in *Proc. ICASSP,* 1989, pp. 1945–1948.

[11] A. J. Ahumada, Jr. and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," in *Proc. SPIE, Human Vision, Visual Processing, and Digital Display III,* vol. 1666, B. E. Rogowitz, Ed., 1992, pp. 365–374.

[12] H. Lohscheller, "Subjectively adapted image communication system," *IEEE Trans. Commun.,* vol. COMM-32, pp. 1316–1322, 1984.

[13] N. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Commun.,* vol. COMM-33, pp. 551–557, 1985.

[14] H. Peterson, A. J. Ahumada, Jr., and A. Watson, "An improved detection model for DCT coefficient quantization," in *Proc. SPIE,* vol. 1913, pp. 191–201, 1993.

[15] ——, "Visibility of DCT quantization noise: Spatial frequency summation," in *Proc. Soc. Inform. Display Int. Symp. Digest Tech. Papers,* Santa Ana, CA, 1994.

[16] ——, "The visibility of DCT quantization noise," in *Proc. Soc. Inform. Display Int. Symp. Digest Tech. Papers,* 1993, vol. 24, pp. 942–945.

[17] J. A. Solomon, A. B. Watson, and A. J. Ahumada, Jr., "Visibility of DCT quantization noise: Contrast masking," in *Proc. Soc. Inform. Display Int. Symp. Digest Tech. Papers,* Santa Ana, CA, 1994.

[18] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE*, vol. 1913, pp. 202–216.

[19] A. B. Watson, J. A. Solomon, and A. J. Ahumada, Jr., "The visibility of DCT basis functions: Effects of display resolution," in *Proc. Data Compression Conf.,* J. A. Storer and M. Cohn, Eds. Los Alamitos, CA: IEEE Computer Society Press, 1994, pp. 371–379.

[20] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math.,* vol. 45, pp. 485–560, 1992.

[21] C. A. Poynton, "Gamma and its disguises," *J. Soc. Motion Picture Television Eng.,* vol. 102, pp. 1099–1108, 1993.

[22] D. G. Pelli and L. Zhang, "Accurate control of contrast on microcomputer displays," *Vision Res.,* vol. 31, pp. 1337–1350, 1991.

[23] *TIFF Revision 6.0,* Aldus Corporation, Seattle, WA, Final Rep., June 1992.

[24] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard.* New York: Van Nostrand Reinhold, 1993.

[25] A. B. Watson, K. R. K. Nielsen, A. Poirson, A. Fitzhugh, A. Bilson, K. Nguyen, and J. A. J. Ahumada, "Use of a raster framebuffer in vision research," *Behav. Res. Methods, Instrum., Comput.,* vol. 18, pp. 587–594, 1986.

[26] A. B. Watson and D. G. Pelli, "QUEST: A Bayesian adaptive psychometric method," *Percept. Psychophys.,* vol. 33, pp. 113–120, 1983.

[27] A. B. Watson, "Probability summation over time," *Vision Res.,* vol. 19, pp. 515–522, 1979.

[28] J. B. Mulligan and L. S. Stone, "Halftoning method for the generation of motion stimuli," *JOSA A,* vol. 6, pp. 1217– 1227, 1989.

[29] J. G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," *Vis. Res.,* vol. 21, pp. 409–418, 1981.

[30] F. L. van Nes and M. A. Bouman, "Spatial modulation transfer in the human eye," *J. Opt. Soc. Amer.,* vol. 57, pp. 401–406, 1967.

[31] A. B. Watson, "Estimation of local spatial scale," *J. Opt. Soc. Amer. A,* vol. 4, pp. 1579–1582, 1987.

[32] M. A. Berkley, F. Kitterle, and D. W. Watkins, "Grating visibility as a function of orientation and retinal eccentricity," *Vis. Res.,* vol. 15, pp. 239–244, 1975.

[33] D. C. Hood and M. A. Finkelstein, "Sensitivity to light," in *Handbook of Perception and Human Performance,* K. Boff, L. Kaufman, and J. Thomas, Eds. New York: Wiley, 1986, vol. 1, ch. 5.

[34] A. B. Watson, "Perceptual optimization of DCT color quantization matrices," in *Proc. IEEE Int. Conf. Image Processing,* Austin, TX, 1994.

[35] R. Rosenholtz and A. B. Watson, "Perceptual adaptive JPEG coding," in *Proc. IEEE Int. Conf. Image Processing,* Lausanne, Switzerland, 1996, vol. 1, pp. 901–904.

[36] A. B. Watson and J. A. Solomon, "A model of contrast gain control and pattern masking," *J. Opt. Soc. Amer*, vol. 14, 1997.

[37] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual Neurosci.,* vol. 9, pp. 181–198, 1992.

[38] J. M. Foley, "Human luminance pattern mechanisms: Masking experiments require a new model," *J. Opt. Soc. Amer. A,* vol. 11, pp. 1710–1719, 1994.

[39] A. B. Watson, "Image data compression having minimum perceptual error," U.S. Patent 5,426,512, 1995.

[40] A. J. Ahumada, Jr., A. B. Watson, and A. M. Rohaly, "Models of human image discrimination predict object detection in natural backgrounds," in *Proc. Human Vision, Visual Processing, Digital Display VI (IS&T/SPIE),* San Jose, CA, 1995.

[41] A. B. Watson, *Digital Images and Human Vision.* Cambridge, MA: MIT Press, 1993.

Andrew B. Watson studied perceptual psychology and physiology at Columbia University, New York, and at the University of Pennsylvania, Philadelphia, where he received the Ph.D. degree in 1977, and conducted post-doctoral research at Cambridge University, Cambridge, U.K., and Stanford University, Stanford, CA.

He is Senior Scientist for Vision Research at NASA Ames Research Center, Moffett Field, CA, where he conducts research on human visual perception and its application to coding, understanding, and display of visual information.

Dr. Watson is a Fellow of the Optical Society of America, and an editor of the journals *Visual Neuroscience*, *Journal of Mathematical Psychology*, and *Displays: Technology and Applications*. He is editor of *Digital Images and Human Vision*, (Cambridge MA: MIT Press, 1993).

Gloria Y. Yang received the B.S. degree in electrical engineering and computer science from the University of California, Berkeley, in 1993, and the M.S. degree in electrical engineering from the University of California, Los Angeles, in 1995. Her thesis research focused on perceptually optimized quantization for wavelet-based color image compression.

She is currently developing internet applications for IBM, Mountain View, CA.

Joshua A. Solomon studied experimental psychology at New York University, received the Ph.D. degree in 1992, and conducted post-doctoral research at Syracuse University, Syracuse, NY, and NASA Ames Research Center, Moffett Field, CA.

He is currently a Research Fellow at the University College London's Institute of Ophthalmology. He conducts research on visual detection and discrimination, and motion perception and attention.

John Villasenor (S'83–M'89) received the B.S. degree from the University of Virginia, Charlottesville, in 1985, the M.S. degree from Stanford University, Stanford, CA, in 1986, and the Ph.D. degree from Stanford in 1989, all in electrical engineering.

From 1990 to 1992, he was with the Radar Science and Engineering Section, Jet Propulsion Laboratory, Pasadena, CA, where he developed interferometric terrain mapping and classification techniques using synthetic aperture radar data. He is currently Associate Professor and Vice Chair of Electrical Engineering at the University of California, Los Angeles, and is conducting research on adaptive computing and on image and video coding and communications.