Visually Lossless Encoding for JPEG2000

Han Oh, Ali Bilgin, Senior Member, IEEE, and Michael W. Marcellin, Fellow, IEEE

Abstract—Due to exponential growth in image sizes, visually lossless coding is increasingly considered as an alternative to numerically lossless coding, which has limited compression ratios. This paper presents a method of encoding color images in a visually lossless manner using JPEG2000. In order to hide coding artifacts caused by quantization, visibility thresholds (VTs) are measured and used for quantization of subbands in JPEG2000. The VTs are experimentally determined from statistically modeled quantization distortion, which is based on the distribution of wavelet coefficients and the dead-zone quantizer of JPEG2000. The resulting VTs are adjusted for locally changing backgrounds through a visual masking model, and then used to determine the minimum number of coding passes to be included in the final codestream for visually lossless quality under the desired viewing conditions. Codestreams produced by this scheme are fully JPEG2000 Part-I compliant.

Index Terms—human visual system, contrast sensitivity function, visibility threshold, visually lossless coding, JPEG2000

I. INTRODUCTION

The contrast sensitivity function (CSF), which models the varying sensitivity of the human eye as a function of spatial frequency, has been widely exploited in many applications such as image/video quality assessment, perceptual image/video compression, and watermarking. The CSF shows some variations according to the age and visual acuity of the subject, the stimulus used in the experiment, and the viewing conditions [1], [2]. However, it is generally known that the sensitivity has a maximum at 2-6 cycles/degree and is limited at both high and very low frequencies. This is mainly due to the frequency-selective responses of the neurons in the primary visual cortex (also known as V1) and low-pass optical filtering [3].

The sensitivity profile can be measured through psychophysical experiments by finding the just-noticeable points (i.e., visibility thresholds) of the stimulus in a pre-defined contrast unit. A simple way of modeling contrast sensitivity is to use a sinusoidal grating such as the Campbell-Robson chart [4], while a more sophisticated two-dimensional CSF model can be obtained using stimuli that have specific ranges of frequency and orientation. Such stimuli are typically generated by transforms that simulate the V1 reception field of the human visual system (HVS). The frequency bandwidth to which a V1 neuron responds is known to be proportional to its center frequency, and its orientation bandwidth is distributed broadly around 40°. The V1 reception fields are well described by the Gabor filter. The Gabor filter is ideal for space-frequency localization; however, it is difficult to compute. Alternatively, Watson proposed the cortex transform, which is invertible, easy to implement, and able to precisely model V1 reception fields with adjustable parameter values [5].

Over-complete transforms such as the Gabor filter or cortex transforms are good tools for modeling the HVS, but they are not appropriate for image/video compression, since they increase the number of coefficients encoded [6]. As a result, in perceptual image/video compression, complete (critically-sampled) transforms such as the discrete cosine transform (DCT) or discrete wavelet transform (DWT), which are already embedded in the codec, are typically used for both modeling and coding.

JPEG2000 employs the separable DWT to decompose an image into several subbands, each having different spatial frequency and orientation [7]. As a decomposition method for simulating the HVS, the DWT has many desirable properties: linearity, invertibility, logarithmically spaced spatial frequencies, and four orientations of 0°, 90°, 45° and 135°. However, two of the orientations, 45° and 135°, are overlapped in the HH subband and the transform is shift-variant. Despite these shortcomings, the separable DWT is widely used in various vision models [8]–[11], as well as in JPEG2000.

In JPEG2000, a color image which has three color components (red, green, and blue) is typically transformed to an image with one luminance component (Y) and two chrominance components (Cb and Cr) for improved compression performance. Each component is then independently transformed to the wavelet domain by the DWT, and compression is achieved via bit-plane coding of wavelet coefficients quantized by a dead-zone quantizer. Coding artifacts in JPEG2000 are caused by errors in the reconstructed wavelet coefficients, induced by quantization. These coding artifacts are perceived differently based on the sensitivity of the HVS to the subband where the quantization distortion occurs.

In [12], visual weighting factors for wavelet subbands were derived from contrast sensitivity curves obtained from sinusoidal gratings. These weighting factors are included in JPEG2000 Part I as examples that can be used to aid in perceptual coding of images. The Kakadu implementation of JPEG2000 [13] can (optionally) use weighting factors to perform rate allocation in its post-compression rate-distortion optimization process. A visual masking process is described in [7] that can be implemented in a JPEG2000 Part I compatible manner, while JPEG2000 Part II includes visual masking tools based on the work reported in [14]. These tools have all proven useful to improve the visual quality of an image compressed at a given bitrate, but do not control the quality directly.

Work on direct quality control in wavelet image coding includes the work of Watson et al, where visibility thresholds...
(VTs) were measured for individual wavelet subbands using randomly generated uniform noise as a surrogate for quantization distortion [15]. The resulting VTs have since been employed in wavelet-based codecs (e.g., [16]). These codecs produce superior results when compared to codecs based on conventional MSE/PSNR metrics. However, as shown below, uniform noise does not well model the distortion produced by the dead-zone quantizer of JPEG2000. Thus, a direct use of these VTs in JPEG2000 may fail to produce a truly visually lossless encoded image. Indeed, rather than “visually lossless coding,” the work of [16] refers to “almost transparent coding.” For example, a mean opinion score of about 4.6 (out of 5) is reported for the Goldhill image in that work. Other efforts to measure VTs [17]–[19] have quantized actual wavelet coefficients obtained from natural images, rather than the synthetic distortions used by Watson. These natural VTs may be more accurate because they take into account spatial correlations between wavelet coefficients. However, these VTs were developed using a uniform quantizer. This paper shows below, that for use in JPEG2000, a distortion model based on a dead-zone quantizer yields superior results.

Other approaches to controlling JPEG2000 image quality directly (rather than controlling bit-rate) exist in the literature. For example, the use of post compression rate-distortion optimization to achieve a target MSE or PSNR is discussed in [7] and in [20]. Predicting MSE/PSNR prior to compression is discussed in [21]. In that work, predictions of MSE are based on target compression ratios and image activity measures which require significantly lower complexity than compression itself. This prediction could be used to choose a compression ratio (a priori) to achieve a desired MSE/PSNR. However, for the purpose of visually lossless compression, methods that target a given MSE/PSNR will not generally be optimal. This follows from the fact that the MSE/PSNR needed to achieve visually lossless coding is not known, and varies from image to image.

Building on our previous work [22]–[24], this paper proposes a visually lossless JPEG2000 encoder for color images. In the YCbCr color space, VTs are measured for a realistic quantization distortion model, which takes into account the statistical characteristics of wavelet coefficients as well as the dead-zone quantization of JPEG2000. The measured VTs are then adjusted to account for the visual masking effects of the underlying background image. The proposed coding scheme is implemented without violating the JPEG2000 Part I standard and automatically produces visually lossless imagery at lower bitrates than other visually lossless schemes from the literature.

This paper is organized as follows. The quantization distortions that occur in JPEG2000 are modeled in Section II. Section III describes a psychophysical experiment that employs the quantization model to determine VTs. Section IV describes a model for visual masking effects. The proposed visually lossless coding method and results are presented in Section V. Section VI summarizes the work.

II. QUANTIZATION DISTORTION MODELING

Wavelet coefficients are often modeled by a generalized Gaussian distribution [25] with probability density function

\[ f(y) = \frac{\alpha \cdot A(\alpha, \sigma)}{2\Gamma(1/\alpha)} \exp(-A(\alpha, \sigma)|y - \mu|^\alpha) \]  

where

\[ A(\alpha, \sigma) = \sigma^{-1} \left( \frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)} \right)^{1/2} \]

and \( \Gamma(\cdot) \) is the Gamma function. The parameters \( \mu \) and \( \sigma \) are the mean and standard deviation, respectively. The parameter \( \alpha \) is called the shape parameter. Wavelet coefficients for the HL, LH, and HH subbands, whose distributions have high-kurtosis and heavy-tailed symmetric densities, are well modeled by the Laplacian distribution with \( \mu = 0 \) and \( \alpha = 1 \), as shown in Fig. 1 (a) for a variance of \( \sigma^2 = 50 \).

JPEG2000 quantizes these wavelet coefficients using the following dead-zone quantizer:

\[ q = Q(y) = \text{sign}(y) \cdot \left| \frac{y}{\Delta} \right| \]  

Here, \( q \) is the quantization index which is subsequently encoded using embedded bit-plane coding. The dequantization procedure in the decoder is expressed by

\[ \hat{y} = Q^{-1}(q) = \begin{cases} 0 & q = 0 \\ \text{sign}(q) |q| + \delta \Delta & q \neq 0 \end{cases} \]

where \( \delta = 1/2 \) corresponds to mid-point reconstruction [26].

The resulting quantization distortions in the HL, LH, and HH subbands are not uniformly distributed over the interval \((-\Delta/2, \Delta/2)\) as is commonly assumed. A more appropriate model for quantization distortions produced by the dead-zone
quantizer and mid-point reconstruction is given by the PDF

\[
f(d) = \begin{cases} 
\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{d^2}{2\sigma^2}} + \frac{1-p_1}{\Delta} & 0 \leq |d| \leq \frac{\Delta}{2} \\
\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{d^2}{2\sigma^2}} & \frac{\Delta}{2} < |d| \leq \Delta \\
0 & \text{otherwise}
\end{cases}
\] (4)

where \( p_1 = \int_{-\Delta}^{\Delta} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{d^2}{2\sigma^2}} \, dy = 1 - e^{-\sqrt{2}\Delta/\sigma} \). The second term of the first line in (4) follows from assuming that the quantization distortion is uniform only for wavelet coefficients whose magnitudes are larger than \( \Delta \) (i.e., coefficients not in the dead-zone). The first term (in the first two lines of (4)) follows from the observation that the quantization errors of the remaining wavelet coefficients (i.e., those coefficients in the dead-zone interval \((-\Delta,\Delta))\) are equal to the coefficients themselves, since the dead-zone quantizer maps these coefficients to zero. Fig. 1 (b) shows the model of (4) corresponding to \( \Delta = 5 \) for the coefficient distribution shown in Fig. 1 (a).

Wavelet coefficients in the LL subband are often modeled by the Gaussian distribution with \( \mu = 0 \) and \( \alpha = 2 \) in (1). Assuming the standard deviation of the LL subband is large compared to the quantization step size results in the quantization distortion of the LL subband being modeled by

\[
f(d) = \begin{cases} 
\frac{1}{\sqrt{12\pi}\sigma} + \frac{1-p_2}{\Delta} & 0 \leq |d| \leq \frac{\Delta}{2} \\
\frac{1}{\sqrt{2\pi}\sigma} & \frac{\Delta}{2} < |d| \leq \Delta \\
0 & \text{otherwise}
\end{cases}
\] (5)

where \( p_2 = \Delta/\sqrt{3}\sigma \). This model is shown in Fig. 1 (c) for \( \Delta = 5 \) and \( \sigma^2 = 2000 \).

For comparison, Fig. 1 (b) and (c) also show the commonly used uniform distortion model. It is worth noting that the uniform model (incorrectly) implies that quantization errors are bounded by \( \pm\Delta/2 \), while the model of (4) correctly indicates that errors can be as large as \( \pm \Delta \). It is also important to note that the uniform model depends only on the parameter \( \Delta \), while the model of (4) depends on \( \Delta \) as well as the coefficient variance \( \sigma^2 \).

III. VISIBILITY THRESHOLDS FOR QUANTIZATION DISTORTIONS

For irreversible compression in Part I of the JPEG2000 standard, a color image having RGB components is first converted into an image with one luminance component (\( Y \)) and two chrominance components (\( Y_s \) and \( Y_c \)) by the irreversible color transform (ICT) [7]. Each component is then transformed using the dyadic Cohen-Daubechies-Feauveau (CDF) 9/7 DWT [27] and quantized independently. In what follows, a normalized CDF 9/7 DWT designed for efficient implementation of JPEG2000 is assumed. This transform preserves the nominal range of input values since the nominal gains of the analysis kernels are one (i.e., \( h_{n}^{\text{wavelet}} = \sum_{l=-4}^{4} h_{l}[n] = 1 \) and \( h_{n}^{\text{quant}} = \sum_{n=-3}^{3} (-1)^{n} h_{n}[n] = 1 \)). The normalization used in [15] produces wavelet coefficients \( 2^k \) times larger for level \( k \).

A \( K \) level dyadic wavelet decomposition has \( 3K + 1 \) subbands. Since \( K = 5 \) is usually sufficient to obtain near optimal compression performance [7], VTs are estimated for 16 subbands in each of the three color components for this work. The center spatial frequency \( f \) of level \( k \) is defined as

\[ f_k = r \cdot 2^{-k} = dv \tan \left( \frac{\pi}{180} \right) \cdot 2^{-k} \] (6)

where \( r \) is the visual resolution in pixels/degree, \( d \) is the display resolution in pixels/cm, and \( v \) is the viewing distance in cm [15].

A stimulus is an RGB image obtained by applying the inverse wavelet transform and the inverse ICT to wavelet data containing quantization distortions. In this work, stimulus generation begins with wavelet subband data corresponding to a \( 512 \times 512 \) color image. These initial data have all coefficients of all subbands (luminance and chrominance) set to 0. One subband of one component is then selected. Quantization distortion is randomly generated for this subband of interest according to either (4) or (5) depending on the subband. This distortion is added to a region of the subband. An example is shown in Fig. 2 (a) for a subband in the luminance component. In this figure, 0 is represented by mid-gray, while positive and negative values are lighter and darker, respectively. The inverse DWT of each component is followed by the inverse ICT. The distortion depicted in Fig. 2 (a) results in the example stimulus of Fig. 2 (b). In the wavelet domain, the size of the region to which distortion is added is \( N \times N \) with \( N = \min\{64, 512 \times 2^{-k}\} \) which corresponds to a JPEG2000 codeblock of nominal size \( 64 \times 64 \). This covers a supporting region, in the image domain, of nominal size \( M \times M \) with \( M = N \cdot 2^k \).

To measure the visibility threshold for a given subband with assumed subband variance \( \sigma^2_{\text{subband}} \), a two-alternative forced-choice (2AFC) method is used. In this method, an image that contains a stimulus and an image that does not contain a stimulus are presented to the subject. The experiment is iterated while varying the spatial frequency of the stimulus. The experiment is terminated when the subject is able to distinguish between the two images with 100% accuracy. This process is repeated for 32 iterations of the QUEST staircase procedure in the Psychophysics Toolbox [28] are used.
to determine the value of $\Delta$ corresponding to the 82% correct-point of a fitted Weibull function [15]. The obtained value of $\Delta$ is then the VT of the subband for the assumed coefficient variance $\sigma^2_b$. Different values of $\sigma^2_b$ generally result in different VTs as discussed in the next subsection.

The experimental environment was arranged similarly to a typical office environment. Stimuli were displayed on a LCD monitor in ambient light. In this threshold experiment, two LCD monitors, a Dell 1901FP and a Dell U2410, were used. The Dell 1901FP 19-in LCD monitor has a dot pitch of 0.294 mm, a resolution of $1280 \times 1024$, an image brightness of 250 cd/m$^2$, and a contrast ratio of 800:1. The Dell U2410 24-in LCD monitor has an In-Plane Switching (IPS) panel, a dot pitch of 0.27 mm, a resolution of $1920 \times 1200$, an image brightness of 400 cd/m$^2$, and a contrast ratio of 1000:1. Each monitor was connected to the PC through a Digital Visual Interface (DVI) cable. The reasons for using LCDs, unlike previous studies, are 1) the MTF (Modulation Transfer Function) of the rendering device is an important factor in determining thresholds, and LCDs have recently become the most common display device; and 2) a previous study revealed that stimuli are more detectable in LCDs than in CRTs [29]. The viewing distance was 60 cm (23.6 inches) with a resulting visual resolution of 35.62 pixels/degree and 38.72 pixels/degree for the Dell 1901FP and Dell U2410, respectively. This viewing distance was selected because it has been commonly assumed for typical viewing conditions in previous studies [15], [16], [30]. Based on this selection, the goal of this study is to achieve visually lossless performance under these conditions. However, the proposed methodology is general and can be adapted to yield visually lossless compression under other desired conditions.

A. Visibility Thresholds for the Luminance Component

Fig. 3 shows the visibility thresholds (i.e., the maximum quantization step sizes at which quantization distortions remain invisible) obtained for 5 levels of HL, LH, and HH subbands. The thresholds of Fig. 3 correspond to the case when $\sigma^2_b$ is fixed at 50. The $x$-axis and $y$-axis indicate the transform level and the obtained threshold, respectively. The points shown in Fig. 3 are the values determined by the QUEST procedure for individual subbands ((HH,1), (HH,2), and so on). Vertical bars indicate $\pm 1$ standard deviation. Two subjects, denoted by A and B, conducted the experiments using the Dell 1901FP monitor. Subject A also conducted the experiments using the Dell U2410 monitor. Both subjects are familiar with wavelet quantization distortion and have normal visual acuity. The results in Fig. 3 are labeled by the monitor and observer to which they correspond (e.g., U2410 A).

As can be seen in Fig. 3, all three monitor/observer combinations produced similar results. In most cases, intersecting the error intervals for the three measurements yields a non-empty set. To ensure a conservative design, the smallest of the three measured values is used as the visibility threshold for each subband in all subsequent discussions. As expected, threshold values in the HH subband are larger than those in the HL and LH subbands at the same level, which is known as the oblique effect [31]. Also, changes in threshold values as a function of spatial frequency are more pronounced in the HH subband than in other subbands. Thresholds in the HL and LH subbands are very similar. Hereafter, the HL and LH subbands are regarded as equivalent and one set of visibility thresholds is reported and used for both.

Changing the assumed variance of the wavelet coefficients can have a dramatic effect on the visibility thresholds. For example, Fig. 4 shows the measured thresholds, as a function of $\sigma^2_b$ for the (HL/LH,3) subband. The four data points denoted by asterisks represent thresholds measured for four different values of $\sigma^2_b$. The error bars represent $\pm 1$ standard deviation. From Fig. 4, it can be seen that the value of the threshold increases as the coefficient variance increases. This trend can be explained by a simple analysis of the quantization error.

Integrating the second line of (4), the probability of “large” distortions, which lie in $(-\Delta, -\Delta/2) \cup (\Delta/2, \Delta)$, is $e^{-\frac{2\Delta^2}{\Delta_b}} - e^{-\frac{2\Delta^2}{\Delta_b}}$ for the HL, LH, and HH subbands. These large distortions are caused by the dead-zone quantizer and tend to be more visible than the smaller distortions in $[-\Delta/2, \Delta/2]$. For a fixed value of $\Delta$, the probability of large distortions decreases as the variance $\sigma^2_b$ increases, provided that the step size $\Delta$ is sufficiently small (i.e., $\Delta < \sqrt{2} \sigma_b \ln 2$). Furthermore, the variance of the quantization distortion generated according to
The variance of the (LL,5) subband is usually much larger than that of the other subbands, and the distortion is not significantly affected by variance changes. This results in a fixed threshold of $t_{LL,5} = 0.63$ for the LL band.

### B. Visibility Thresholds for Chrominance Components

It is well known that the human visual system is less sensitive to quantization distortion in the chrominance components than in the luminance component. Also, the variance of the chrominance components is much smaller than that of the luminance component. The average variances calculated from 10 natural images (not used in subsequent calibration or evaluation) are listed in Table II. These images are available for viewing at [32]. Table III contains measured threshold values for the chrominance components, determined in the same fashion as for the luminance component using the values from Table II for $\sigma_b^2$ when generating the stimulus images. Experimental results indicate that the chrominance thresholds are insensitive to variance changes. Thus, the thresholds for the chrominance components are fixed to the values given in Table III. From this table, it can be seen that threshold values for the chrominance components are larger than those for the luminance component, and that the Cb component is the least sensitive component.

### IV. VISIBILITY THRESHOLD ADJUSTMENT BASED ON VISUAL MASKING EFFECTS

In actual image coding, unlike the psychophysical experiments described above, all subbands are quantized simultaneously and quantization distortion is superimposed on a “background” image. Threshold changes for the compound

---

Table I: Linear Parameters $a_b$ and $r_b$

<table>
<thead>
<tr>
<th>subband</th>
<th>$a_b$</th>
<th>$r_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL1</td>
<td>$105.67 \times 10^{-4}$</td>
<td>4.85</td>
</tr>
<tr>
<td>HL2</td>
<td>$19.94 \times 10^{-4}$</td>
<td>0.92</td>
</tr>
<tr>
<td>HL/LH1</td>
<td>$46.03 \times 10^{-4}$</td>
<td>1.98</td>
</tr>
<tr>
<td>HL/LH2</td>
<td>$13.84 \times 10^{-4}$</td>
<td>0.64</td>
</tr>
<tr>
<td>HL/LH3</td>
<td>$10.83 \times 10^{-4}$</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Table II: Average Variances of Wavelet Coefficients for Chrominance Components

<table>
<thead>
<tr>
<th>subband</th>
<th>Cb</th>
<th>Cr</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL1</td>
<td>1.33</td>
<td>1.43</td>
</tr>
<tr>
<td>HL/LH1</td>
<td>0.72</td>
<td>0.74</td>
</tr>
<tr>
<td>HL/LH2</td>
<td>3.06</td>
<td>3.52</td>
</tr>
<tr>
<td>HL/LH3</td>
<td>1.16</td>
<td>1.20</td>
</tr>
<tr>
<td>LL,5</td>
<td>150.08</td>
<td>109.51</td>
</tr>
</tbody>
</table>

---

1In the case of (HH,1), (HL/LH,1) and (HH,2), the measured thresholds were infinite. Adopting this result would effectively result in chrominance subsampling. Instead, large but finite values were chosen in Table III to provide a more conservative approach to these subbands, resulting in negligible file size increases.
distortions caused by quantizing subbands simultaneously have been studied in [18]. Visibility of quantization distortions from similar subbands is known to linearly increase when the distortion is highly visible. However, such compound distortions are negligible at the thresholds chosen for this work, and therefore threshold adjustments for compound distortion are not needed. This result is consistent with [15], which provides individual quantization step sizes for visually lossless compression without considering other subbands.

On the other hand, VTs can vary significantly with image background. In this context, the image background is called the masker, while the distortion is referred to as the target. The change of threshold values according to the contrast of the image background is represented by the target threshold versus masker contrast (TvC) function [33]. As the contrast of the masker increases, the threshold decreases slightly and then begins to increase. The target becoming less visible due to a high-contrast masker is called masking, while the opposite effect, induced by a low-contrast background is called facilitation. The effect of facilitation is generally insignificant, and only the masking effect is considered in this work. In what follows, visibility thresholds are modified to exploit the masking effect. In particular, thresholds are increased within spatial regions depending on the relevant image background. These threshold increases decrease file size without introducing visible distortion increases.

In coding of color images, the chrominance components typically occupy a much smaller portion than the luminance component in the overall file. Thus, for simplicity, threshold adjustments are made only for the luminance component in this work.

The contrast of a luminance masker within a spatial region is determined by the luminance wavelet coefficients in the corresponding spatial region of all subbands. However, since masking occurs most strongly when the target and masker have the same frequency and orientation, this work considers only wavelet coefficients in the subband to be encoded, and the visually lossless masked threshold \( t_b \) is determined as

\[
\hat{t}_b = t_b \cdot m_b
\]

where \( t_b \) is the base threshold value calculated from (7) and \( m_b \) is a masking factor calculated from the magnitudes of wavelet coefficients in subband \( b \). This intra-band masking model may provide slightly lower accuracy than models that consider both intra- and inter-band masking effects, but has the advantage of simplicity, as well as enabling parallel processing because subbands can be encoded independently.

### Table III

<table>
<thead>
<tr>
<th>subband</th>
<th>( C_b )</th>
<th>( C_r )</th>
<th>subband</th>
<th>( C_b )</th>
<th>( C_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH,1</td>
<td>24.72</td>
<td>15.49</td>
<td>HH,4</td>
<td>4.95</td>
<td>1.25</td>
</tr>
<tr>
<td>HL/LH,1</td>
<td>14.50</td>
<td>6.35</td>
<td>HL/LH,4</td>
<td>3.26</td>
<td>0.69</td>
</tr>
<tr>
<td>HH,2</td>
<td>14.77</td>
<td>7.40</td>
<td>HH,5</td>
<td>1.05</td>
<td>0.56</td>
</tr>
<tr>
<td>HL/LH,2</td>
<td>6.36</td>
<td>2.60</td>
<td>HL/LH,5</td>
<td>1.05</td>
<td>0.58</td>
</tr>
<tr>
<td>HH,3</td>
<td>11.55</td>
<td>2.57</td>
<td>LL, 5</td>
<td>1.32</td>
<td>0.66</td>
</tr>
<tr>
<td>HL/LH,3</td>
<td>4.03</td>
<td>1.23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The masking factor \( m_b \) is calculated using two visual masking models; the self-contrast masking model and the texture masking model. The self-contrast masking model approximates the change of threshold in the TvC function according to the magnitude of wavelet coefficients. The corresponding self-contrast masking factor, \( s_b[n] \), at two-dimensional location \( n \) in subband \( b \) is defined by

\[
s_b[n] = \max \left\{ 1, w_1 \left( \frac{|y[n]|}{\hat{y}_b + \epsilon} \right)^{\rho_1} \right\}
\]

where \( y[n] \) is the wavelet coefficient at location \( n \) and \( \hat{y}_b \) is the conditional expectation of \( y \) given \( y \geq 0 \), calculated from (4) as \( E[y|y \geq 0] = \sqrt{2 \sigma_b}/2 \). The small constant \( \epsilon > 0 \) is included for stability of the equation. The parameter \( \rho_1 \) reflects nonlinearity of self-contrast masking and has a value between 0 and 1. The parameter \( w_1 \) adjusts the degree of self-contrast masking and together with \( \rho_1 \) prevents over-masking.

The self-contrast masking factor, as shown in Fig. 6, takes a value of 1 (no masking) for small values of \( |y[n]| \). On a log-log scale, it increases linearly with slope \( \rho_1 \) when \( \log |y[n]| \) exceeds \( \log \hat{y}_b \) by more than \( \log w_1^{-1/\rho_1} \). This self-contrast masking model can have problems at edges, because edges generate large coefficients, but are perceptually more sensitive [14]. Therefore, care is needed in choosing the parameters \( w_1 \) and \( \rho_1 \) to ensure visually lossless quality in all regions.

In addition to self-contrast, texture activity can significantly affect distortion visibility. Specifically, VTs increase as the texture beneath the distortion becomes more difficult to predict. In this work, the texture masking factor, \( \tau_b[n] \), for the texture activity of small local texture-block \( j \) is given by

\[
\tau_b[n] = \max \left\{ 1, w_2 \left( \frac{\sigma_j^2}{\hat{\sigma}_b^2} \right)^{\rho_2} \right\}
\]

where \( \sigma_j^2 \) is the variance of reconstructed wavelet coefficients in texture-block \( j \). The parameters \( w_2 \) and \( \rho_2 \) play similar roles to those of \( w_1 \) and \( \rho_1 \) in (9).

The texture masking model is similar to the neighbor masking models in [16] and [34], but computes the masking factor only once per texture-block for computational efficiency. Every location in the texture-block is then assigned the same

![Fig. 6. Log-log plot of self-masking factor vs. wavelet coefficient magnitude for \( \sigma_b^2 = 50 \), \( w_1 = 1 \) and \( \rho_1 = 0.4 \).](image-url)
value. A texture-block of size $W \times W$ in the spatial domain implies a texture-block of size $N \times N$ in the wavelet domain, with $N = W \cdot 2^{-k}$ for DWT level $k$. Since $N$ is usually smaller than the size of a JPEG2000 codeblock, each codeblock can contain several texture-blocks. Note that although the texture masking factor is constant over a texture-block, it is not generally constant over a codeblock. Textural masking is applied only to the high-frequency subbands ($k \leq 3$) because lower frequency subbands have insufficiently large texture-blocks to calculate a variance. Because textures can be affected by quantization, reconstructed wavelet coefficients are used when calculating the variance $\sigma^2$ [35].

Fig. 7 depicts the image horse and the product of the two masking factors, $s_b[n] \cdot \tau_b[n]$, in the wavelet domain. Brighter intensities represent higher masking effects.


g_{\text{visual}}(x) = \max \{ g(x), \text{value} \}
\end{equation}

where $g(x)$ is the image intensity and value is a threshold. The value of $g(x)$ is compared with value to determine the visual quality of the image.

V. VISUALLY LOSSLESS CODING AND EXPERIMENTAL RESULTS

A. Visually Lossless Coding

In JPEG2000, wavelet coefficients in subband $b$ are initially quantized with a quantization step size $\Delta b$. The subband is partitioned into codeblocks, and the quantization indices of each codeblock are bit-plane coded. Each bit-plane is coded in three coding passes, except the most significant bit-plane (MSB), which is coded in one coding pass. Thus, a codeblock with $M$ bit-planes has $3M - 2$ coding passes. One or more least significant coding passes may optionally be omitted from the final codestream to increase the effective quantization step size within a codeblock. Omitting the $p$ least significant bits of the quantization index of a coefficient results in an effective quantization step size of $2^p \Delta b$ for that coefficient.

To ensure that the effective quantization step sizes are less than the thresholds derived previously, the following procedure is followed for all luminance subbands except (LL,5). First, the variance $\sigma^2_{b,i}$ for codeblock $i$ in subband $b$ is calculated. Then a base threshold $t_{b,i}$ for that codeblock is determined using (7). The self-masking factor $s_{b,i}[n]$ is calculated for each wavelet coefficient in the codeblock according to (9). During the bit-plane coding, the texture-masking factor $\tau_{b,i}[n]$ is calculated for each coefficient in the codeblock according to (10), followed by $m_{b,i} \left( (11) \right)$ and $t_{b,i} \left( (8) \right)$. The maximum absolute error for the codeblock is calculated at the end of each coding pass $z$ as

\begin{equation}
D(z) = \max_{n \in B} \left( |y[n] - \hat{y}(z)[n]| \right) \tag{12}
\end{equation}

where $\hat{y}(z)[n]$ denotes the reconstructed value of $y[n]$ using the quantization index $\hat{q}(z)[n]$, which has been encoded only up to coding pass $z$. Coding is terminated when $D(z)$ falls below the masked threshold $t_{b,i}$.

In principle, $\tau_{b,i}[n]$ must be recalculated at the end of each coding pass, since it is computed using reconstructed coefficients. In practice, to reduce the computation load, the texture masking factors are not evaluated (and are set to 1.0) until $D(z)$ falls below a prescribed maximum masked threshold, $t_{b,max}$. In addition to complexity reduction, $t_{b,max}$ sets a conservative upper bound on the effective quantization step size that can be applied to any given codeblock, possibly at the expense of some small increase in file size.

Pseudo-code for the proposed algorithm is given in Fig. 8. It is worth emphasizing that the base VTs computed in the second line of Fig. 8 are adaptive codeblock-by-codeblock and image-by-image. This is in contrast to [16] where a fixed set of base VTs (one per subband) is employed, independent of the image to be encoded. The adaptivity of the base VTs proposed here stems from the dependence of the quantization distortion model on coefficient variance.

For (LL,5) of the luminance component, the visibility threshold $t_{LL,5} = 0.63$ is used directly as the quantization step size $\Delta_{LL,5}$, and all bit-planes are included in the codestream. The same procedure is applied for all chrominance subbands. That is, the thresholds of Table III are used as quantization step sizes, and all bitplanes are included in the codestream.

Fig. 7. (a) 512 × 512 reference image horse and (b) the product of its two masking factors, $s_b[n] \cdot \tau_b[n]$, in the wavelet domain for $W = 32$. Brighter intensities represent higher masking effects.
The proposed algorithm does not require post-compression rate-distortion (PCRD) optimization [7] to select which coding passes are included in the final codestream. The algorithm automatically computes only compressed coding passes that will be included in the final codestream. Thus, as in [16], the “overcoding” that is typical of many JPEG2000 implementations is avoided, which can significantly reduce computational complexity. It is important to note that all codestreams produced by the proposed algorithm are compatible with Part-I of the JPEG2000 standard, and can be decoded by any JPEG2000 decoder.

As mentioned previously, prior to bit-plane coding, all wavelet coefficients in luminance subband b are quantized using a step size of $\Delta_b$. An ideal value of $\Delta_b$ would be the final masked threshold. However, this value is unknown before encoding and varies from codeblock to codeblock. A simple strategy would be to select $\Delta_b$ as the minimum possible value of the final masked threshold, i.e., $\tau_b$ in Table I. In this work, a slightly larger value of $\Delta_b = 1.03\tau_b$ was chosen. This choice yields slightly better compression performance with no impact on visual quality. Indeed, this value is well within the error bars depicted in Fig. 4 for all subbands.

Calibration of the masking models was performed using several 8-bit monochrome images. The purpose of this calibration was to find a set of model parameters which yields the largest masking factors (i.e., the smallest file sizes), while maintaining visually lossless quality via compression as described above, followed by decomposition using an unmodified JPEG2000 Part I decoder. Table IV shows the parameter values obtained by visual inspection using numerous combinations of values. As verified by validation experiments described below, these parameter values provide visually lossless quality for a wide range of natural images. Calibration by image type is possible, and different parameter values for different image types may provide better compression performance, but at the expense of the compressor becoming image type specific.

### B. Coding Results

The proposed encoder was implemented in Kakadu V6.1 [13]. All reported results were generated with this encoder and the unmodified decoder, kduexpand. Tables V and VI, respectively, show the bitrates obtained for encoding 8-bit monochrome and 24-bit color images using the proposed JPEG2000 coding scheme. For comparison, bitrates obtained for numerically lossless JPEG2000 compression are also included. Images used to generate these results were obtained from the USC database [36], LIVE database [37], Kodak PhotoCD [38], and ISO JPEG2000 test suite. Images marked with an asterisk in Table V were used during masking model calibration.

For monochrome images, the numerically lossless coding method of JPEG2000 yields an average bitrate of 4.80 bits-per-pixel (bpp), while the proposed visually lossless coding method achieves an average bitrate of 1.91 bpp, an improvement in compression ratio of 2.5 to 1, without perceivable quality degradation. In the case of color images, the numerically lossless coding method and the proposed visually lossless coding method achieve, respectively, 10.50 bpp and 1.91 bpp on average, for an improvement in compression ratio of 5.5 to 1.

As can be seen in the tables, different images encoded in the proposed visually lossless manner have significantly different bitrates. Also, peak-signal-to-noise ratios for the luminance component (Y PSNRs) vary widely depending on the image.
TABLE VI
BITRATES AND PSNRs FOR THE PROPOSED VISUALLY LOSSLESS JPEG2000 ENCODER FOR 24-BIT COLOR IMAGES

<table>
<thead>
<tr>
<th>Image</th>
<th>Dimension (W×H)</th>
<th>Lossless (bpp)</th>
<th>Proposed (bpp)</th>
<th>Ratio</th>
<th>Y PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bike</td>
<td>2048×2048</td>
<td>11.96</td>
<td>1.72</td>
<td>6.95</td>
<td>40.44</td>
</tr>
<tr>
<td>bikes</td>
<td>768×512</td>
<td>10.81</td>
<td>2.40</td>
<td>4.51</td>
<td>39.53</td>
</tr>
<tr>
<td>building2</td>
<td>640×512</td>
<td>13.90</td>
<td>2.79</td>
<td>4.99</td>
<td>36.06</td>
</tr>
<tr>
<td>buildings</td>
<td>768×512</td>
<td>11.13</td>
<td>2.30</td>
<td>4.84</td>
<td>37.96</td>
</tr>
<tr>
<td>caps</td>
<td>768×512</td>
<td>8.09</td>
<td>1.10</td>
<td>7.35</td>
<td>43.52</td>
</tr>
<tr>
<td>carnival</td>
<td>610×488</td>
<td>9.67</td>
<td>1.75</td>
<td>5.53</td>
<td>41.40</td>
</tr>
<tr>
<td>cemetery</td>
<td>627×482</td>
<td>12.69</td>
<td>2.37</td>
<td>5.35</td>
<td>38.86</td>
</tr>
<tr>
<td>church</td>
<td>634×505</td>
<td>10.98</td>
<td>1.85</td>
<td>5.93</td>
<td>39.96</td>
</tr>
<tr>
<td>coins</td>
<td>640×512</td>
<td>11.88</td>
<td>2.35</td>
<td>5.05</td>
<td>40.39</td>
</tr>
<tr>
<td>dancers</td>
<td>618×453</td>
<td>11.98</td>
<td>2.28</td>
<td>5.26</td>
<td>38.82</td>
</tr>
<tr>
<td>flowers</td>
<td>640×512</td>
<td>12.36</td>
<td>2.36</td>
<td>5.23</td>
<td>38.47</td>
</tr>
<tr>
<td>house</td>
<td>768×512</td>
<td>10.08</td>
<td>1.99</td>
<td>5.06</td>
<td>41.22</td>
</tr>
<tr>
<td>lighthouse</td>
<td>480×720</td>
<td>10.19</td>
<td>1.73</td>
<td>5.88</td>
<td>40.11</td>
</tr>
<tr>
<td>lighthouse2</td>
<td>768×512</td>
<td>9.75</td>
<td>1.71</td>
<td>5.72</td>
<td>40.26</td>
</tr>
<tr>
<td>manfish</td>
<td>634×438</td>
<td>10.06</td>
<td>1.79</td>
<td>5.61</td>
<td>40.08</td>
</tr>
<tr>
<td>monarch</td>
<td>768×512</td>
<td>8.98</td>
<td>1.21</td>
<td>7.41</td>
<td>42.83</td>
</tr>
<tr>
<td>ocean</td>
<td>768×512</td>
<td>8.77</td>
<td>1.50</td>
<td>5.83</td>
<td>41.71</td>
</tr>
<tr>
<td>painted</td>
<td>768×512</td>
<td>10.16</td>
<td>2.14</td>
<td>4.74</td>
<td>40.43</td>
</tr>
<tr>
<td>parrots</td>
<td>768×512</td>
<td>8.48</td>
<td>0.94</td>
<td>9.06</td>
<td>43.46</td>
</tr>
<tr>
<td>plane</td>
<td>768×512</td>
<td>8.06</td>
<td>1.21</td>
<td>6.64</td>
<td>43.09</td>
</tr>
<tr>
<td>rapid</td>
<td>768×512</td>
<td>10.16</td>
<td>2.35</td>
<td>4.33</td>
<td>40.54</td>
</tr>
<tr>
<td>sailing1</td>
<td>768×512</td>
<td>9.59</td>
<td>1.91</td>
<td>5.02</td>
<td>39.98</td>
</tr>
<tr>
<td>sailing2</td>
<td>480×720</td>
<td>9.39</td>
<td>1.22</td>
<td>7.67</td>
<td>41.85</td>
</tr>
<tr>
<td>sailing3</td>
<td>480×720</td>
<td>9.57</td>
<td>1.39</td>
<td>6.87</td>
<td>41.96</td>
</tr>
<tr>
<td>sailing4</td>
<td>768×512</td>
<td>9.29</td>
<td>1.91</td>
<td>4.86</td>
<td>41.26</td>
</tr>
<tr>
<td>statue</td>
<td>480×720</td>
<td>9.55</td>
<td>1.52</td>
<td>6.27</td>
<td>41.70</td>
</tr>
<tr>
<td>stream</td>
<td>768×512</td>
<td>11.86</td>
<td>2.84</td>
<td>4.17</td>
<td>36.39</td>
</tr>
<tr>
<td>sculpture</td>
<td>632×505</td>
<td>13.31</td>
<td>2.83</td>
<td>4.71</td>
<td>37.87</td>
</tr>
<tr>
<td>woman</td>
<td>2048×2560</td>
<td>11.50</td>
<td>1.67</td>
<td>6.90</td>
<td>40.57</td>
</tr>
<tr>
<td>woman2</td>
<td>480×720</td>
<td>11.61</td>
<td>2.38</td>
<td>4.87</td>
<td>39.88</td>
</tr>
<tr>
<td>womanhat</td>
<td>480×720</td>
<td>9.75</td>
<td>1.67</td>
<td>5.82</td>
<td>41.76</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>10.50</td>
<td>1.91</td>
<td>5.50</td>
<td>40.34</td>
</tr>
</tbody>
</table>

In particular, the resulting bitrates range from 0.91 bpp to 2.97 bpp for monochrome images, and from 0.94 bpp to 2.84 bpp for color images. The monochrome images have a minimum PSNR of 35.41 dB and a maximum PSNR of 44.60 dB. The luminance PSNR values for color images are similar. These results demonstrate that bitrate and PSNR are not effective criteria for determining visual lossless quality.

Table VII compares the proposed coding method with the visually lossless method proposed in [30]. This table reports bitrates obtained by each method for encoding five 512 × 512 8-bit digitized radiographs used in [30]. As described in that work, the images were obtained by cropping 512 × 512 diagnostically relevant regions of 2048 × 2056 radiographs acquired using a Lumisys 75 laser film scanner (8-bits/pixel grayscale, 146 dots/inch). The five images are images of the hip, chest, neck, leg, and abdomen. As can be seen from the table, the method proposed here provides significantly lower bitrates. These lower bitrates result in a lower average PSNR, but do not result in visual artifacts.

Comparison with other methods from the literature is more difficult. As compared to [16], the bitrates obtained here for monochrome images are considerably higher. However, as discussed in the introduction, the encoder of [16] does not provide truly visually lossless quality as evidenced by their use of the term “almost transparent coding” rather than “visually lossless coding,” and by their reported mean opinion scores. This is likely caused by the use of a uniform probability model for quantization distortion, as well as the use of a CRT for the psychovisual experiments in [15], from where the VTs used in [16] are drawn.

As also mentioned in the introduction, the Kakadu implementation of JPEG2000 provides a CSF-based visual weighting option that can enhance the visual quality of compressed imagery. Command line options for the application kdu_compress allow the user to select a target bitrate, or alternatively, a target distortion-rate slope. But, no guidance is given as to how a target bitrate or distortion-rate slope should be chosen to yield visually lossless quality. If an “Oracle” provides the bitrates from the proposed method (from Tables V and VI) as target bitrates to kdu_compress (with visual weighting), very high quality imagery results. In fact, the quality rivals that of the proposed method. However, it must be emphasized that kdu_compress has no such Oracle, and that a major contribution of the proposed method is that it automatically adjusts the compression on an image-by-image basis to achieve low bitrates while preserving visually lossless quality.

It is worth noting that the base visibility thresholds can be used alone (without masking) to achieve visually lossless encoding. Indeed, results for this case were reported in [22]. Specifically, for a given codeblock, the coefficient variance is computed and used to compute the base threshold \( b_0 \), from (7). Bitplane coding is then carried out until the maximum absolute coefficient error in the codeblock falls below \( b_0 \). Visually lossless encoding is achieved, but with about 10% larger file sizes than in the masked case. The masked case can be understood to be uniformly more aggressive than the unmasked case. This is seen by noting that the masking model can only increase the base threshold. That is, from (8) and (11), it follows that \( b_2 ≥ b_0 \). Thus, for each codeblock, the number of coding passes included in the masked case is no more than in the unmasked case. In turn, the quantization error of each coefficient in the masked case is lower bounded by that in the unmasked case. Accordingly, only the masked version of the encoder is considered in the validation experiments below.

C. Validation Experiments

To verify that the images encoded with the proposed scheme are visually lossless, a three-alternative forced-choice (3AFC)
method was used. For each test image, two original copies and one compressed copy were displayed side by side with the position of the compressed copy chosen randomly. The test images were selected for presentation in a fixed order. For each test image, the subject was given an unlimited amount of time to choose the copy that looked different from the other two copies. If the images are compressed in a visually lossless manner, the rate at which the subject correctly chooses the compressed copy should be 1/3.

3AFC testing has been used previously in [30]. Evidently, 3AFC testing can provide a more rigorous validation than 2AFC. This can be seen by considering the 2AFC procedure.

In 2AFC testing, an original copy is displayed together with only one compressed copy. The subject is asked to choose the copy which has been compressed. If the images are compressed losslessly, the rate at which the subject chooses the correct copy should be 1/2. However, consider the case when the images are compressed with very high quality, but are not visually lossless. In the case of 2AFC testing, the subject would be able to perceive that differences exist between the two copies, but might not be able to tell which copy has been compressed. In this situation, the rate of correct response might still be 1/2, even though the images are not visually lossless. On the other hand, for 3AFC testing, the subject can identify
the compressed image as the one that differs from the other two, resulting in a correct response rate of greater than 1/3, indicating the fact that the imagery is not visually lossless.

As test images, 8 monochrome and 11 color images were selected from the images in Tables V and VI. Each has a low luminance PSNR (less than 40 dB) and none were used during algorithm design/calibration. Additionally, the two radiographs from Table VII having the lowest PSNRs were included in the validation test set. In the validation experiment, the Dell U2410 LCD monitor with a resolution of 1920 × 1200 was used. The maximum dimension of images displayed during testing was 570 × 975 so that three copies fit side by side across the screen. Test images exceeding these maximum dimensions were cropped after decompression. Sixteen observers participated in the experiment. All were familiar with compression artifacts that occur in JPEG2000 and have normal or corrected-to-normal vision. Each subject performed five trials for each of the 21 images for a total of 1680 evaluations. During the experiment no feedback was provided on the correctness of choices. Although the proposed encoder was designed to achieve visually lossless quality at a typical viewing distance of 60 cm, observers were allowed to view the images as closely as desired.

Under the null hypothesis that the three images are indistinguishable, the compressed image should be chosen correctly with a frequency of 1/3. Images compressed using the proposed method were selected with a mean frequency of 0.3494, and a standard deviation of 0.041. Table VIII shows the results of a one-sample t-test that measures whether the mean frequency for images encoded using the proposed method significantly differs from the hypothesized frequency of 1/3. In this test, the hypothesis that the responses were randomly chosen could not be rejected at the 5% significance level. Based on these results, it is claimed that the proposed coding method is visually lossless under the viewing conditions set during the design and validation studies.

Fig. 9 shows an original monochrome image together with a version that has been encoded by the proposed method. The encoded image exhibits a PSNR of 39.53 dB, and no differences are visible when the images are displayed using a 1:1 scale. Because PDF viewers may inappropriately scale images according to window size or user settings, and unintended distortions may be introduced during PDF creation, more examples of encoded images (both monochrome and color) are provided online [32] for proper comparison.

<table>
<thead>
<tr>
<th>Test Value</th>
<th>Degrees of freedom</th>
<th>Significance (2-tailed)</th>
<th>Mean difference</th>
<th>95% confidence interval of the difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.333</td>
<td>20</td>
<td>0.087</td>
<td>0.0167</td>
<td>-0.0026 to 0.0347</td>
</tr>
</tbody>
</table>

VI. Conclusions

The HVS has varying sensitivity to different color components, spatial frequencies, orientations, and underlying background images. Using this fact, a visually lossless coding algorithm has been presented for 8-bit monochrome and 24-bit color images. Visibility thresholds were measured for statistically modeled quantization distortion, and have different values depending on the local variances within each subband. Since quantization distortion appears on various background images, the threshold values are adjusted using a self-masking model and a texture masking model to cope with spatially changing visual masking effects. The resulting thresholds are used to determine the maximum quantization for visually lossless coding. The proposed JPEG2000 Part I compliant coding scheme successfully yields compressed images, whose quality is indistinguishable to those of the original images, at significantly lower bitrates than those of numerically lossless coding and other visually lossless algorithms in the literature.

To improve the performance of the proposed algorithm, the distribution model for wavelet coefficients could be replaced by a more sophisticated model such as Gaussian Scale Mixture (GSM) [39], which takes into account spatial correlation between wavelet coefficients. However, a generalized Gaussian model that only uses variance as a parameter was used here to avoid the intricacy of a multi-variable experiment. Also, to preserve independent coding of codeblocks, light adaptation using the LL subband [16] and inter-band masking effects were excluded. Nevertheless, the proposed coding scheme well predicts the appropriate compression parameters required for visually lossless coding.

References


