A Wavelet Visible Difference Predictor

Andrew P. Bradley, *Member, IEEE*

***Abstract*— In this paper, we describe a model of the human visual system (HVS) based on the wavelet transform. This model is largely based on a previously proposed model, but has a number of modifications that make it more amenable to potential integration into a wavelet based image compression scheme. These modifications include the use of a separable wavelet transform instead of the cortex transform, the application of a wavelet contrast sensitivity function (CSF), and a simplified definition of subband contrast that allows us to predict noise visibility directly from wavelet coefficients. Initially, we outline the luminance, frequency, and masking sensitivities of the HVS and discuss how these can be incorporated into the wavelet transform. We then outline a number of limitations of the wavelet transform as a model of the HVS, namely the lack of translational invariance and poor orientation sensitivity. In order to investigate the efficacy of this wavelet based model, a wavelet visible difference predictor (WVDP) is described. The WVDP is then used to predict visible differences between an original and compressed (or noisy) image. Results are presented to emphasize the limitations of commonly used measures of image quality and to demonstrate the performance of the WVDP. The paper concludes with suggestions on how the WVDP can be used to determine a visually optimal quantization strategy for wavelet coefficients and produce a quantitative measure of image quality.**

***Index Terms*—Human visual system, image compression, image quality, visible difference prediction, wavelet transform.**

I. Introduction

THE QUALITY of an image is a difficult concept to quantify. Image quality has traditionally had different definitions depending upon the context or application in which it is being used. For example, image formation [17, ch. 10], reconstruction [17, ch. 8], compression [5, ch. 6], enhancement [45, ch. 4], and display [45, ch. 4] all have different definitions of and techniques for measuring image quality. This paper, however, is primarily concerned with image quality as it is defined for image compression, where quality is measured relative to an original, uncompressed, image. This is done despite the fact that the observer may never see the original, uncompressed, image. The reasons for this are as follows.

1) The *visibility* of errors in the compressed image depend strongly on their location in the original image, i.e., whether they are in smooth or highly textured areas.
2) The visual *importance* of errors depends on their location in the original image, e.g., errors on salient edges, or on the face of a portrait will effect recognition more than errors in the background.
3) The compressed image should be a true representation of the original image and relay the same "high level" information as the original.

For an image compression scheme to be successful, it must exploit all of the redundancies inherent in digital images, namely the coding, interpixel, and psychovisual redundancies [14]. Both coding and interpixel redundancies have been successfully exploited in numerous compression schemes. However, psychovisual redundancies have often been more difficult to fully exploit. The reason for this difficulty is the lack of a reliable and complete model of the human visual system (HVS) that can be directly applied to image compression. This problem is also directly related to the lack of an adequate measure of image quality. In rate-distortion optimization, for example, a meaningful measure of image quality is required in order to jointly minimize rate and distortion. Currently, measures based on mean squared error (MSE) are used, e.g., [35].

Rate-distortion optimization attempts to minimize distortion globally throughout an image. A HVS model however, can potentially be used to vary the distortion spatially over the image. This is desirable because of the observation that more noise can be tolerated in highly textured image areas that in smooth areas. An example where this is important is in visually lossless compression where ideally one should keep the probability of detecting an error at each pixel *just* below the visual threshold [11]. A conventional compression scheme however, must keep the noise threshold below the global minimum, i.e., so that the noise is not visible in the smoothest areas of the image. This means that compression gains could potentially be made if the noise were allowed to increase in the more textured regions. To do this however, the compression scheme must have direct access to a model of the HVS so that the visibility of noise added in each area can be reliably judged. This type of approach has been successfully applied to 8×8 discrete cosine transform (DCT) subblocks [37] and is now an extension to the JPEG standard. This paper details an initial step in applying these same principles to wavelet based compression schemes.

It is the *long term* goal of this research to model characteristics of the HVS directly into a wavelet image compression scheme. This allows us to implicitly take advantage of effects such as frequency sensitivity and visual masking during the quantization process. Effectively, this means that we measure image quality in the transform domain. This approach is

Manuscript received June 27, 1997; revised July 14, 1998. This research was supported by the Motorola Wireless Data Group and the Canadian Natural Sciences and Engineering Research Council under Grant OGP9198. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Brian Funt.

The author was with the Department of Computing Science, University of Alberta, Edmonton, Alta., Canada T6G 2H1. He is now with Canon Information Systems Research Australia, North Ryde, NSW 2113, Australia (e-mail: bradley@research.canon.com.au).

Publisher Item Identifier S 1057-7149(99)03436-3.

conceptually more efficient than the alternative, which is to iteratively perform an inverse transform, measure the image quality, and then adjust quantization in the transform domain again. The focus of the current paper is a preliminary task toward this goal. We discuss a number of simplifications to a previously proposed model of the HVS [7] so that the model becomes more amenable to direct integration into a wavelet based image coder. These modifications include the following.

- Removal of the light adaption preprocessing.
- Use of a separable wavelet transform [6] instead of the cortex transformation [50]. Both critically sampled and over complete transforms are considered.
- Application of the contrast sensitivity function directly on the wavelet coefficients [37].
- A simplified definition of subband contrast that allows us to predict noise visibility directly from wavelet coefficients.

To demonstrate the efficacy of these modifications, and to highlight deficiencies in current measures of image quality, we have constructed a model that can be used to predict the visible differences between an original and compressed (or noisy) image.

The paper is structured as follows. Section II introduces some of the properties of the HVS that a model must incorporate if it is to reliably guide an image compression scheme. It also reviews some of the image quality measures commonly used in the image compression literature both to improve compression performance and as quality metrics. Section III discusses the wavelet transform and its similarities and differences to some proposed multiple channel models of the HVS. Section IV introduces visible difference prediction as a way both to measure image quality and to demonstrate the efficacy of the proposed model. The wavelet visible difference predictor (WVDP) is then described in detail. Section V discusses results obtained using the WVDP and suggests possibilities for further research, including a "visually optimal" quantization strategy for wavelet image compression and a quantitative measure of image quality.

II. HVS Models and Image Compression

This section is not intended to be a thorough review of the properties of the HVS; for that the reader is directed to [16], [45], and [49]. However, this section will outline some of the fundamental properties and terminology required as background to the visual model described later in the paper. The discussion is specific to *photopic,* or bright light vision. In this way, an image is viewed with the maximum visual acuity by the most densely packed, color sensitive, *cone* cells at the *fovea* of the eye. This means that there must be sufficient image illumination and that the image is either far enough away from the observer for its image to fall completely on the fovea, or the observer can visually roam the image for maximum detail.

A. Some Fundamental Properties of the HVS

It has been found that the HVS has a number of fundamental properties which, though not independent, have often been studied and modeled separately. In the following section we shall briefly mention these properties and discuss, in general terms, how they have been modeled. We shall discuss these properties roughly in the order in which they occur in the HVS and therefore the order in which they appear in HVS models.

- *Luminance Sensitivity:* Subjective brightness is known to be a nonlinear function of the light intensity incident on the eye [14, ch. 2, p. 34]. At practical light levels of around 100 cd/m^2 it is most commonly modeled by either a logarithmic or power law model [7, 17, ch. 3, p. 25].
- *Frequency Sensitivity:* The HVS is not only sensitive to the luminance levels in an image, but also to spatial changes of these luminance levels. For example, at a sharp intensity ramp in an image, the perceived brightness along the bright side of the edge is heightened, while along the dark edge the perceived darkness is deepened. This is often called the *Mach band* effect and implies that the modulation transfer function (MTF) of the HVS, often called the contrast sensitivity function (CSF), is band pass in nature. The CSF is defined to be the inverse of the contrast threshold function (CTF), which in turn is normally measured from threshold detection experiments [41], [45]. That is, by measuring the contrast required for an observer to detect sine wave gratings of various frequencies. The CSF has a peak response of between 2 and 10 cycles per degree (c/d), depending on the viewer and the viewing conditions [17]. It is also dependent on orientation, being most sensitive in the horizontal and vertical directions and least sensitive at oblique angles [12].
- *Signal Content Sensitivity: Contrast masking* is a phenomenon whereby noise can be masked, i.e., its visibility reduced, by the underlying image signal. For an image signal to mask a noise signal, both signals must occur in approximately the same spatial location, be of approximately the same spatial frequency, and their spatial frequencies must be of approximately the same orientation. These observations have led to the development of multichannel models of the HVS [7], [22], [41], [54], [55], often generally referred to as a *cortex transform* [50]. However, it should be noted that the HVS frequency sensitivity and masking are closely related, in that both effects suggest we are less sensitive to noise in high frequency, i.e., busy, parts of the image. This means that if we take advantage of the frequency sensitivity, outlined previously, this will directly reduce the amount of redundancy that masking can take advantage of [10], [20]. In addition, masking is a very complex process that cannot always be reliably modeled. For example, a *facilitation* effect may occur whereby the signal actually increases the visibility of the noise, i.e., the noise is effectively *unmasked* [46]. Also, the masking effect is more pronounced on noisy or random backgrounds than it is on regular, easily learned, backgrounds [53].

In addition, there are a number of other sensitivities, e.g., shape [16], [56], color [17, ch. 3], and temporal sensitivities [13], [21], that are not relevant to the model described in Section IV-A and so will not be discussed further.

B. HVS Properties Used in Image Compression

Models of the HVS have been used extensively in the image compression literature. The most common HVS phenomena considered are the luminance and frequency sensitivities [2], [4], [19], [25], [28], [30]. Visual masking has been used successfully in predictive coding [11], [42], vector quantization [15], and to some extent in block transform coding [51]. However, its use in subband coding has still not been fully exploited [2], [37], [38].

Measures used to evaluate image compression schemes are still primarily limited to the mathematical functions, such as MSE, peak signal-to-noise ratio (PSNR) and L_p-norm metrics [40], [43], [48]. Some more complex measures that use frequency weighting to take advantage of the CSF have also been used with some success [25], [26], [28], [29], [39]. Of interest is a technique that measures a number of both local and global properties of the error to obtain a quantitative picture quality score (PQS) [27]. Some of these properties relate to the luminance and frequency sensitivities of the HVS, whilst others are more *ad hoc* in nature, such as correlated, edge, and block errors. It should be noted that PQS does not attempt to explicitly model the multiple channel nature of the HVS. However, it has been shown to closely approximate the mean opinion score of observers (MOS) on a limited set of images, with correlation of more than 0.92 as compared to 0.57 for weighted mean square error (WMSE) [27]. How well this result extends to different image types, compression techniques, viewers, and viewing conditions has yet to be demonstrated. Other techniques, based more closely on the multiple channel models of the HVS, have also been shown to be highly correlated to subjective observer scores (0.84 HVS model versus 0.78 PSNR [54] and 0.74 HVS model versus 0.46 PSNR [55]). However, the use of such measures is yet to gain wide acceptance among image compression researchers. This is mainly due to their computational complexity and the fact that they can not be directly applied to image compression schemes.

There are a plethora of mathematical measures of image quality that measure quantities, such as *correlation quality*, *structural content*, or *image fidelity* [8]. However, in this paper we are primarily concerned with measures that directly model the HVS in order to obtain greater agreement with human rating of image quality. We believe that it is only by modeling properties of the HVS that an image quality measure will be able to reliably detect the most important visible differences between the compressed and original images. The parameters of these models will then mirror those specified for human observers, such as viewing distance and luminance levels, and be directly applicable to a broad range of image types, such as natural, remotely sensed, or computer generated images.

III. Wavelets and the HVS

The wavelet transform is one of the most powerful techniques for image compression [3], [40], [43]. Part of the reason for this is its similarities to the multiple channel models of the HVS [9], [23], [47]. In particular, both decompose the image into a number of spatial frequency channels that respond to an explicit spatial location, a limited band of frequencies, and a limited range of orientations. Fig. 1 shows an ideal spatial frequency decomposition of a four-level separable wavelet decomposition. It can be seen that subband decomposition provides a representation that mimics the multiple channel models of the HVS, as shown in Fig. 2. This structure has the potential to allow prediction of the visibility of errors in the wavelet domain and thus allow us to take advantage of visual masking in the quantization of wavelet coefficients.

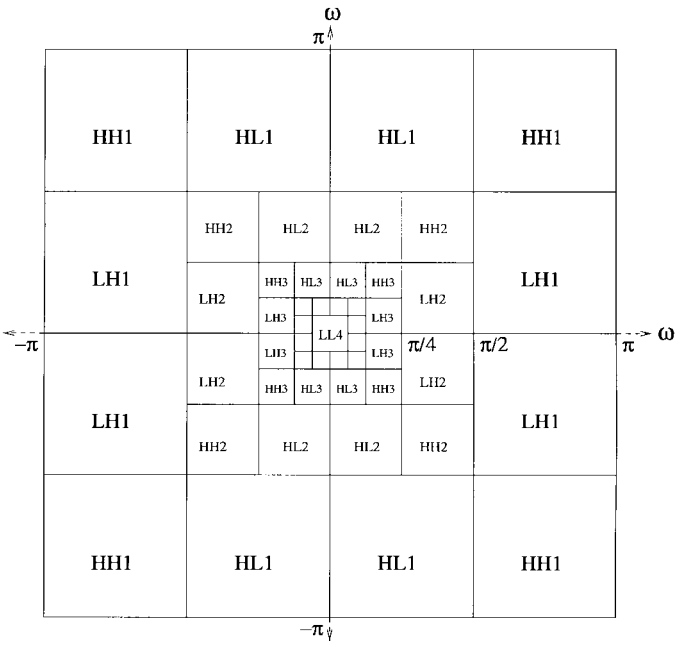

Fig. 1. Four-level wavelet decomposition in the frequency domain. Each subband is labeled with its orientation and decomposition level, e.g., HH1 indicates a first level diagonal subband.

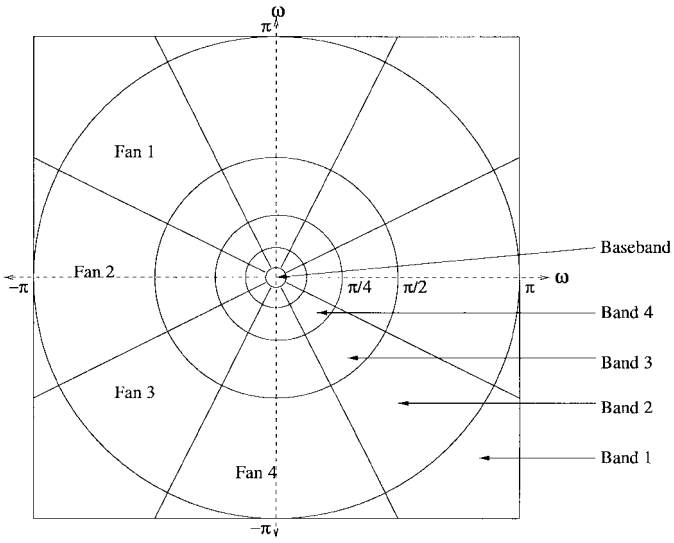

Fig. 2. Typical cortex transform decomposition in the frequency domain.

The similarities between the wavelet transform and multiple channel models of the HVS have been noted before [9], [23], [47]. In addition, a number of desirable properties of a spatial-frequency hierarchy used to model the HVS have been discussed [7], [50]. These properties are that the transform

1) must be invertible, so that the image response in the frequency channels can be directly related to the spatial response in the image;

2) should have unity frequency response, so that the channel model does not interfere with the HVS frequency response modeled (independently) in the CSF;
3) have an orientation selectivity that is sensitive to *at least* four different (45° equally spaced) orientations;
4) should have minimal overlap between adjacent frequency channels (this will result in negligible aliasing when/if the subbands are subsampled);
5) be shift invariant, or at least power "shiftable" [44];
6) be orthogonal, in order to reduce computational complexity;
7) have limited spatial extent (high-frequency basis functions should be of smaller spatial extent than low frequency ones);
8) basis functions should also have linear phase, each adjacent frequency scale (highpass lowpass pair) being in quadrature phase.

The biorthogonal wavelet transform considered here, like all of the linear transforms commonly used in image compression, does not have all of the above properties. The fact that it only has three orientation sensitive channels has been the main objection to the use of (separable) wavelet transforms, and other quadrature mirror filters, as a model of the HVS [7]. The wavelet transform has only one diagonal channel which effectively combines the responses from both 45° and 135°. Therefore, there is a danger of predicting a masking effect from an error that is in fact oriented at 90° to the underlying signal.

An additional problem with the critically sampled wavelet transform is its lack of translational invariance [44]. This drawback has limited the use of critically sampled wavelet transforms in image vision applications, such as edge detection and motion compensation [23]. Translational invariance in the wavelet transform is caused by the critical sampling of each subband, the very feature that makes it computationally attractive. This critical sampling violates the Nyquist criterion. However, because the aliasing errors introduced in each subband cancel out when they are recombined in the inverse transform there is no loss of information. Translations of the input signal will only produce simple translations of the wavelet coefficients when they are a multiple of *all* of the sampling factors, e.g., multiples of 8 (2^3) for a three-level decomposition. In general, when the translation is not a multiple of the sampling factors, the energy in each subband will move between each of the other subbands in an unpredictable manner [44]. For a visible difference predictor, like any other image analysis application, this effect is *not* desirable and will lead to incorrect predictions about the visibility of image errors. For example, changing just one pixel in an image will produce different coefficient errors depending on where, relative to the sampling lattices, the pixel is changed.

IV. Visible Difference Prediction

From the multiple channel models of the HVS it is possible to predict, on a pixel by pixel basis, if the noise introduced in the compressed image will be visible to a human observer. However, even from this point it is still no simple task to combine these visible errors and obtain one quantitative measure of image quality. In fact in its general sense, image quality cannot be given just one single value. This is because our definition of image quality is as a *subjective* measure and will therefore vary considerably between different viewers, their expectations, image type, and final application [1], [20]. However, despite this note of caution, it may still be possible to develop a one-dimensional (or perhaps multidimensional) quantitative measure of image quality for a specific application and viewing conditions.

Of particular interest in the design and evaluation of image compression schemes is the idea of comparing the original and compressed image and obtaining a probability map of visible differences between the two. This has been called *visible difference prediction* (VDP) [7], [22] and can be used either, as a starting point for summing the errors [54], [55], or as an error metric in itself. The VDP map provides an indication of the location and degree of visual errors in a compressed (noisy) image. It allows a system designer to see the effects of design changes without having to use expensive, time consuming, human observer trials. A VDP describes the threshold behavior of the HVS, but does not attempt to discriminate among different suprathreshold errors, this task is left to the system designer.

For a model of the HVS to be suitable for direct integration into a wavelet based image compression scheme the HVS model should use the same linear transform as the compression scheme. In this way, coefficient quantization errors can be directly related to error detection probabilities in the model. To do this, the model should work directly on the transform coefficients rather than local band contrast values. In addition, the computational burden of the model can be reduced if we remove the need for CSF prefiltering of the images. This can be done by including the CSF directly into the contrast masking function. Computation can further be reduced if we assume the grey levels in the image relate directly to perceived brightness on the image display. This approximation may not always be applicable, but when it is, it removes the need to model both the display luminance and observer luminance sensitivity functions. In the next section we describe a model that applies all of these ideas and we demonstrate the efficacy of the approach by constructing a visible difference predictor.

A. Wavelet Visible Difference Predictor

Fig. 3 shows the component parts of the WVDP model, which is based on the VDP model described in [7]. The wavelet transform used is the linear-phase 9/7 biorthogonal wavelets [6] previously used in psychovisual quantization experiments [52]. In this paper the wavelets can be either *critically sampled*, or *overcomplete*, i.e., with no subsampling. We shall discuss the WVDP in general terms pointing out any differences that may be required for the critically sampled or overcomplete transforms. In the examples in this paper we have used a four-level wavelet decomposition which, although is less than the five to six levels thought to exist in the HVS [49], is thought to trade off computational complexity and accuracy. In addition, a four-level decomposition allows the

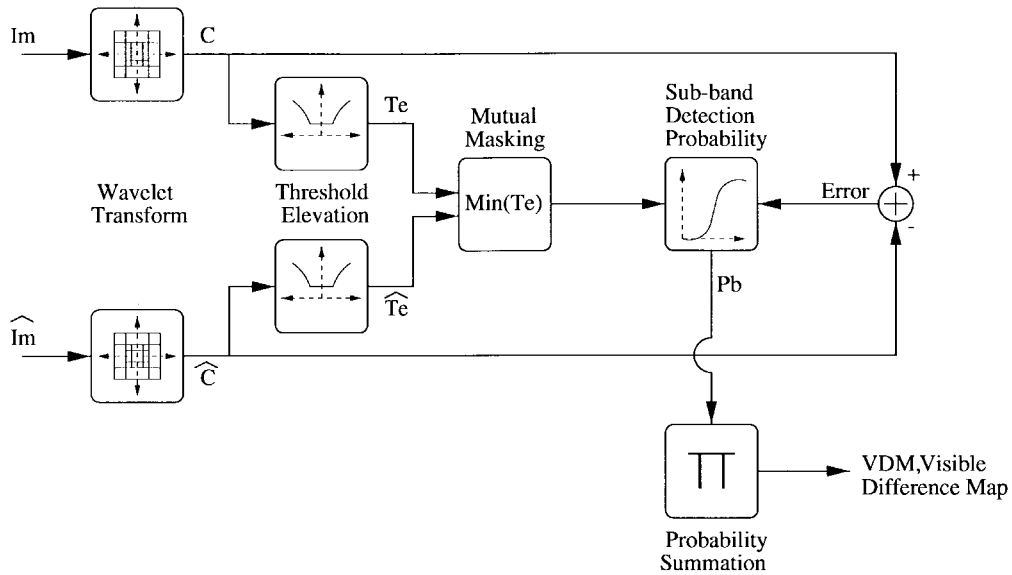

Fig. 3. Wavelet visible difference predictor (WVDP).

wavelet coefficients to be rearranged into 16 × 16 blocks, which are then suitable for block adaptive quantization.

1) Outline of the WVDP: The original image, I_m, and the noisy image, $\hat{I}_m$, are first transformed to the wavelet domain. The differences between I_m and $\hat{I}_m$ (the errors) are tested against a contrast masking, or threshold elevation function, T_e. The contrast masking function defines the amount of error that can be added to a wavelet coefficient, e.g., during quantization, without it being visible after reconstruction. The contrast masking function used is based both on psychovisual experiments with the wavelet transform [52] and the idea that image areas of large contrast (large coefficients) can tolerate more noise. A psychometric function is then used to estimate the error detection probability for each wavelet coefficient and the detection probabilities are combined to give the error detection probability at each pixel.

2) Luminance Sensitivity: It is assumed that both the original image and the distorted (or compressed) image are displayed on a monitor that has been gamma corrected, i.e., that the display luminance is a nonlinear function of image grey level. This assumption means that image grey level is directly proportional to perceived brightness on the monitor and allows us to standardize the WVDP model.[1] If a monitor is used that is not gamma corrected the model requires a suitable nonlinear transform to be applied to the images before being applied to the WVDP model, e.g., see [7].

3) Frequency Sensitivity: The frequency sensitivity of the HVS (the CSF) is not explicitly modeled in the WVDP, but implicitly built into the threshold elevation function. This has the advantage of reducing computational complexity of the WVDP, but does mean that the threshold elevation function is no longer the same in each subband.[2] The threshold elevation function has two parts, the first is the minimum threshold when there is little or no image contrast, i.e., no contrast masking. The second is an increasing function of image contrast that defines the effect of contrast masking.

The minimum of the threshold elevation function is based on results from psychovisual detection of noise added directly to wavelet coefficients and subsequently viewed on a gamma corrected monitor [52]. The noise was added to the wavelet coefficients of a blank image of grey level 128. This gives a *minimum* threshold based on no contrast masking. The model fitted to the data, to calculate the minimum threshold, y, was as follows [52]:

$$\log(y(\theta, f)) = \log(a) + k \cdot (\log(f) - \log(g_\theta f_0))^2 \quad (1)$$

where a is the minimum (0.495), k the width (0.466), and $g_\theta f_0$ the minimum of the parabola (f_0 = 0.401 and g_θ is 1.501, 1, and 0.534 for the LL, LH/HL, and HH subbands, respectively). An example of this function is shown in Fig. 7. The spatial frequency, f, is determined by both the viewing conditions (maximum display resolution and viewing distance) and the wavelet decomposition level. The minimum threshold estimated from this model, y, is expressed as the peak amplitude of the signal detected by an observer. For the case when a single coefficient was set (the experiment was also done for uniform noise added to a subband) this relates to the peak of the impulse response when the wavelet pyramid is reconstructed. Examples of impulse responses, i.e., wavelet basis functions, from the first three levels of LH subbands are shown in Figs. 4–6.

In order to apply the results presented in [52] to a model of the HVS, we need to convert the measured detection thresholds from the spatial domain to the wavelet domain. In this way we estimate the size of the wavelet coefficient that produced the detected spatial (impulse) response. To do this we have a "worst case" formula that estimates the coefficient detection

[1] Often, the gamma of the display will not perfectly match the gamma of the eye and so there is reduced sensitivity at both the dark and bright luminance extremes [10].

[2] The reasons why the masking function can be the same for each subband are detailed in [7].

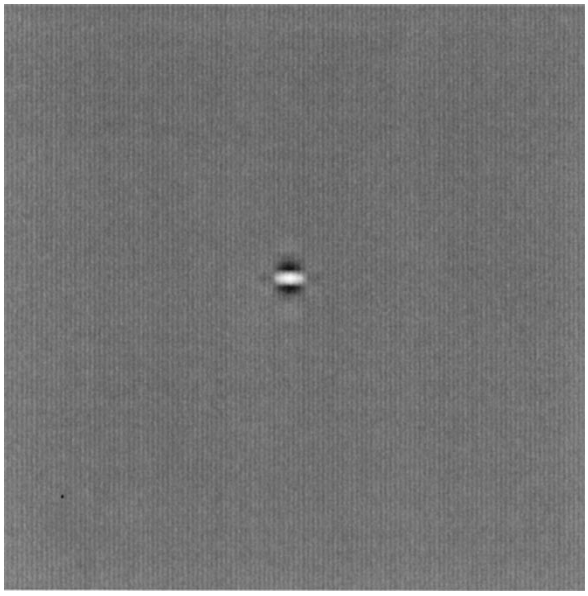

Fig. 4. Synthesis filter, first level impulse response.

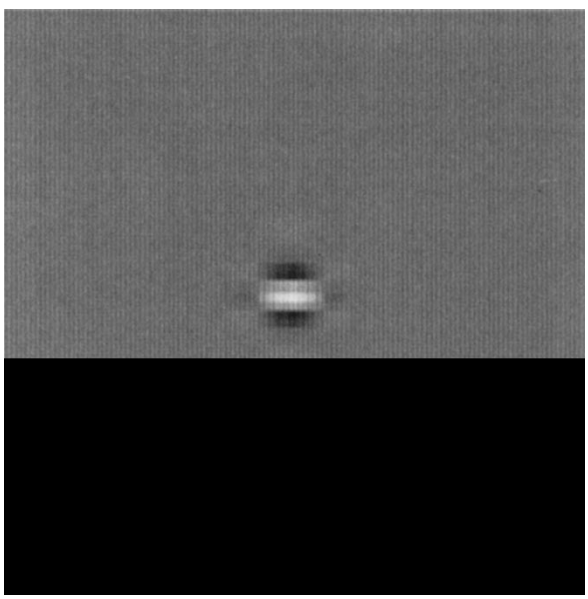

Fig. 5. Impulse response, second level.

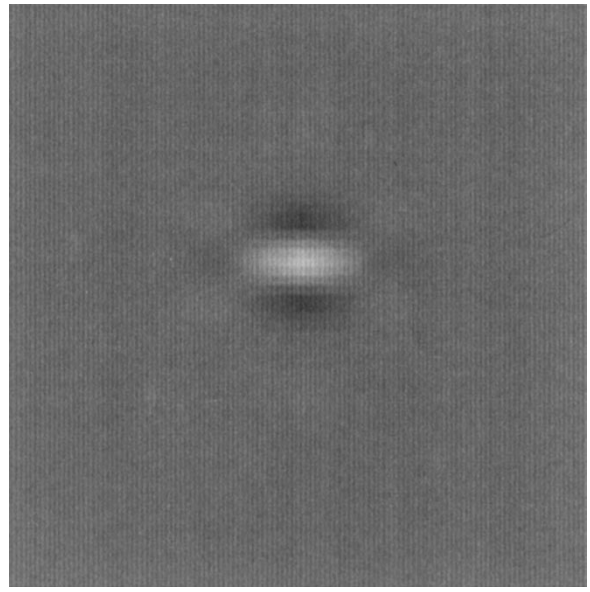

Fig. 6. Impulse response, third level.

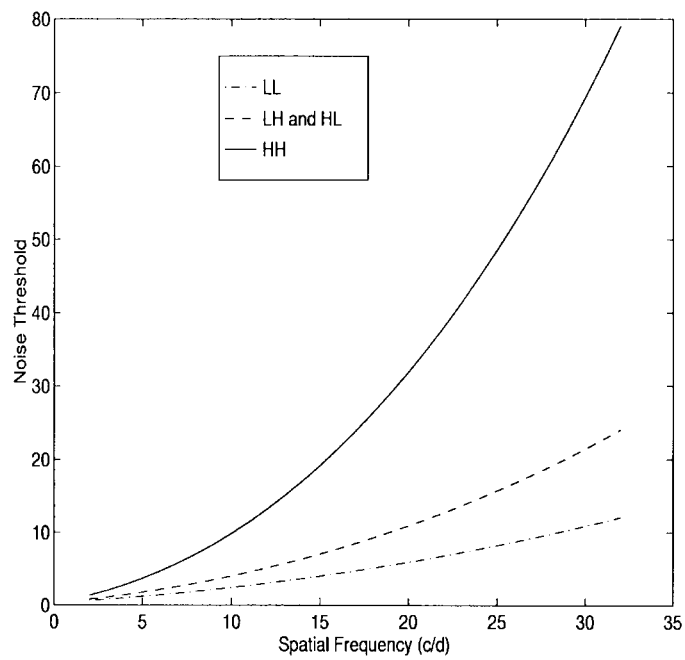

Fig. 7. Detection thresholds for HH, LH, and LL subbands.

threshold, n_c:

$$n_c(\theta, f) = \frac{y}{i_\theta \cdot p_l^{2(l-1)}} \tag{2}$$

where l is wavelet decomposition level, i_θ is either p_l^2, p_h^2, or $p_l \times p_h$ for the LL, HH, or LH/HL subbands, respectively. The maximum coefficient amplitudes for the low pass and high pass linear-phase 9/7 synthesis filters are $p_l = 0.788\,485$ and $p_h = 0.852\,699$, respectively [6].

4) Signal Content Sensitivity: The complete masking function is determined from (2) and from the masking effect due to the *contrast* in the subband. This means that the minimum threshold, n_c, is a function of frequency level (f) and orientation (θ) only, while the masking effect is also a function of the actual value of the wavelet coefficients, $C(i, j)$. The threshold elevation function, $T_e(\theta, f)$, which is a simplification of the one used in [7], is given by

$$T_e(\theta, f, i, j) = \begin{cases} \max(n_c(\theta, f), b(f) \cdot C(i, j)), & \text{if } C(i, j) > 0 \\ \max(n_c(\theta, f), b(f) \cdot \|C(i, j)\|), & \text{otherwise.} \end{cases} \tag{3}$$

The constant $b(f)$ can be used to alter the slope of the masking function at each frequency level of the decomposition. We use a constant value of $b(f) = 1$ for all results reported in this paper, which corresponds to experimentally derived slopes for phase-incoherent (noise) masking [7]. An example threshold elevation function is shown in Fig. 8.

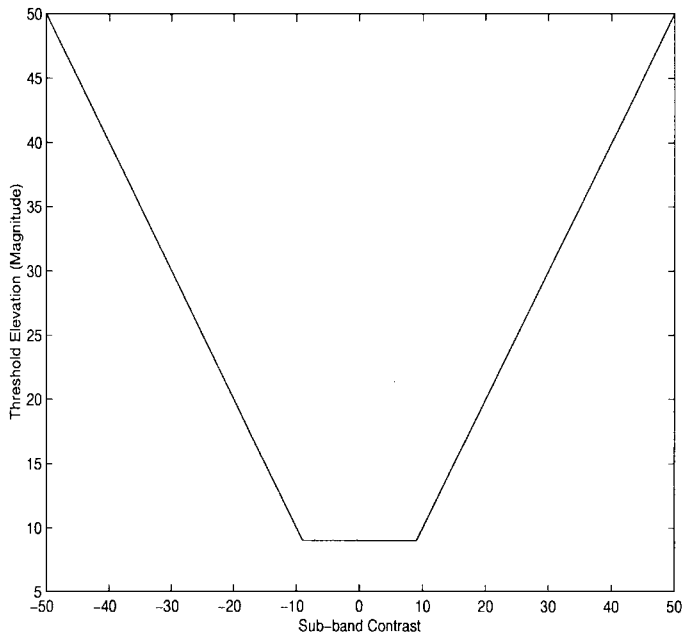

Fig. 8. Example threshold elevation function for $n_c = 9$, $b(f) = 1$.

5) Mutual Masking: The amount of masking that occurs in a compressed image is dependent not only on the content of the original image, but also on what happens to the image contents once they are compressed. For example, if the original image contained a highly textured (busy) area, one might assume that we could expect a significant masking effect to occur in that area. However, if the compression scheme effectively smooths that busy area, then this assumption would be incorrect and little, or no, visual masking would occur. In addition, if a smooth area in the original is made highly textured by the compression scheme, then again no visual masking should occur. This effect is called *mutual masking* [7] and is usually accounted for by taking the minimum (T_{em}) in each band for both the original, T_e, and noisy, $\hat{T}_e$ thresholds. In this way only areas that are highly textured in *both* the original and compressed images produce a significant masking effect.

6) Simplified Definition of Band Contrast: Image contrast is not, strictly speaking, equal to the value of the wavelet coefficient, $C(i, j)$. A measure such as band limited contrast, $c(i, j)$ [31], should be used in each channel of the HVS model

$$c(i, j) = \frac{a(i, j)}{l(i, j)}. \tag{4}$$

Here, $a(i, j)$ is the filter response at location (i, j) in the subband of interest and $l(i, j)$ is the local luminance mean at this location, i.e., the total response at this location of all the subbands below the subband of interest. However, if we assume the local luminance to be constant across the image and equal to the average value of the coefficients in the lowest frequency subband [in this case the low frequency coefficients at the fourth level, $E(C_{LL4})$] then (4) becomes

$$c(i, j) = \frac{C(i, j)}{E(C_{\mathrm{LL4}})}. \tag{5}$$

Because $E(C_{\mathrm{LL4}})$ is a constant applied to each wavelet coefficient, the band limited contrast, $c(i, j)$, can be assumed to be directly proportional to the value of the wavelet coefficients, $C(i, j)$. Therefore, $C(i, j)$ can be used as a measure of subband contrast in the WVDP. This simplification is necessary because (1) was determined from actual wavelet coefficient values. Moreover, it has the advantage that the WVDP model can be directly applied to the quantization of wavelet coefficients in an image compression scheme.

7) Probability Summation: Once the threshold elevation, T_{em}, has been calculated, it is used in conjunction with the coefficient differences ($\Delta C(i, j) = C(i, j) - \widehat{C}(i, j)$) to calculate a detection probability for each coefficient in each subband. A psychometric function then converts these differences, as a ratio of the threshold elevation, to subband detection probabilities. Applying the psychometric function to the contrast difference between the original and noisy image ($\Delta C(i, j)$) and the threshold elevation function ($T_{em}(i, j)$) gives [7]

$$P_b(i, j) = 1 - e^{-|\Delta C(i, j)/(T_{em}(i, j)\cdot\alpha)|^{\beta}}. \tag{6}$$

Here, the slope of the psychometric function, β, is 2.0 and the decision threshold, α, is 4.0. The value of α was set so that an error of $n_c(\theta, f)$ on a uniform field resulted in a detection probability of 0.5 [7]. Therefore, this relates to one just noticeable difference (JND).

$P_b(i, j)$ calculates the probability of detecting a visible difference in *each* subband, i.e., for each coefficient in the wavelet decomposition. The final output of the WVDP is a probability map, i.e., the detection probability at each *pixel* in the image. Therefore, the probability of detection in each of the subbands must be combined for every spatial location in the image. This is done using a product series [7]:

$$P_d(i, j) = 1 - \prod_b (1 - P_b(i, j)). \tag{7}$$

Here, the probability of detecting a visible difference at each location in the image, $P_d(i, j)$, is defined to be the product of the probability of detection in every subband. This means that for the critically sampled transform every sibling and parent is used, at each location, while for the overcomplete transform we know the subband detection probability at each pixel.

Note that in (6) we have taken the absolute value of the ratio of the coefficient differences, ΔC, and the threshold elevation function, T_{em}. This effectively removes the sign of the differences between the original and noisy images. This information could be used to indicate whether a visible difference is lighter or darker than the original [7]. However, at this stage we are primarily concerned with an error being visible or not.

V. Visualization of Detection Maps

In this section we will compare the WVDP to a number of more conventional measures of image quality such as mean squared error and weighted mean square error. To do this we will show error map images that either show the probability of detecting an error or the value of the *squared* error at each pixel.

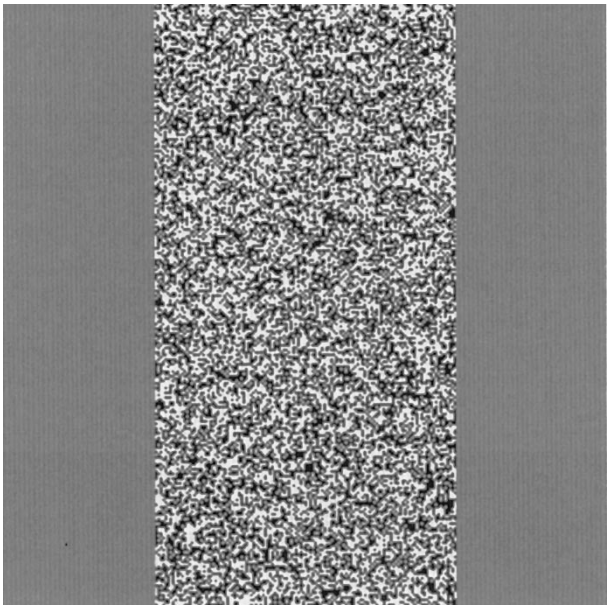

Fig. 9. Original simple test image.

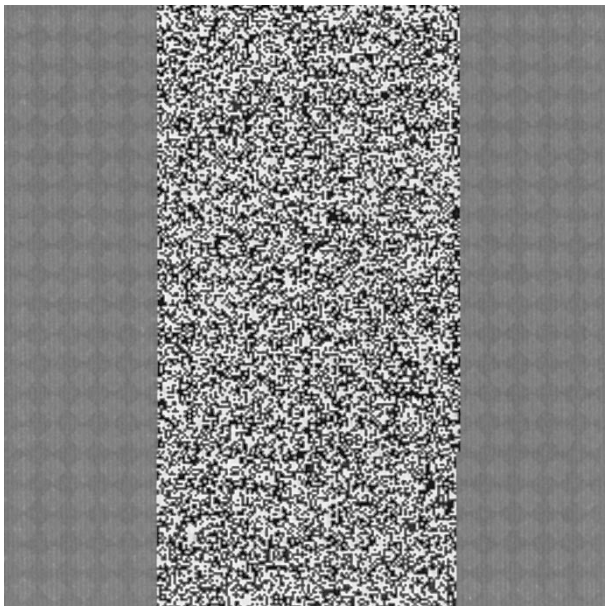

Fig. 10. Test image plus checker noise.

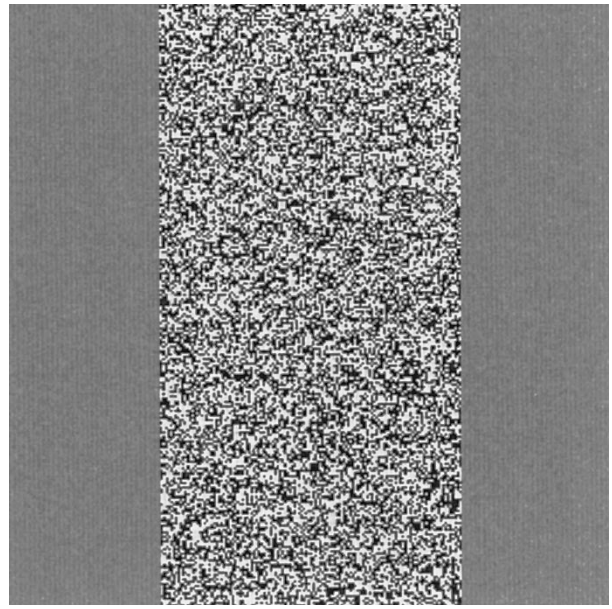

Fig. 11. Test image plus random noise.

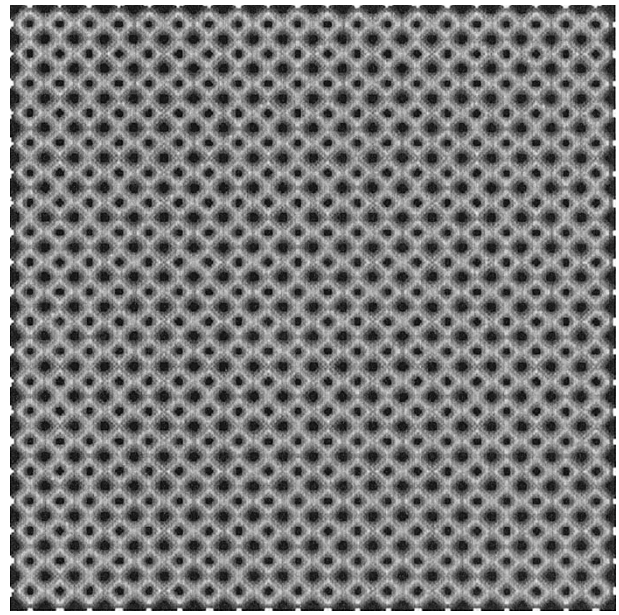

Fig. 12. CSF weighted squared error, checker noise.

A. A Simple Test Image

Initially, we will use a simple test image that consists of a smooth background of grey level 128, with a vertical bar of random texture in the middle. This test image is shown in Fig. 9. Next we have created two noisy versions of the test image. The first has a diagonal checkerboard pattern added to it with peak amplitude $[-4, 4]$, producing a PSNR of 36.25 dB. We shall refer to this image as the checker image. It is shown in Fig. 10. The second image is created by adding zero-mean Gaussian noise with variance estimated from the checker image. Therefore, this image will also have a PSNR of approximately 36.25 dB (36.27 dB). We shall refer to this image as the random noise image. It is shown in Fig. 11.

Despite the fact that both of the checker and random images were created to have the same MSE, it is fairly obvious to see that the checker error is more visually objectionable than the random error. This immediately illustrates the limitations of using simple mathematical measures of image quality, such as MSE and PSNR, and highlights the need to use HVS models in image quality analysis. The other point to note is that there is significant visual masking in the central textured stripe of the test image. This means that even though the error images are both added equally over the whole test image, the errors are significantly less visible in this central portion. Therefore, we can see that for any quantitative measure of image quality to be effective it must consider the spatial variability of the image contents.

Figs. 12 and 13 show the squared error image produced by a frequency weighted signal to noise measure (WPSNR). The frequency weighting is done according to the following

Fig. 13. CSF weighted squared error, random noise.

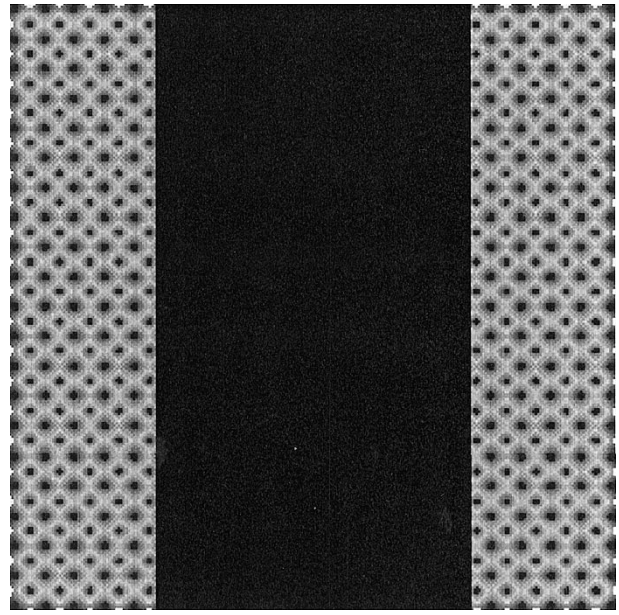

Fig. 14. As Fig. 12, but weighted by local contrast.

approximation to the CSF [27]:

$$S(\omega) = 1.5e^{-\alpha^2\omega^2/2} - e^{-2\alpha^2\omega^2} \tag{8}$$

where $\alpha = 2$, $\omega = 2\pi f/60$, and $f = \sqrt{u^2 + v^2}$, u and v being the horizontal and vertical spatial frequencies, respectively, in c/d.

Because of the bandpass nature of the CSF, WPSNR correctly gives the most weight to the edges of the checks and relatively little weight to the central portion of each check. However, WPSNR does not take the image contents into consideration and so weights all spatial locations equally. This is obviously incorrect in this test image due to the masking effect of the central stripe. Also of interest are the actual values given by WPSNR, in that it prefers the checker error (37.21 dB) to the random error (36.73 dB). This is obviously incorrect and is due to the fact that in the checker image the most frequently occurring errors (on a per pixel basis) are in the central portion of the checks and because these are low frequency errors they are given less weight in WPSNR. The edges of the checks are the most visually annoying part of the error. However, they occur less frequently and so contribute less in total to WPSNR. This is a case where the actual shape of the error, i.e., the linear edges, increases the visibility of the error with a subsequent reduction in image quality. This effect cannot be accounted for by a purely frequency-based error measure such as WPSNR.

A simple way in which to take account of image contents (and therefore visual masking) is to weight a measure such as WPSNR by the inverse of image contrast in a local neighborhood.[3] In this way the error gets a large weighting in smooth areas, where the contrast is low, while at edges or in textured parts of the image the increased contrast reduces the error weighting. An example of such a contrast weighted WPSNR, in this case weighted by the inverse of the variance in a 5×5 neighborhood, is shown in Fig. 14. This simple masking function however, will not be able to distinguish edges from textures, or patterns of different orientations. It will therefore incorrectly predict a masking effect when there may not be one. Also because it does not use multiple frequency detection channels it will not be able to distinguish the masking differences between the random and structured (checker) errors used here.[4]

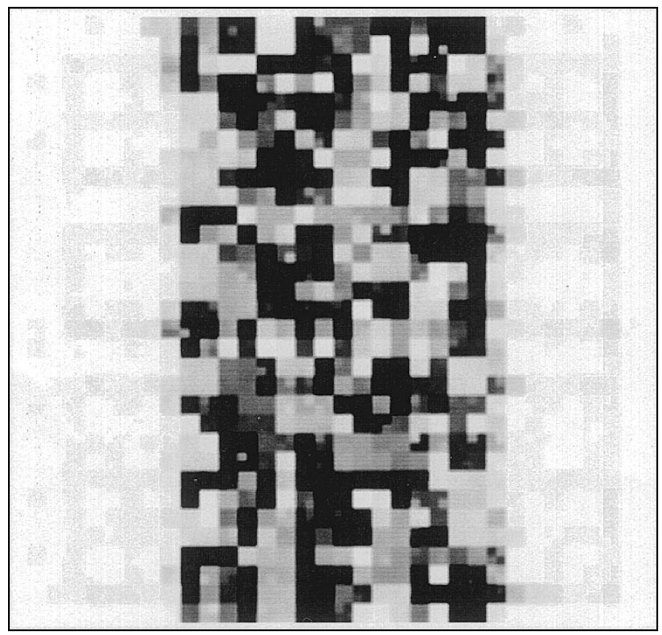

Fig. 15. WVDP probability map, checker noise.

Figs. 15 and 16 show the probability of detecting a visible difference between the test image and the checker image and the test image and the random image, respectively. The results were obtained using the *orthogonal* wavelet transform WVDP. Figs. 17 and 18 show the same probability maps, but for

[3] A sort of "engineering approximation" for visual masking.

[4] The random error has a flat frequency spectrum. However, the frequency spectrum of the checker error has a number of peaks that will increase its probability of detection.

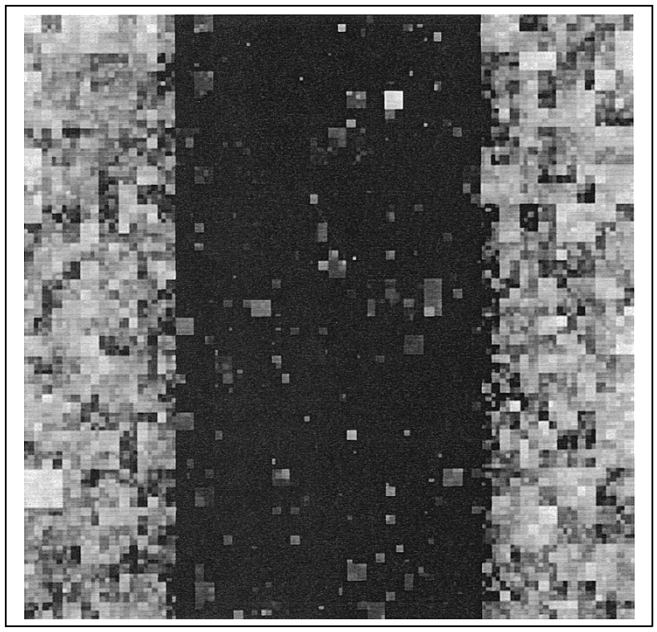

Fig. 16. WVDP probability map, random noise.

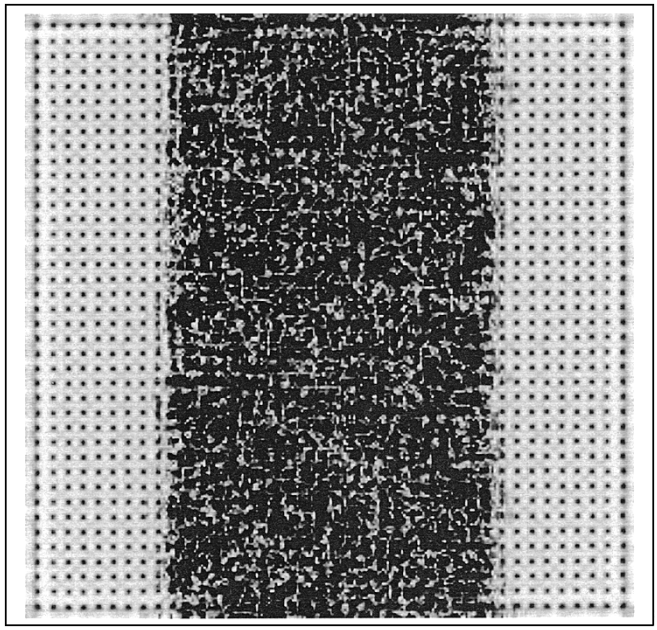

Fig. 17. Same as Fig. 15, but overcomplete wavelet.

the *overcomplete* wavelet transform WVDP. It can be seen that both the orthogonal and overcomplete WVDP's correctly predict the visual masking effect of the central stripe and that the `checker` error is more visible than the `random` error. However, the orthogonal WVDP has a number of false alarms in the central stripe, where due to the masking effect the errors should not be visible. The overcomplete WVDP also has some incorrect predictions in this central area, however, these are far less objectionable, i.e., they are from the highest frequency subbands.

The reasons that the orthogonal wavelet decomposition has more prediction errors than the overcomplete transform are primarily those discussed in Section III, as follows.

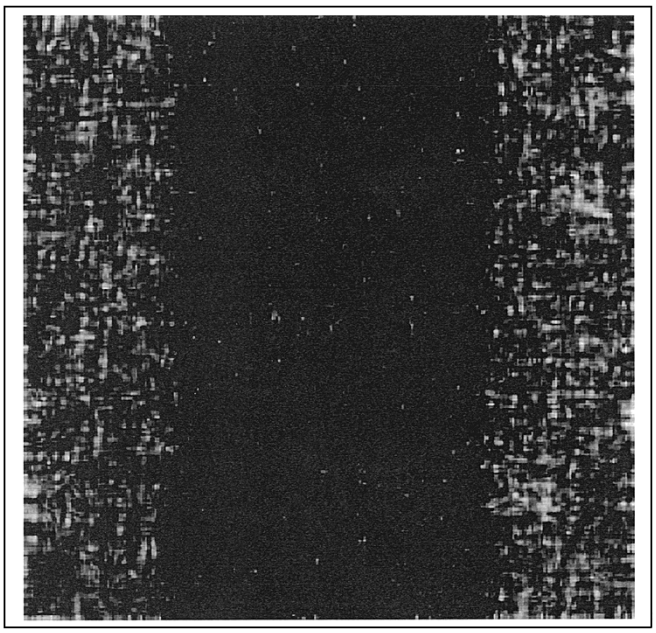

Fig. 18. Same as Fig. 16, but overcomplete wavelet.

- Due to the critical sampling and resultant aliasing both the masking function and the coefficient errors will be unreliable. For example, a masking effect may not be predicted in a busy area, or large coefficient errors may appear when there are not large image errors (or vice versa).
- In addition, the reduced spatial resolution of the lower frequency subbands means that when unreliable predictions are made they affect a large area in the probability map.

The overcomplete wavelet VDP does not subsample and so has both more reliable masking/error detection and full spatial resolution at each frequency subband. For these reasons the VDP maps given in the next section will be from the overcomplete wavelet transform only.

B. A Natural Image

Fig. 19 shows the face portion of the Barbara image commonly used in the image compression literature. Figs. 20–22 show the Barbara image corrupted with the checker noise, random noise, and noise added as a result of compression by the JPEG algorithm [32], respectively. All three of these noise-corrupted images have approximately the same error measured using PSNR (checker = 36.25 dB, random = 36.27 dB, JPEG = 36.15 dB). The JPEG quality factor is chosen to give a PSNR as close as possible to 36.25 dB. A visual inspection of these images would rate the JPEG image as the highest quality, random next, and checker as the lowest quality. However, due to the limited quality of printing, these differences may be difficult to see. Therefore, picture quality scores (PQS) [27] were also measured to confirm our visual judgment (checker = 3.50, random = 3.83, and JPEG = 3.87). Note that the JPEG error *should* be the least visible error of the three because the JPEG algorithm has been designed to quantize image areas that have high spatial frequency with the most severity. Therefore, the compression noise added by

Fig. 19. Original Barbara image.

Fig. 21. Barbara plus random noise.

Fig. 20. Barbara plus checker noise.

Fig. 22. Barbara plus JPEG noise.

JPEG varies spatially and takes advantage of the CSF and visual masking to some extent [20], [51]. This can also account for the good subjective performance of WPSNR over PSNR [25], [26], [28], [29], as most compression schemes will place the largest errors in the highest frequency areas of an image. These errors are then likely to be highly correlated to the image area, i.e., be high frequency also, and so will receive a low weighting from WPSNR.

Figs. 23–25 show detection probability maps from the overcomplete WVDP for the checker, random, and JPEG images, respectively. For the spatially invariant noise images (checker and random), these figures show the following.

- A reduced probability of detecting an error in the busy areas of the image, such as the striped head scarf and the wicker chair.
- An increased probability of detection in the smooth areas of the image, such as the background, arms, forehead, and cheeks.
- Increased detection probability of the checker noise over the random noise.

For the JPEG image the situation is reversed. In the smooth areas, where there is little error, there is only a small probability of detection, while in the busy areas, where the error is high, there is an increased probability of detection. However, because of the visual masking effects in these areas, this probability is not high (<0.5).

Despite our previous objections to single number quality metrics (see Section IV) they are often necessary for quantitative evaluation and comparison. Here we will define a simple WVDP *impairment* score, P_s, that is the Minkowski sum of

Fig. 23. WVDP map, Barbara, checker noise.

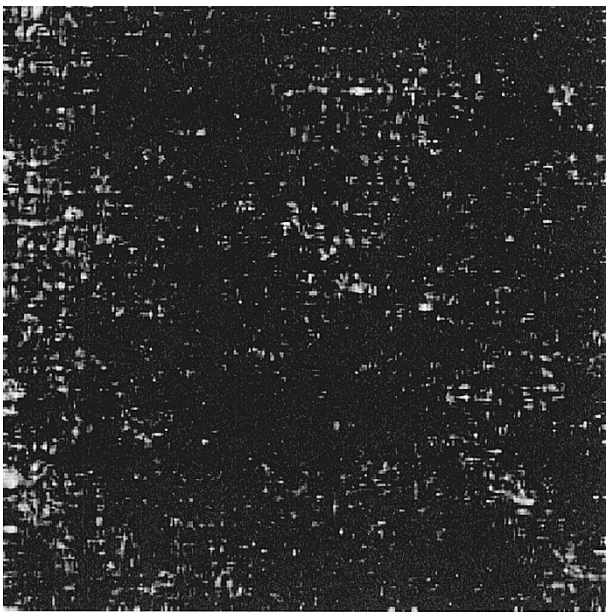

Fig. 24. WVDP map, Barbara, random noise.

the detection probabilities, $P_d(i, j)$:

$$P_s = \left\{ \sum_{image} |P_d(i, j)|^\beta \right\}^{1/\beta}. \tag{9}$$

The Minkowski metric is a standard feature of many current vision models [51] and here we have chosen to do probability summation over the whole image with $\beta = 4$. This WVDP score may not necessarily be the best way of obtaining a single figure of merit from the WVDP map images. Alternative measures, such as the *peak probability of detection* (Minkowski sum with $\beta = \infty$) or *critical distance* [the minimum viewing distance where $P_d(i, j) < 0.5 \forall (i, j)$] have also been proposed and may be more suitable depending on the application [7]. The WVDP *impairment* score was calculated for the test

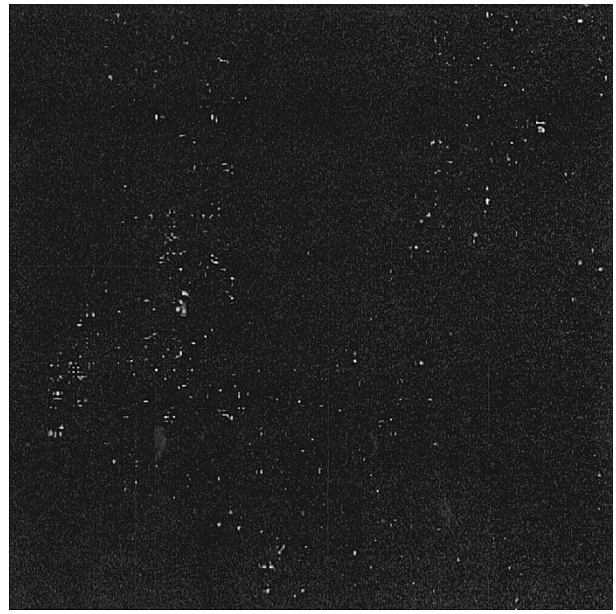

Fig. 25. WVDP map, Barbara, JPEG noise.

images and confirms the subjective quality ordering (JPEG = 1.934, random = 3.035, and checker = 9.285). This ordering also held for the orthogonal wavelet transform, though due to the reduced size of the probability map the scores are significantly smaller (JPEG = 1.239, random = 1.76, and checker = 3.12). Note that would not expect these values to be significantly different from the PQS scores. However, unlike PQS the WVDP score comes from a model based directly on a multiple channel model of the HVS.

C. Limitations of the WVDP

The visible difference maps produced by the WVDP are not perfect. There are a number of areas where the probability of detection is over 0.5 when there are no visible differences between the images. We believe that this is primarily due to an unacceptable amount of overlap between adjacent subbands, not only in different levels in the wavelet pyramid, but also between the sibling (orientation) subbands. This has undoubtedly affected the reliability of the WVDP as errors that should only appear in one subband are now spread across a number of subbands. Compared to the *cortex* filters used previously [7], [50], [54], which are specified in the frequency domain using *mesa* filters, the dilated wavelets used here [6] are specified by their impulse response only. Though having filters represented by only a small number of taps (7 and 9 for the lowpass and highpass, respectively) may agree more closely with the mechanisms found in the HVS [33], it does mean that their frequency response may not be as well suited to a HVS model as filters designed in the frequency domain.

The results presented in this paper may not be as good as those of the original VDP model [7]. However, this paper demonstrates that the WVDP predicts the majority of image distortions with a simple model based directly on the wavelet transform. The fact that we have tried to make the WVDP model directly applicable to image compression has meant that we have made a number of simplifying assumptions that have

also penalized prediction performance. For example, we have no CSF prefiltering, which though it reduced computational complexity and allowed direct application of the previous psychovisual model [52], means that different masking functions had to be used at each level of the decomposition. In the absence of any psychovisual data on which to base these masking functions, they were based on ones used previously in a different HVS model [7]. However, there is no guarantee that these masking functions are in any sense the best ones to use and the values of the masking parameters used in this paper will require more formal calibration using visual trials. In addition, the fact that we are using the actual values of the wavelet coefficients to estimate the subband contrast is an approximation that may not be valid for a lot of images. However, we feel that these modifications are necessary for the model to be suitable for integration into a wavelet compression scheme.

Perceptual models, such as the WVDP, may have difficulty in predicting image quality when the visual distortion is well over the visual threshold, i.e., is suprathreshold. This is because the models are formed using perceptual factors measured near the visual threshold. One way of addressing this problem is to weight the objectionability of compression artifacts in a linear regression analysis, e.g., PQS [27]. However, the limitation of this approach have are that different observer groups may judge artifacts using different weights and so the model becomes specific to a particular observer group and their perceptual preferences. In addition, visual thresholds can vary considerably over different observer groups[5] and over time as observers get better at detecting artifacts. Therefore, care must be taken when comparing or calibrating the models to specific groups of observers.

D. Extension to Wavelet Coefficient Quantization

Taking account of properties of the HVS in image coding, often called *perceptual coding* [18], is important if we are to take advantage of all of the available sources of image redundancy [14]. To gain maximum compression, it is important that the wavelet coefficients are quantified with the maximum severity whilst keeping the perceptual error below the visual threshold, or constant across the image. The WVDP model provides a framework in which to achieve this. A simple method to implement visually lossless wavelet compression would be to keep the probability of detection of each coefficient in each subband to below say 0.2. From (7), the total probability of detection will therefore be limited to less than $1-(1-0.2)^4 = 0.59$, i.e., slightly more than one JND. Using this methodology it is unlikely that an average observer will see any visible distortion. Another possibility is to calculate P_s on subblocks of the image and iteratively adjust quantization to obtain uniform impairment over the whole image. This is already done using a DCT-based visual model and applying uniform scaling of the quantization matrix [51]. Subblocks of wavelet coefficients can either be obtained by rearranging a "whole image" transform or by implementing a lapped orthogonal transform [24].

The masking functions described in (3) could be implemented using a nonuniform linear quantizer, i.e., quantizer bin width increases (increasing quantization error) for larger wavelet coefficients. However, an optimal uniform quantizer, perhaps designed using (2) alone, coupled with entropy encoding may have similar performance to this nonuniform quantize [14], [36]. For this reason a vector quantization scheme may be preferable to take advantage of visual masking effects [15].

[5] For example, image compression experts will have a much lower threshold than, say, radiologists who may be familiar with the images, but who may not be familiar with image compression artifacts.

VI. Conclusions

This paper has highlighted the limitations of some common measures of image quality and discussed the need to take account of properties of the HVS to produce more reliable, quantitative, measures of image quality. We have discussed multiple channel models of the HVS and their relationship to the wavelet transform. We highlighted two failings of the wavelet transform as a model of the HVS, translational variance and poor orientation sensitivity. We introduced the concept of visible difference prediction and described a VDP based on the wavelet transform. The WVDP based on an overcomplete wavelet transform was then demonstrated to reliably predict visible differences between an original and noisy image. This WVDP can then be applied directly to a wavelet image compression scheme to control coefficient quantization. However, whether this model is best applied to the critically sampled wavelet coefficients directly, with a corresponding reduction in reliability, or indirectly via the overcomplete transform is still a matter for further research.

Acknowledgment

The author wishes to thank N. Roeder, X. Li, and J. Andrew for useful discussions during the course of this research, and the anonymous referees for helpful comments on earlier drafts of this paper.

References

[1] A. J. Ahumada aud C. H. Null, "Image quality: A multidimensional problem," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 141–148.
[2] M. G. Albanesi, "Wavelets and human visual perception in image compression," in *Proc. Image Processing and its Applications*, 1995, pp. 859–863.
[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] S. Bertoluzza and M. G. Albanesi, "On the coupling of human visual system model and wavelet transform for image compression," *SPIE*, vol. 2303, pp. 389–397, 1994.
[5] R. J. Clarke, *Transform Coding of Images*. New York: Academic, 1985.
[6] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 45, pp. 485–560, 1992.
[7] S. Daly, "The visible difference predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 179–206.
[8] A. M. Eskicioglu and P. S. Fisher, "A survey of quality measures for grey scale image compression," in *Proc. NASA Space Earth Science Data Compression Workshop*, 1993, pp. 49–61.
[9] L. Gaudart, T Grebassa, and J. P. Petrakian, "Wavelet transform in human visual channels," *Appl. Opt.*, vol. 32, pp. 4119–4127, 1993.

[10] B. Girod, "The information theoretical significance of spatial and temporal masking in video signals," *SPIE, Human Vision, Visual Processing, and Display*, vol. 1077, pp. 178–187, 1989.
[11] ——, "What's wrong with mean-square error," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 207-220.
[12] W. E. Glenn, "Digital image compression based on visual perception," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 63–71.
[13] ——, "Review of video compression which depends on visual perception limits," *SPIE, High Definition Video*, vol. 1976, pp. 118–126, 1993.
[14] R. C. Gonzalez and R. E. Woods, *Digital Image Processing.* Reading, MA: Addison-Wesley, 1992.
[15] R. M. Gray, P. C. Cosman, and K. L. Oehler, "Incorporating visual factors into vector quantizers for image compression," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 35–51.
[16] D. H. Hubel, in *Eye, Brain and Vision.* New York: Scientific American Library, 1988.
[17] A. K. Jain, in *Fundamentals of Digital Image Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1989.
[18] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, pp. 1385–1421, 1993.
[19] Y. Kim *et al.*, "Wavelet transform image compression using human visual characteristics and a tree structure with a height attribute," *Opt. Eng.*, vol. 35, pp. 204–212, 1996.
[20] S. A. Klein, "Image quality and image compression: A psychophysicist's viewpoint," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 73–88.
[21] C. J. VDB. Larnbrecht and O. Verscheure, "Perceptual quality measure using a spatio temporal model of the human visual system," in *Proc. IS&T Symp. Electronic Imaging: Science and Technology*, San Jose, CA, Feb. 1996, vol. 2668.
[22] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 163–178.
[23] S. Mallat, "Wavelets for vision," *Proc. IEEE*, vol. 84, pp. 604–614, 1996.
[24] H. S. Malvar, "Lapped transforms for efficient transform/subband coding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, 1990.
[25] J. L. Mannos and D. J. Sakrison, "The effects of visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory*, vol. 20, pp. 525–536, 1974.
[26] H. Marmolin, "Subjective MSE measures," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-16, pp. 486–489, 1986.
[27] M. Miyahara, K. Kotani, and V. R. Algazi, "Objective picture quality scale (PQS) for image coding," *IEEE Trans. Commun.*, vol. 6, pp. 1215–1232, 1998.
[28] 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.
[29] N. B. Nill and B. H. Bouzas, "Objective image quality measure derived from digital image power spectra," *Opt. Eng.*, vol. 31, pp. 813–825, 1992.
[30] T. P. O'Rourke and R. L. Stevenson, "Human visual system based wavelet decomposition for image compression," *J. Vis. Commun. Image Represent.*, vol. 6, pp. 109–121, 1995.
[31] E. Peli, "Contrast in complex images," *J. Opt. Soc. Amer.*, vol. 7, pp. 2032–2040, 1990.
[32] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard.* New York: Van Nostrand Reinhold, 1993.
[33] G. C. Phillips and H. R. Wilson, "Orientation bandwidths of spatial mechanisms measured by masking," *J. Opt. Soc. Amer. A*, no. 1, pp. 226–232, 1984.
[34] C. A. Poynton, "Gamma and its disguises: The nonlinear mappings of intensity in perception, CRT's, film, and video," *J. Soc. Motion Picture Television Eng.*, pp. 1099–1108, 1993.
[35] K. Ramchandran and M. Wetterli, "Best wavelet packet bases in a rate distortion sense," *IEEE Trans. Image Processing*, vol. 2, pp. 160–175, 1993.
[36] N. A. Roeder, "Image compression and quality assessment: An investigation exploring wavelet packets and human visual system characteristics," M.S. thesis, Univ. Alberta, Edmonton, Alta., 1997.
[37] 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.
[38] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *Proc. ICASSP*, 1989, vol. 3, pp. 1945–1948.
[39] J. A. Saghri, P. S. Cheatham, and A. Habibi, "Image quality measure based on a human visual system model," *Opt. Eng.*, vol. 28, pp. 813–818, 1989.
[40] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, 1996.
[41] D. J. Sakrison, "On the role of the observer and a distortion measure in image transmission," *IEEE Trans. Commun.*, vol. COMM-25, pp. 1251–1267, 1977.
[42] R. Schafer, "Design of adaptive and NA-adaptive quantizers using subjective criteria," *Signal Process.*, vol. 5, pp. 333–345, 1983.
[43] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients (EZW)," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, 1993.
[44] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Hegger, "Shiftable multiscale transforms," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, 1992.
[45] H. L. Snyder, "Image quality: Meaures and visual performance," in *Flat Panel Displays and CRT's*, L. E. Tannas, Ed. New York: Van Nostrand Reinhold, 1985, pp. 70–90.
[46] C. F. Stromeyer and S. Klein, "Spatial frequency channels in human vision as asymmetric (edge) mechanisms," *Vis. Res.*, vol. 14, pp. 1409–1420, 1974.
[47] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proc. IEEE*, vol. 84, pp. 626–638, 1996.
[48] R. A. deVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Inform. Theory*, vol. 38, pp. 719–746, 1992.
[49] B. A. Wandell, *Foundations of Vision.* Sunderland, MA: Sinauer, 1995.
[50] A. B. Watson, "The cortex transform: Rapid computation of simulated neural images," *Comput. Vis., Graph., Image Process.*, vol. 39, pp. 311–327, 1987.
[51] ——, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE Human Vision, Visual Processing, and Digital Display IV*, 1993, pp. 1913–1927.
[52] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Processing*, vol. 6, no. 8, pp. 1164–1175, 1997.
[53] A. B. Watson, R. Borthwick, and M. Taylor, "Image quality and entropy masking," in *Proc. SPIE Human Vision, Visual Processing, and Digital Display VIII*, San Jose, CA, 1997,
[54] S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Perceptual image quality based on a multiple channel HVS model," in *Proc. ICASSP*, 1995, pp. 2351–2354.
[55] C. Zetzsche and G. Hauske, "Multiple channel model prediction of subjective image quality," in *Proc. SPIE, Human Vision, Visual Processing, and Display*, 1989, vol. 1077, pp. 209–215.
[56] C. Zetzsche, E. Barth, and B. Wegmann, "The importance of intrinsically two-dimensional image features in biological vision and picture coding," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 109–138.

Andrew P. Bradley (M'96) was born in Whitby, North Yorkshire, U.K., in 1966. He received the first class B.Eng. Honors degree in electrical and electronic engineering in 1989 from the University of Plymouth, U.K., and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Queensland, Australia, in 1996.

He worked as a Systems Engineer in northeast England until 1992. After completing the Ph.D. degree, he spent a year as a Post-doctoral Fellow in the Department of Computing Science, University of Alberta, Edmonton, Alta., Canada. He is currently a Research Engineer with Canon Information Systems Research Australia (CISRA), North Ryde, Sydney, Australia. His research interests include image analysis, machine learning and classification, the human visual system, scale-space and multiresolution methods, and image and video compression.

Dr. Bradley is an Associate Member of the Institution of Electrical Engineers and an active member of the Australian Pattern Recognition Society.