A Methodology for Visually Lossless JPEG2000 Compression of Monochrome Stereo Images

Hsin-Chang Feng, Michael W. Marcellin, *Fellow, IEEE*, and Ali Bilgin, *Senior Member, IEEE*

***Abstract*— A methodology for visually lossless compression of monochrome stereoscopic 3D images is proposed. Visibility thresholds are measured for quantization distortion in JPEG2000. These thresholds are found to be functions of not only spatial frequency, but also of wavelet coefficient variance, as well as the gray level in both the left and right images. To avoid a daunting number of measurements during subjective experiments, a model for visibility thresholds is developed. The left image and right image of a stereo pair are then compressed jointly using the visibility thresholds obtained from the proposed model to ensure that quantization errors in each image are imperceptible to both eyes. This methodology is then demonstrated via a particular 3D stereoscopic display system with an associated viewing condition. The resulting images are visually lossless when displayed individually as 2D images, and also when displayed in stereoscopic 3D mode.**

***Index Terms*— Stereoscopic images, visually lossless coding, JPEG2000, crosstalk, human visual system.**

I. Introduction

STEREOSCOPIC 3D imaging has been applied in diverse fields such as aerial stereo photography, stereoscopic surgery, and digital cinema [1]–[3]. Accordingly, it has received considerable attention over the last few decades. Recently, consumers are viewing many different types of stereoscopic content in television and gaming applications due to inexpensive consumer-grade 3D displays becoming widely available.

In order to simulate the parallax between the left and right eyes of a human observer, left and right images of a scene are taken from two different positions. These two images make up a stereo pair (stereo image). In a stereo viewing system, the left and right images of a stereo pair are presented to the left and right eyes of the observer, respectively. The human brain fuses the two images to create the perception of 3D [4].

Manuscript received March 13, 2014; revised August 7, 2014 and November 19, 2014; accepted December 6, 2014. Date of publication December 18, 2014; date of current version January 8, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Anthony Vetro.

H.-C. Feng and M. W. Marcellin are with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721 USA (e-mail: feng1@email.arizona.edu; mwm@email.arizona.edu).

A. Bilgin is with the Department of Biomedical Engineering, the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721 USA (e-mail: bilgin@email.arizona.edu).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes stereo images. The total size of the file is 85.3 MB. Contact feng1@email.arizona.edu, mwm@email.arizona.edu, or bilgin@email.arizona.edu for further questions about this work.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2014.2384273

Typical display technologies for stereoscopic 3D viewing employ passive or active glasses, although other technologies are currently available [5], [6]. For passive systems, the left and right images are displayed sequentially on a display which polarizes them clockwise and anti-clockwise, respectively. Polarized glasses enable only the correct image to pass through to each eye. For active display technologies, active shutter glasses are synchronized to the display [6]. The left and right lenses of these glasses switch to an ON state and a DARK state alternately to allow only the correct signals to be passed through to each eye [7]. In this paper, we focus on a system which employs active shutter glasses, based on liquid crystal technology [7]–[11].

Since two 2D images together form a stereoscopic 3D pair, the amount of data for an uncompressed stereo image is doubled compared to that for an uncompressed 2D image. Thus efficient compression techniques are of paramount importance. This paper considers visually lossless compression of stereoscopic 3D images.

A simple approach to this problem might be to compress each of the component images of a stereoscopic pair in a visually lossless manner. To this end, our initial experiments used the method of [12] to compress the left and right images of a stereo pair independently. It was verified that even experienced viewers were unable to perceive any coding artifacts when the original and compressed left images were viewed side-by-side as 2D images. The same held for the right image. Interestingly however, when viewed as a stereo pair, visual artifacts were readily apparent. Indeed, three separate observers examined the images. All three could easily identify significant differences between the original and compressed stereo pairs when viewed side-by-side in 3D mode.

To investigate the source of these artifacts, a compressed left image (as prepared above) was paired with an original uncompressed right image. This pair was then displayed in 3D mode. The display was then observed with the left eye covered, yet artifacts were still perceived (by the right eye viewing the right image), despite the fact that no such artifacts were present in the right image. This experiment, coupled with the fact that only still images are considered, firmly establishes that compression noise leakage is present from one channel to the other due to crosstalk in the stereoscopic display system (and not due to any persistence properties of the human visual system). This phenomenon is discussed further in Section V.

From the preceding discussion, it can be concluded that crosstalk must be carefully considered in the design of visually lossless 3D compression systems. Crosstalk occurs when one

eye can see a signal that is intended to be seen only by the other eye. Since these unintended signals are usually dim, the crosstalk effect is also sometimes called the ghost effect. It can lead not only to degradation in the perceived quality of 3D images, but also to discomfort in some individuals. Crosstalk is caused by leakage between the left and right channels. For active display technologies, crosstalk is contributed to by a glasses-dependent factor and a display-dependent factor [7]. The first factor is related to synchronization between active shutter glasses and 3D displays, response time between the ON state and the DARK state of active shutter glasses, and the various transmittance ratios of the center parts and marginal parts of active shutter glasses. The second factor corresponds to the response time of the display not allowing the correct luminance values of pixels to be reached during the short transition time of switching between the left and right images.

There are ongoing efforts to quantify the amount of crosstalk between the two channels in systems using active shutter glasses [8], [9], [11], [13]. In these four works, constant gray level images are displayed in the left and right channels. The gray levels of these two images are set to various combinations. A photometer is placed behind a lens of the active shutter glasses (where an eye would normally be) to measure luminance values. The crosstalk is then characterized by comparing various measured luminance values. In [14], crosstalk is quantified via subjective experiments for a CRT display system based on liquid crystal shutter glasses. Two side-by-side spatial regions R_1 and R_2 in the left image are set to constant gray levels g_1 and g_2, respectively. Two spatial regions R_3 and R_4, at the same locations as R_1 and R_2, but in the right image are set to constant gray levels $g_3 = 0$ and g_4, respectively. Since $g_3 = 0$, it is hypothesized that R_3 does not induce crosstalk in R_1. However, R_4 does induce crosstalk in R_2. Subjects are asked to adjust g_1 until the pixel intensity of R_1 and R_2 appear identical to the left eye. The crosstalk in R_2 due to R_4 is then computed as the gray level difference g_1–g_2. These works and others (see [15], [16]) have consistently demonstrated that crosstalk from one channel to another is a function of the gray levels in both channels. Additionally, it was found in [10] that crosstalk is also a function of spatial frequency, which is consistent with earlier results published in [17].

As mentioned above, we investigate here the visually lossless compression of 3D stereoscopic images. To this end, we consider the contrast sensitivity function (CSF) for stereoscopic 3D images in the presence of crosstalk. The CSF describes the sensitivity of the human visual system (HVS) to different spatial frequencies (in cycles/degree) and has been a popular avenue of investigation with respect to perceptually based compression of 2D images. In early work, sinusoidal gratings were used in perceptual experiments to measure the CSF. Peak contrast sensitivity has been found to lie between 2 and 6 cycles/degree [18], [19]. More recently, the CSF has been modeled using the discrete wavelet transform [20]. In that work, uniform noise was added to each wavelet subband (one at a time) of an 8-bit constant 128 grayscale image to generate a stimulus image. Visibility thresholds (VTs) were measured for each subband by subjective experiments in which the level of the noise was adjusted to the point where it just became imperceptible to a human viewer.

The method of [20] was extended to a more realistic noise model in [12] and [21]. Specifically, a quantization noise model was developed for the dead-zone quantization of JPEG2000 as applied to wavelet transform coefficients. Then, rather than adding uniform noise to a wavelet subband as in [20], stimulus images were produced by adding noise generated via the dead-zone quantization noise model. The resulting visibility thresholds are functions not only of quantization step size Δ and spatial frequency/orientation as in [20], but also of the variance σ^2 of the wavelet data in each subband. A 2D image compression system was then developed to ensure that the magnitudes of all quantization errors in a wavelet subband fall below the corresponding VT.

Building on previous efforts [17], [22], [23], the work presented in this paper provides a methodology for measuring VTs in the presence of crosstalk via stimulus images. Since the crosstalk observed in one channel is a function of the gray levels in both channels, it is reasonable to suspect that the resulting VTs depend not only on the parameters studied previously for 2D images in [12] and [20], but also on the combination of gray levels displayed in the left and right channels. Indeed, experimental results presented in Section IV confirm this suspicion.

Based on all relevant parameters, a model for VTs in stereoscopic 3D images is proposed. Appropriate VTs derived from this model are then used to design a JPEG2000 coding scheme which compresses the left and right images of a stereo pair jointly. The performance of the proposed JPEG2000 coding scheme is demonstrated by compressing monochrome stereo pairs. The resulting left and right compressed image files can be decoded separately by a standard JPEG2000 decoder. The decompressed images are visually lossless when viewed separately as 2D images, and also when viewed as a 3D stereo pair. This claim is validated via subjective experiments.

The remainder of this paper is organized as follows. In Section II, the modeling of visibility thresholds is described. Methods for measurement of visibility thresholds are presented in Section III. The proposed visually lossless coding scheme for stereo pairs is given in Section IV. In Section V, we present the resulting parameters for the proposed VT model, results from our proposed coding scheme, as well as validation that the resulting compressed stereo pairs are visually lossless. Section VI concludes the work.

II. Modeling of Visibility Thresholds

To facilitate visually losslessly compression via JPEG2000, VT models are developed in this section, for both the left and right images of stereo pairs. Since the procedures for measuring VTs for one image are identical to the other, the focus of the following discussion is the measurement of VTs for the left image.

The VT for a JPEG2000 codeblock of a 2D image is defined [12] as the largest quantization step size Δ for which quantization distortion remains invisible. It is modeled as a function of the wavelet coefficient variance σ^2 within the

codeblock, as well as the orientation $\theta \in \{\text{LL, LH, HL, HH}\}$, and level k of the (dyadic wavelet decomposition) subband to which the codeblock belongs.

As mentioned in the introduction, extension to the case of stereoscopic 3D images requires careful consideration of the crosstalk effect. The crosstalk effect is considered here to be a luminance leakage from one channel to the other channel [7]. When crosstalk occurs, signals intended for the left eye can be seen by right eye. Therefore, VTs for the left image must be chosen so that any quantization distortion (in the left image) is invisible to both the left and right eyes. To this end, we find two VTs t' and t''. We emphasize that both thresholds correspond to the visibility of distortion due to quantization of the left image. The first VT, t', is chosen so that the resulting distortion is invisible to the left eye, while the second VT, t'', ensures that the distortion is invisible to the right eye. The final VT t is then taken as the minimum of t' and t''.

Consistent with the properties of crosstalk observed in [7], [11], and [13], our experiments indicate that the perceived visibility of distortion by the left and right eyes is a function of luminance levels in both the left and right images. Thus, t' and t'' are functions not only of σ^2, θ and k in the left image, but also of the gray level I in both the left and right images. Mathematically, the VT for a given codeblock in the left image at level k and orientation θ is then

$$t_{\theta,k,l}\left(\sigma^2_{\theta,k,l}, I_{\theta,k,l}, I_{\theta,k,r}\right) = \min\left\{t'_{\theta,k,l}\left(\sigma^2_{\theta,k,l}, I_{\theta,k,l}, I_{\theta,k,r}\right), t''_{\theta,k,l}\left(\sigma^2_{\theta,k,l}, I_{\theta,k,l}, I_{\theta,k,r}\right)\right\}, \tag{1}$$

where $\sigma^2_{\theta,k,l}$ is the variance of wavelet coefficients within the codeblock and $I_{\theta,k,l}$ is a gray level for the left image representative of the spatial region (in the image domain) associated with the codeblock. Similarly, $I_{\theta,k,r}$ is a gray level representative of the same spatial region, but in the right image. In our experiments, we employ $K = 5$ levels of dyadic 9/7 wavelet transform so that $1 \leq k \leq 5$. The four orientations θ are indexed as {LL, HL, LH, HH}. For example, the VT associated with the visibility by the *right* eye of distortion introduced in a codeblock of the HL subband of level 2 of the *left* image is denoted by $t''_{HL,2,l}(\sigma^2_{HL,2,l}, I_{HL,2,l}, I_{HL,2,r})$. The computation of $I_{\theta,k,l}$ and $I_{\theta,k,r}$ is discussed in Section IV.

The measurement of VTs via psycho-visual experiments is discussed in Section III. However, exhaustively measuring VTs for all possible parameter choices would be a daunting task. For a given σ^2, there are 2 thresholds $\times$ 16 subbands $\times$ 256^2 combinations of I, for a total of more than 2×10^6 thresholds to be measured. When σ^2 is varied, the total number of thresholds to be measured increases proportionally. Measuring this number of thresholds through subjective experiments is prohibitive. Thus, to reduce the number of experiments to a manageable level, we consider a separable model for VTs given by

$$\hat{t}_{\theta,k,l}\left(\sigma^2_{\theta,k,l}, I_{\theta,k,l}, I_{\theta,k,r}\right) = S_{\theta,k,l}\left(\sigma^2_{\theta,k,l}\right) \times T_{\theta,3,l}\left(I_{\theta,3,l}, I_{\theta,3,r}\right). \tag{2}$$

In this model, we begin with "nominal thresholds" $T_{\theta,3,l}(I_{\theta,3,l}, I_{\theta,3,r})$ measured for level $k = 3$ and variance $\sigma^2 = 50$ via

$$T_{\theta,3,l}\left(I_{\theta,3,l}, I_{\theta,3,r}\right) = \min\left\{t'_{\theta,3,l}\left(50, I_{\theta,3,l}, I_{\theta,3,r}\right), t''_{\theta,3,l}\left(50, I_{\theta,3,l}, I_{\theta,3,r}\right)\right\}. \tag{3}$$

For the 5 level wavelet decomposition employed here, $k = 3$ represents the median transform level. As will be evident from the results presented in Figure 5 of Section V, the value of $\sigma^2 = 50$ is chosen as being near the middle of the range in which the VTs vary most as a function of variance.

The nominal thresholds $T_{\theta,3,l}(I_{\theta,3,l}, I_{\theta,3,r})$ attempt to model the effect of crosstalk caused by different intensities in the left and right images for different orientations, but fixed variance and transform level. The nominal thresholds are then scaled by a factor $S_{\theta,k,l}(\sigma^2_{\theta,k,l})$ which attempts to model the effects of orientation, variance, and transform level (as considered in [12]), but with no crosstalk present. To achieve a state of no crosstalk, the intensities in the left and right images are set to be equal [11]. To limit the number of required measurements, as described below, only the median gray level is employed in $S_{\theta,k,l}(\sigma^2_{\theta,k,l})$. Specifically,

$$S_{\theta,k,l}\left(\sigma^2_{\theta,k,l}\right) = \frac{t_{\theta,k,l}\left(\sigma^2_{\theta,k,l}, 128, 128\right)}{T_{\theta,3,l}(128, 128)}. \tag{4}$$

A methodology for the measurement of $T_{\theta,3,l}$ and $S_{\theta,k,l}$ is provided in the next section.

III. Measurement of Visibility Thresholds

The 3D vision system used in this work includes the Nvidia Quadro FX 3800 graphics card and active shutter glasses. A USB IR transmitter is used to synchronize the glasses with a Samsung SyncMaster 2233 RZ display (22" with 1680 $\times$ 1050 resolution, 120 Hz refresh-rate, 300 cd/m^2 brightness, 1,000:1 typical contrast ratio, and 170°/160° viewing angle). The display resolution d for the SyncMaster display is 35.45 pixels/cm. When the viewing distance υ between subject and display is 60 cm, the resulting visual resolution r is 37.12 pixels/degree, which is derived via in [20, eq. 1].

Since the above-mentioned 3D stereoscopic display system is frequently used in the office/home, we chose a goal of ensuring visually lossless quality in the work environment [24]. To this end, all visual experiments were conducted with normal office lighting conditions and a viewing distance of 60 cm. The results presented herein are only guaranteed to hold for the specific stereo display system and particular viewing conditions considered. On the other hand, the methodology presented can be used to obtain VTs for different display devices and/or more critical viewing conditions such as those described in [25] and [26]. It is worth noting that a straightforward application of the proposed methodology would result in a fixed design for each display device and viewing condition. An interesting avenue for future work is the development of a more general model for VTs that could take as input such parameters as display resolution, veiling luminance, viewing distance, etc. [27].

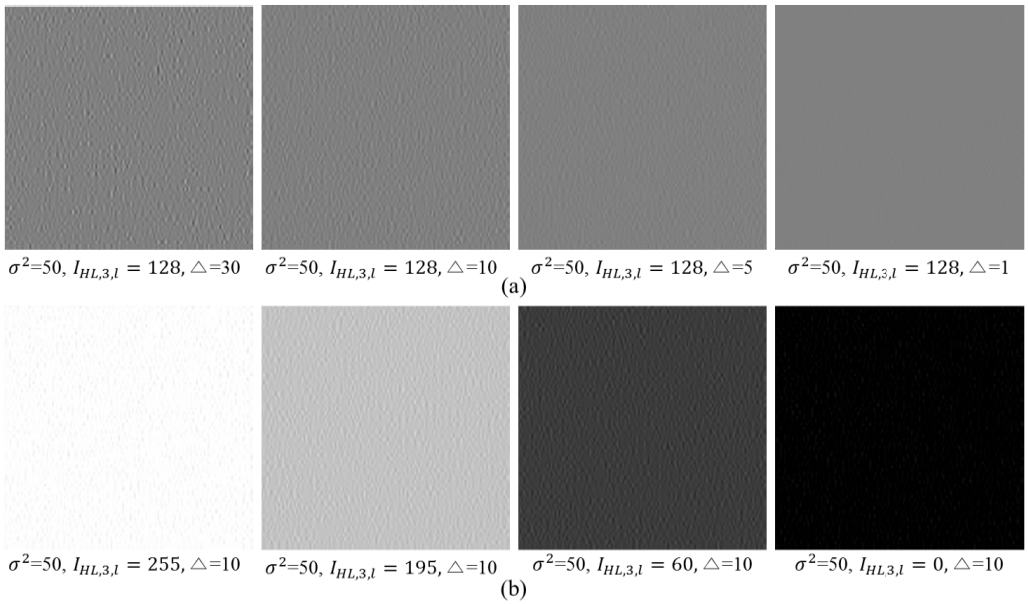


Fig. 1. Stimulus images derived by adding simulated quantization distortion in the HL3 subband with $\sigma^2 = 50$. (a) $I_{HL,3,l}$ is fixed to 128 but the value of Δ varies among 30, 10, 5, and 1. (b) The value of Δ is fixed to 10 but the value of $I_{HL,3,l}$ is varied among 255, 195, 60, and 0.

For the purpose of measuring VTs, stimulus images are generated by performing the inverse wavelet transform of wavelet coefficient data containing simulated quantization distortion. Specially, the wavelet data for all subbands of an image of size 512×512 are initialized to 0, and then noise is added to one subband. This noise is generated pseudo-randomly according to the JPEG2000 quantization distortion model of [12]. The inverse wavelet transform is performed, and the result is added to a *constant* gray level image having all pixel intensities set to a fixed value $I_{\theta,3,l}$ between 0 and 255. After the addition, the value of each pixel in this stimulus image is rounded to the closest integer between 0 and 255.

In order to measure the visibility thresholds $t'_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$ and $t''_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$, as required by (3), *left* stimulus images are created by adding noise (as described above) to a level 3 subband with orientation θ. Figure 1 illustrates stimulus images generated by adding noise to the HL3 subband ($k = 3$ and $\theta = \text{HL}$). In particular, Figure 1 (a) demonstrates stimulus images for various values of Δ with $\sigma^2 = 50$ and $I_{HL,3,l} = 128$. As described above, the value of $I_{HL,3,l}$ corresponds to the background gray level of the stimulus image. The values of Δ and σ^2 correspond to an assumed quantization step size and variance of wavelet coefficients to be quantized, respectively. From this figure, we see that when σ^2 and $I_{HL,3,l}$ are fixed, noise visibility decreases with Δ as expected.

In Figure 1 (b), the assumed coefficient variance and quantization step size are held fixed at $\sigma^2 = 50$ and $\Delta = 10$, while the background intensity $I_{HL,3,l}$ is varied among 255, 195, 60, and 0. This figure demonstrates, as is well known in the literature [28], [29], that noise visibility is a function of the background gray level $I_{HL,3,l}$. In fact, as will be seen in subsequent sections, experiments indicate that when viewed in 3D mode, the visibility of noise (introduced in only the left image) is a function of both $I_{\theta,k,l}$ and $I_{\theta,k,r}$.

Accordingly, to find $t'_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$ and $t''_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$, a value of Δ is chosen and a left stimulus image (containing noise) is created with parameters, Δ, θ, $I_{\theta,3,l}$, $k = 3$, and $\sigma^2 = 50$. This image is paired with a right constant gray image with all pixels set to $I_{\theta,3,r}$ (and no noise). The value of Δ in the left stimulus image is then adjusted until the noise is just invisible to the left eye, with the right eye covered. The resulting value of Δ is $t'_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$. The experiment is then repeated with the same setup, but with the left eye covered. The value of Δ in the *left* stimulus image is then adjusted so that the noise is just invisible to the *right* eye. The resulting value of Δ is $t''_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$. We emphasize again that noise is introduced only in the left image for the measurement of both t' and t''.

The thresholds $t'_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$ and $t''_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$ are measured for five values of $I_{\theta,3,l}$ and nineteen values of $I_{\theta,3,r}$. Specifically, $I_{\theta,3,l} \in \{0, 60, 128, 195, 255\}$ while $I_{\theta,3,r}$ can take any of the eighteen integer multiples of 15 between 0 and 255, plus the additional value of 128. That is $I_{\theta,3,r} \in \{0, 15, \ldots, 120, 128, 135, \ldots, 240, 255\}$.

We now proceed to the measurement of $S_{\theta,k,l}(\sigma^2_{\theta,k,l})$ as given in (4). The denominator of (4) is a special case of $T_{\theta,3,l}(I_{\theta,3,l}, I_{\theta,3,r})$ and is included in the measurements described above. The numerator $t_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$ remains to be discussed. To this end, the gray levels for both the left and right images are fixed to 128 but the value of the assumed (left) codeblock variance $\sigma^2_{\theta,k,l}$ is varied. For a given choice of θ, k, and $\sigma^2_{\theta,k,l}$, the value of Δ in the left stimulus image is again adjusted to the point that the noise is just

imperceptible for each eye, resulting in $t'_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$ and $t''_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$. Six values of $\sigma^2_{\theta,k,l} \in \{10, 50, 100, 200, 600, 1000\}$ are tested for each value of θ and k. The total number of thresholds measured for the numerator of (4) is then 6 variances $\times$ 2 thresholds $=$ 12 per subband. Six values of $t_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$ are then obtained for each subband as the minimum between $t'_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$ and $t''_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$ for each θ and each $k \in \{1, 2, \ldots, 5\}$.

During the experiments described above, an opaque mask is used to cover the right eye of the subject during the measurement of t'. Similarly, the left eye is covered during the measurement of t''. Due to the fact that the central part of the active shutter lenses is darker than the peripheral part, signals correctly blocked by the central part may be perceived as crosstalk through the peripheral part. Thus, in order to obtain conservative values for the VTs, the subject is allowed to tilt their head during measurements.

Spatial three-alternative forced-choice (3AFC) testing is employed to find the value of Δ for which the distortion is just invisible in each case. Three stereoscopic images are shown on the display concurrently. One is placed at the top center of the screen, and the other two are arranged at the bottom left and bottom right, respectively. A stereoscopic stimulus image (containing noise in only the left channel) is displayed randomly at one of three locations. The other two stereoscopic images contain no noise. The subject is asked to decide which stereoscopic image contains the noise by means of keyboard input.

For a given combination of θ, k, $\sigma^2_{\theta,k,l}$, $I_{\theta,k,l}$ and $I_{\theta,k,r}$, 32 trials (each with a different value of Δ) are conducted. The value of Δ in each trial is controlled by the QUEST staircase procedure obtained from the Psychophysics Toolbox [30]. In each trial, the three stereoscopic images are displayed for 20 seconds. The subject can take an unlimited amount of time to make a decision as to which stereoscopic image contains the stimulus after the stereoscopic images are removed from the display. The VT is then taken as the value of Δ derived from the 75% correct-point on the Weibull function fitted from the 32 trails [31]. It is worth emphasizing that all VTs are measured using only (pseudo-randomly generated) stimulus images. No actual images are used in this regard.

In what follows, 4$^{\text{th}}$ order polynomial interpolation is used to obtain values of the nominal thresholds $T_{\theta,3,l}(I_{\theta,3,l}, I_{\theta,3,r})$ for values of $I_{\theta,3,l}$ and $I_{\theta,3,r}$ not employed in the visual measurements described above. Negative values may result from this interpolation. Since VTs should be positive, the nominal threshold is taken as the maximum between its interpolated value and a small constant $\delta = 0.25$, which corresponds to approximately 70% of the minimum of all nominal thresholds measured via psycho-physical tests. Similarly, two term power series interpolation is employed to obtain values of the numerator in (4) for values of $\sigma^2_{\theta,k,l}$ not employed in the measurements. The resulting values, when employed in (2), may cause $\hat{t}_{\theta,k,l}(\sigma^2_{\theta,k,l}, I_{\theta,k,l}, I_{\theta,k,r})$ to be negative or overly small. Thus, the larger of $\hat{t}_{\theta,k,l}(\sigma^2_{\theta,k,l}, I_{\theta,k,l}, I_{\theta,k,r})$ and a small constant $\epsilon = 0.025$ is taken as the final VT. This value of ϵ corresponds to approximately 10% of the minimum $\hat{t}_{\theta,k,l}(\sigma^2_{\theta,k,l}, I_{\theta,k,l}, I_{\theta,k,r})$ calculated using only measured thresholds (i.e., before interpolation). The value of 10% is chosen as an overly conservative value, due to the fact that it is applied to the final VTs.

IV. Visually Lossless Coding of Stereo Pairs

The proposed coding method for visually lossless compression of 8-bit monochrome stereoscopic images is adapted from the coding method of [12]. In JPEG2000, a subband is partitioned into rectangular codeblocks. The coefficients of each codeblock are then quantized and encoded via bit-plane coding. Three coding passes (significance propagation, magnitude refinement, and cleanup) are performed for each bit-plane except the most significant bit-plane which only has a cleanup pass. The maximum possible number of coding passes for a codeblock is then $3M - 2$ where M denotes a number of bit-planes sufficient to represent the magnitude of all quantized coefficients in a codeblock [32]. The actual number of coding passes included in a compressed code stream can vary from codeblock to codeblock and is typically selected to optimize mean squared error over the entire image for a given target bit rate [32].

Rather than minimizing mean squared error, the method proposed in [12] includes the minimum number of coding passes necessary to achieve visually lossless encoding of a 2D image. This is achieved by including a sufficient number of coding passes for a given codeblock such that the absolute error of every coefficient in that codeblock is less than the VT for that codeblock. This is extended to the coding of 3D images in what follows. Specifically, the coding of the left image of a stereo pair is carried out using the left VT for each codeblock of the left image as computed via the process described in Section III. The right image is then encoded using the right VT for each codeblock of the right image. Asymmetries [33] are not considered in this work. Thus, we compute right VTs in the identical manner as left VTs, by simply reversing the roles of the left and right images.

For simplicity, we do not consider visual masking, nor other more sophisticated perceptual coding tools such as those described in [12], [21], and [34]. Thus, the resulting compression ratios reported in subsequent results should be considered as lower bounds to what might be possible. Indeed, in [12] the VT for each codeblock is modified by a multiplicative factor to account for visual masking. This masking factor is computed based on the image data in the supporting region of a codeblock, and is always greater than or equal to 1.0. The resulting increased VTs yield an increase in the compression ratio of approximately 10% over the unmodified VTs, while still maintaining visually lossless quality.

From (3) and (4), the VT of a codeblock in the left image depends not only on the variance of the wavelet coefficients within the codeblock, but also on gray levels from the left and right images representative of the spatial region corresponding to the codeblock. In our initial experiments, the requisite gray levels were computed as the average pixel intensity (in the left and right images, respectively) over the supporting region

```
For each sub-codeblock
{
        α = 0.0001, ϵ = 0.025, and δ = 0.25
        Compute I_{θ,k,l} and I_{θ,k,r} for the supporting region of the sub-codeblock
        Compute σ²_{θ,k,l} for the sub-codeblock
        if σ²_{θ,k,l} < α
                σ²_{θ,k,l} = α
        Compute t̂_{θ,k,l} for the sub-codeblock:
                Determine T_{θ,3,l}(I_{θ,3,l}, I_{θ,3,r}) and S_{θ,k,l}(σ²_{θ,k,l}), interpolating as necessary
                if T_{θ,3,l}(I_{θ,3,l}, I_{θ,3,r}) < δ
                        T_{θ,3,l}(I_{θ,3,l}, I_{θ,3,r}) = δ
                t̂_{θ,k,l}(σ²_{θ,k,l}, I_{θ,k,l}, I_{θ,k,r}) = S_{θ,k,l}(σ²_{θ,k,l}) · T_{θ,3,l}(I_{θ,3,l}, I_{θ,3,r})
                if t̂_{θ,k,l}(σ²_{θ,k,l}, I_{θ,k,l}, I_{θ,k,r}) < ϵ
                        t̂_{θ,k,l}(σ²_{θ,k,l}, I_{θ,k,l}, I_{θ,k,r}) = ϵ
}
t̂_{θ,k,l} = min{t̂_{θ,k,l}(σ²_{θ,k,l}, I_{θ,k,l}, I_{θ,k,r})} over all sub-codeblocks
for coding pass Z = 0 to 3M − 1
{
        if D^(Z) < t̂_{θ,k,l}
                terminate coding
}
```

Fig. 2. Pseudo-code of the proposed coding method for a codeblock from the left image.

of the codeblock. Due to inadequate spatial granularity, this approach results in clearly visible distortion. To illustrate the lack of spatial granularity, consider a codeblock of size 32×32 in level $k = 5$. Such a codeblock has a spatial support in the image domain of size 1024×1024. The average gray level over such a large area provides little information about the potential for crosstalk in disparate regions within this area.

In order to find a conservative VT for a codeblock of the left image, the codeblock is partitioned into sub-codeblocks. As in the example above, codeblocks are taken to be of size 32×32. The dimensions of the sub-codeblocks are taken as $W \times W$, where $W = 32/2^k$. The supporting region of each sub-codeblock in the image domain is then 32×32. For each sub-codeblock, the average pixel intensity is then computed over its supporting region for both the left and right images. These values are taken as $I_{\theta,k,l}$ and $I_{\theta,k,r}$ for the sub-codeblock. Similarly, the variance of the wavelet data is computed for the sub-codeblock and taken as $\sigma^2_{\theta,k,l}$ for the sub-codeblock. In the level $k = 5$ subbands, all sub-codeblocks are of size 1×1, and thus have a variance of zero. Although rare, a variance of zero can also occur in subbands with $k < 5$. To avoid numerical problems, any zero valued variance is replaced by the arbitrarily chosen value $\alpha = 10^{-4}$. The values of $I_{\theta,k,l}$, $I_{\theta,k,r}$, and $\sigma^2_{\theta,k,l}$ are then used to compute $\hat{t}_{\theta,k,l}$ for each sub-codeblock. Finally, the VT for a codeblock is chosen as the minimum $\hat{t}_{\theta,k,l}$ over all sub-codeblocks in the codeblock.

As described previously, the maximum absolute error over all coefficients in a codeblock should be smaller than the VT of the codeblock in order to obtain a visually lossless encoding. In other words, let $D^{(Z)}$ be the maximum absolute error that would be incurred if only coding passes 0 through Z were decoded. Starting with coding pass 0, encoding proceeds until the first coding pass Z such that $D^{(Z)}$ is smaller than the VT for the codeblock. Pseudo-code for the compression of a codeblock from the left image is provided in Figure 2.

V. Results

A. Visibility Thresholds

Figure 3 shows the measured values of $t'_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$ and $t''_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$ for the LL3, HL3, and HH3 subbands of the left image. VTs for LH subbands are similar to those for HL subbands. Thus, thresholds measured for the HL subbands are used for the LH subbands in what follows. In Figure 3, left eye VTs $t'_{\theta,3,l}$ are plotted in blue, while the red curves denote right eye VTs $t''_{\theta,3,l}$ (both due to distortion introduced only in the left image). The VTs for LL3, HL3 and HH3 are represented by circle, triangle and asterisk symbols, respectively. Each subfigure provides graphs of VTs as a function of $I_{\theta,3,r}$ for a fixed value of $I_{\theta,3,l}$.

Consider for a moment only the blue curves, which correspond to the visibility of noise (in the left image) by the left eye, but plotted as a function of the background gray level in the right image $I_{\theta,3,r}$. As might be expected, these curves are relatively flat, indicating low sensitivity to $I_{\theta,3,r}$. Keeping in mind that larger thresholds correspond to lower noise visibility, a comparison between different subplots also yields expected results. That is, the sensitivity of the left eye to noise in the left image is reduced for both low and high background intensities, as discussed previously in connection with Figure 1 (b). To see this, it is important to note the different scales of the vertical axes in each of the subplots.

Consider now the red curves of Figure 3. These depict the sensitivity of the right eye to noise introduced only in the left image. In the absence of crosstalk, there would be no such sensitivity and all corresponding VTs would be infinite. The shape of the red curves indicate that the sensitivity to noise due to crosstalk decreases for both low and high background intensity levels in the *right* image. Careful examination of the vertical axis scales in the different subfigures shows that the visibility of noise in the right eye due to noise in the left image also decreases for both low and high background intensity levels in the *left* image. This indicates that noise visibility due to crosstalk is a function of the intensity levels in both images, as claimed earlier.

As expected, the values of $t'_{\theta,3,l}$ (blue) typically lie below those of $t''_{\theta,3,l}$ (red) indicating that distortion introduced in the left image is *generally* more visible to the left eye. It is important to note however that significant exceptions exist. For some combinations of luminance values, the red curves fall below the corresponding blue curves, indicating that noise in the left image is more easily seen by the right eye. The existence of this effect can be understood as an extension of the fact that noise visibility is a function of background gray level, as discussed previously for 2D images in conjunction with Figure 1. In the measurement of thresholds as discussed above, flat (constant gray level) left and right images are created. Noise of a given variance is then added only in the left image. Consistent with Figure 1, certain choices of background gray level in the left image may render the noise invisible by the left eye, while others may not. Consider selecting for the background gray level of the left image one of the values that conceals the noise from the left eye. Now, due to crosstalk, a certain (reduced) amount of noise from the left image will

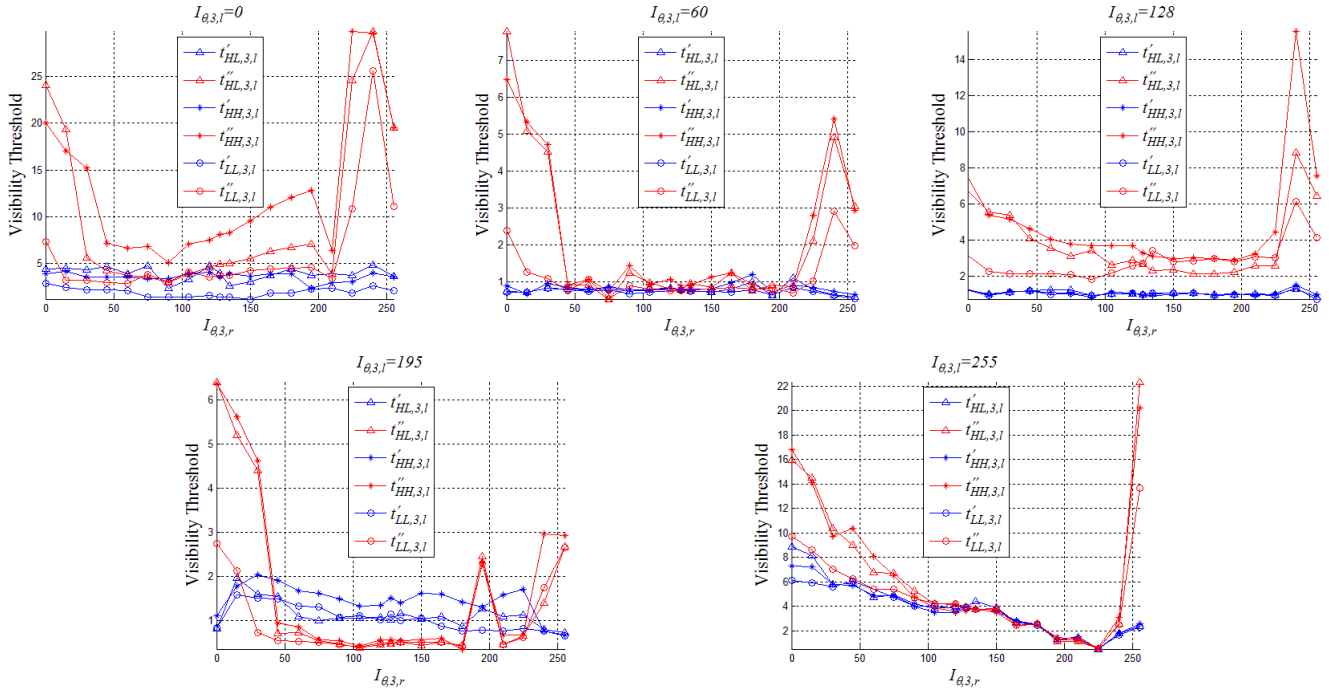


Fig. 3. Visibility thresholds measured via subjective experiments. $t'_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$ and $t''_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$ are represented by blue and red curves, respectively. Circle symbols, triangle symbols, and asterisk symbols are used for LL3, HL3, and HH3, respectively. From (3), it can be seen that the nominal thresholds are equal to the minimum between the red and blue curves for each set of parameter choices.

leak through to the right eye. It is reasonable to expect that certain choices of background gray level in the right image may result in no distortion perceived by the right eye, while other choices may not. That is, there may exist choices for the left and right gray levels which result in the noise being imperceptible by the left eye, yet perceptible by the right eye. This effect is particularly evident in the subfigure with $I_{\theta,3,l} = 195$. This justifies the need to consider both $t'_{\theta,3,l}$ and $t''_{\theta,3,l}$ and to choose $t_{\theta,3,l} = \min\{t'_{\theta,3,l}, t''_{\theta,3,l}\}$ to ensure that noise introduced into the left image is imperceptible to both the left and right eyes.

As mentioned previously, values of $T_{\theta,3,l}(I_{\theta,3,l}, I_{\theta,3,r})$ for values of $I_{\theta,3,l}$ and $I_{\theta,3,r}$ not employed in the psychovisual experiments are rendered using 4$^{\text{th}}$order polynomial interpolation. Specifically,

$$\begin{aligned} T_{\theta,3,l}&(I_{\theta,3,l}, I_{\theta,3,r}) \\ &= p_{00} + I_{\theta,3,l}\,p_{10} + I_{\theta,3,r}\,p_{01} + I^2_{\theta,3,l}\,p_{20} \\ &\quad + I_{\theta,3,l}I_{\theta,3,r}\,p_{11} + I^2_{\theta,3,r}\,p_{02} + I^3_{\theta,3,l}\,p_{30} \\ &\quad + I^2_{\theta,3,l}I_{\theta,3,r}\,p_{21} + I_{\theta,3,l}I^2_{\theta,3,r}\,p_{12} + I^3_{\theta,3,r}\,p_{03} \\ &\quad + I^4_{\theta,3,l}\,p_{40} + I^3_{\theta,3,l}I_{\theta,3,r}\,p_{31} + I^2_{\theta,3,l}I^2_{\theta,3,r}\,p_{22} \\ &\quad + I_{\theta,3,l}I^3_{\theta,3,r}\,p_{13} + I^4_{\theta,3,r}\,p_{04}. \end{aligned} \tag{5}$$

The interpolation parameters p_{ij} are given in Table I, while Figure 4 (a), (b), and (c) depict the interpolated surfaces for $T_{\theta,3,l}(I_{\theta,3,l}, I_{\theta,3,r})$ with θ equal to LL, HL, and HH, respectively. As can be seen in the figures, the VT surfaces are somewhat "bowl shaped". Consistent with previous discussions, this indicates that noise visibility is decreased when the gray levels of either image are low or high, and

TABLE I

PARAMETERS FOR 4$^{\text{th}}$ ORDER POLYNOMIAL INTERPOLATION OF $T_{\theta,3,l}$

	LL3	HL3/LH3	HH3
p_{00}	2.8563	4.3468	3.9231
p_{10}	-0.058	-0.1092	-0.1164
p_{01}	-0.0298	-0.0411	-0.0145
p_{20}	5.126e-4	1.198e-3	1.51e-3
p_{11}	4.925e-4	6.306e-4	4.118e-4
p_{02}	1.817e-4	2.705e-4	-3.869e-5
p_{30}	-2.836e-6	-6.718e-6	-8.673e-6
p_{21}	-8.072e-7	-5.579e-7	-6.727e-7
p_{12}	-2.182e-6	-3.846e-6	-1.853e-6
p_{03}	-4.664e-7	-4.987e-7	4.085e-7
p_{40}	7.497e-9	1.518e-8	1.866e-8
p_{31}	-6.084e-9	-7.397e-9	-5.732e-9
p_{22}	8.656e-9	9.273e-9	7.864e-9
p_{13}	5.094e-10	4.483e-9	2.184e-10
p_{04}	6.647e-10	1.073e-10	-2.161e-10

clearly demonstrates that $T_{\theta,3,l}(I_{\theta,3,l}, I_{\theta,3,r})$ is a function of both $I_{\theta,3,l}$ and $I_{\theta,3,r}$. It is worth noting (see (3)) that the interpolation is performed after taking the point-by-point minimum between the appropriate red and blue curves of Figure 3. Thus, the interpolation does not need to preserve the sudden drops and peaks seen in the individual red and blue curves of Figure 3.

Figure 5 shows $t_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$ as a function of $\sigma^2_{\theta,k,l}$ for the LL, HL, and HH subbands for level $k \in \{1, 2, 3, 4, 5\}$. As discussed previously, these thresholds are used in the numerator of the scaling factor in (4), and are measured for $\sigma^2_{\theta,k,l} \in \{10, 50, 100, 200, 600, 1000\}$. Measured $t_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$ for levels 1, 2, 3, 4, and 5 are represented by magenta asterisk, red triangle, green cross, cyan

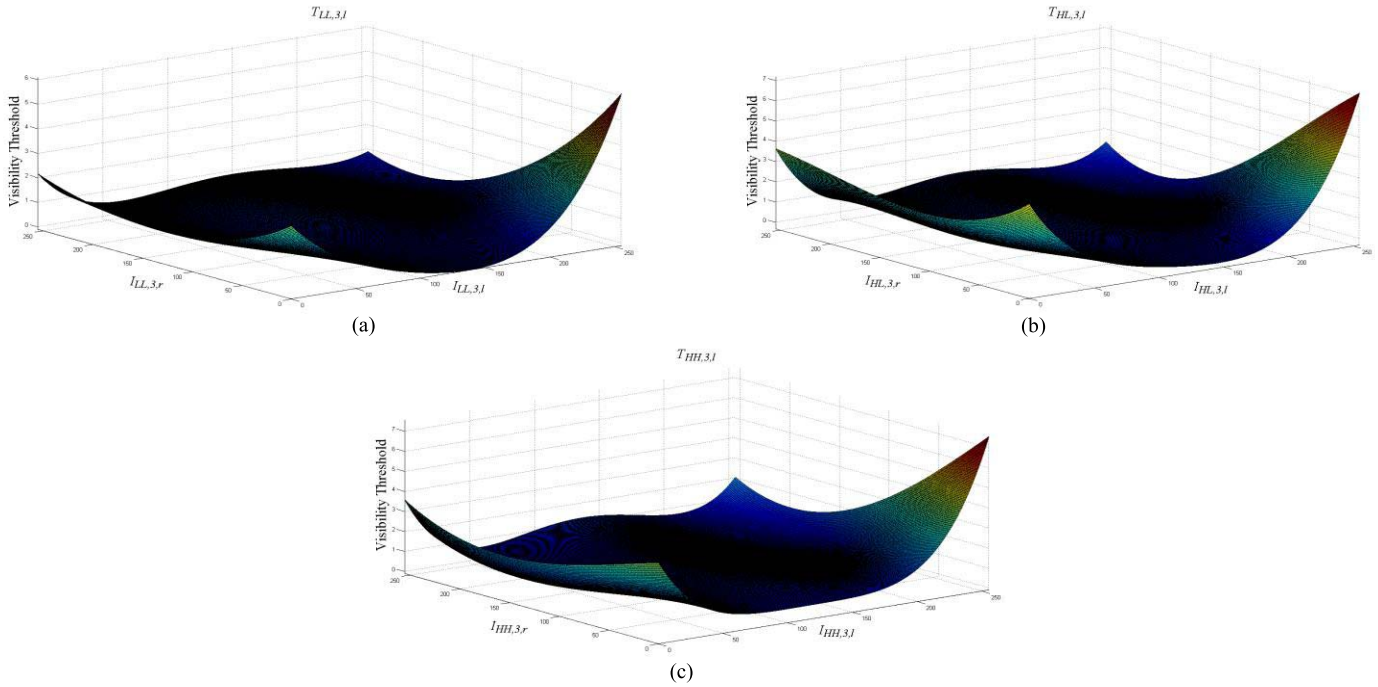


Fig. 4. VTs as a function of $I_{\theta,3,l}$ and $I_{\theta,3,r}$. (a) $T_{LL,3,l}(I_{LL,3,l}, I_{LL,3,r})$, (b) $T_{HL,3,l}(I_{HL,3,l}, I_{HL,3,r})$, and (c) $T_{HH,3,l}(I_{HH,3,l}, I_{HH,3,r})$.

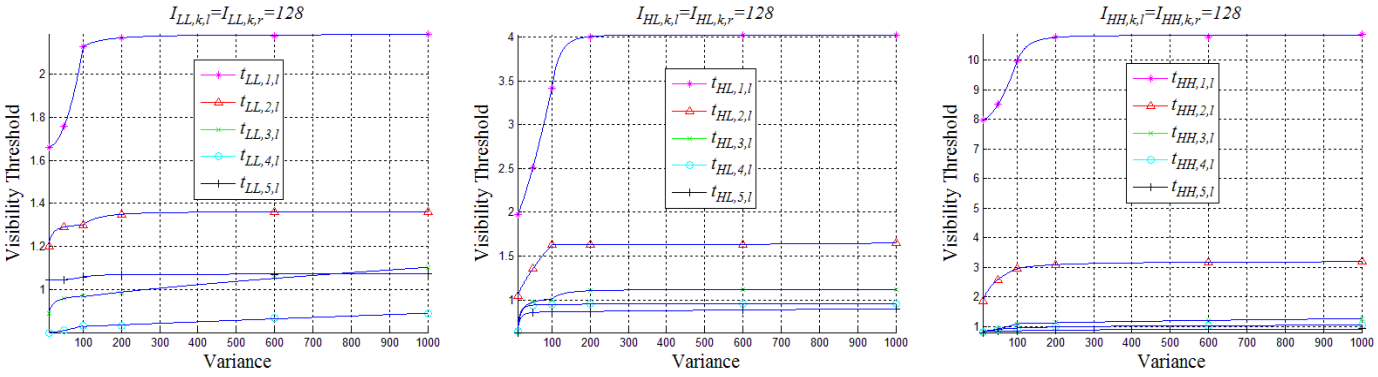


Fig. 5. $t_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$. The measured $t_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$ for $k = 1, 2, 3, 4$, and 5 are represented by magenta asterisk symbols, red triangle symbols, green cross symbols, cyan circle symbols, and black plus sign symbols, respectively. The blue curves in each figure denote interpolated $t_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$.

TABLE II
PARAMETERS FOR POWER SERIES INTERPOLATION ($\sigma^2_{\theta,k,l} \in [0, 100]$)

subband	a	b	c	Subband	a	b	c	Subband	a	b	c
LL1	1.758e-5	2.214	1.656	HL1/LH1	0.003	1.323	1.904	HH1	4.123e-4	1.847	7.934
LL2	-1.298	-1.088	1.309	HL2/LH2	0.034	0.671	0.892	HH2	1.78	0.157	-0.7
LL3	-0.558	-0.759	0.987	HL3/LH3	-5.097	-1.109	1.041	HH3	4.16e-4	1.417	0.813
LL4	1.442e-5	1.674	0.8	HL4/LH4	-38.657	-2.094	0.95	HH4	0.618	0.071	0.095
LL5	3.838e-8	2.789	1.044	HL5/LH5	-19.6	-1.916	0.866	HH5	-4.553	-1.909	0.86

circle, and black plus sign symbols, respectively. The blue curves depicted in Figure 5 are the interpolated versions of $t_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$, rendered using two term power series as

$$t_{\theta,k,l}\left(\sigma^2_{\theta,k,l}, 128, 128\right) = a\left(\sigma^2_{\theta,k,l}\right)^b + c. \quad (6)$$

In order to obtain high-quality fits, the range of the variance $\sigma^2_{\theta,k,l}$ is partitioned into two segments, from 0 to 100 and from 100 to infinity, respectively. The constant parameters a, b, and c for each segment are provided in Tables II and III. Sufficient parameters are provided in the tables to accommodate wavelet transforms employing from 1 to 5 levels. However, exactly 5 levels of transform are employed in all visually lossless compression experiments that follow. As can be seen in the figure, the curves rise steeply for $\sigma^2_{\theta,k,l}$ between 0 and 100, especially for small values of k. When the value of $\sigma^2_{\theta,k,l}$ is large, the curves are relatively

TABLE III
PARAMETERS FOR POWER SERIES INTERPOLATION ($\sigma^2_{\theta,k,l} \in [100, \infty]$)

subband	a	b	c	Subband	a	b	c	Subband	a	b	c
LL1	-346.9	-1.891	2.185	HL1/LH1	-9.122e9	-5.088	4.024	HH1	-1.658e7	-3.644	10.829
LL2	-3.251e3	-2.362	1.362	HL2/LH2	1.97e-10	2.671	1.629	HH2	-29.434	-1.031	3.216
LL3	0.002	0.689	0.932	HL3/LH3	-3.93e5	-3.286	1.115	HH3	0.005	0.552	1.033
LL4	1.181e-4	0.925	0.82	HL4/LH4	-0.095	-0.435	0.96	HH4	-0.754	-0.278	1.161
LL5	-1.202e3	-2.481	1.072	HL5/LH5	2.858e-4	0.731	0.854	HH5	-0.469	-0.109	1.144

constant. In general, the values of $t_{\theta,k,l}(\sigma^2_{\theta,k,l}, 128, 128)$ for lower levels (higher frequency) are larger than those for higher levels.

To conclude this sub-section, we discuss several practical matters regarding the measurement of VTs. First, we note that the measured value of VTs can vary somewhat from observer to observer. The measured VTs may even vary between the two eyes of a single observer. Thus, a very conservative approach might be to employ many observers to obtain a VT for each eye of each observer for each fixed choice of parameter settings. For each such choice, the "final measured VT" could then be taken as the minimum over all measured VTs for that choice. However, we do not follow such a labor intensive approach here. Indeed, Figures 3 and 5 represent measurements from only the right eye of only one observer, yet represent many weeks of observation effort.

Second, also due to labor intensity, we had to carefully consider which parameter choices to employ in the measurement of VTs. The main motivation for this work is the surprising fact that the right eye can sometimes perceive distortion in the left image, even when that distortion is not visible by the left eye. For this reason, we put more effort into the measurement of the right eye thresholds $t''_{\theta,3,l}(50, I_{\theta,3,l}, I_{\theta,3,r})$. From a limited set of experiments, it was determined that the perceived distortion to the right eye is more dependent on $I_{\theta,3,r}$ than $I_{\theta,3,l}$. For this reason, nineteen values of $I_{\theta,3,r}$ were employed, while only five values of $I_{\theta,3,l}$ were utilized.

Although no theoretical justification has been provided to support our use of the proposed separable model, and it is not possible to directly evaluate the "goodness of fit" of the proposed model for parameter values not used in the experiments, it will be demonstrated in Subsection V.C that a compressor based on this model indeed yields visually lossless imagery. On the other hand, that does not imply that a more sophisticated model might not yield visually lossless imagery with smaller compressed file sizes than those presented here. For example, a more sophisticated model might consider both the variance of the wavelet data in left and right images. Similarly, the value of the variance used in the measurement of the nominal thresholds might be selected more carefully, even on a subband by subband basis, rather than using the fixed value of 50 as proposed herein.

B. Coding Results

Thirteen 8-bit monochrome stereo pairs are compressed and results are reported in units of bits/pixel. The stereo pairs used in this experiment are from the Middlebury stereo datasets [35]–[38]. The left image of each stereo pair is shown in Figure 6. Each left image is compressed as described in Section IV. The identical process is carried out for the right image, interchanging the roles of left and right. The encoder is implemented in Kakadu V6.1 [39]. Performance is reported in Table IV for three encoding methods: information lossless JPEG2000, visually lossless JPEG2000 for 2D images [12], and our proposed visually lossless JPEG2000 for stereoscopic 3D images. In every case, decompression can be performed using as unmodified JPEG2000 decoder (kdu_expand).

As validated in the following section, the proposed coding method achieves visually lossless performance for each image when viewed as a stereoscopic 3D pair, (as well as for the left and right images viewed separately in 2D). On the other hand, as mentioned in the Introduction, when the visually lossless coding method for 2D images from [12] is adopted to compress the left and right images independently, compression artifacts are clearly visible when a compressed stereo pair is displayed in 3D mode. This is despite the fact that the left and right images are indeed visually lossless when viewed separately in 2D mode. This result holds for the VTs reported in [12], as well as for VTs designed specifically for the Samsung monitor used in this work. The results reported in Table IV are for the latter VTs. Evidently, larger bitrates are required to achieve visually lossless coding for 3D stereoscopic images than for 2D images for the display and viewing conditions employed. This increased bit rate is due to crosstalk as discussed previously.

Comparing the performance of the method from [12] for 2D images with the performance of our proposed method for stereoscopic 3D images, the average bitrate produced by the proposed method is higher, as expected. On the other hand, the average bitrate of the proposed coding method is lower than that for information lossless coding by roughly a factor of 1.36, for a compression ratio of about 4 to 1 without any loss in visual quality.

C. Validation

In [12] and [40], spatial 3AFC testing was performed to validate that 2D images obtained after decompression were visually lossless. As a part of that procedure, three copies of a 2D image were displayed side-by-side on the screen. Unfortunately, three full resolution stereo pairs do not fit side-by-side on the screen. Visibility of artifacts, as well as 3D perception of stereoscopic images can be adversely impacted if the dimensions of the images are decreased by

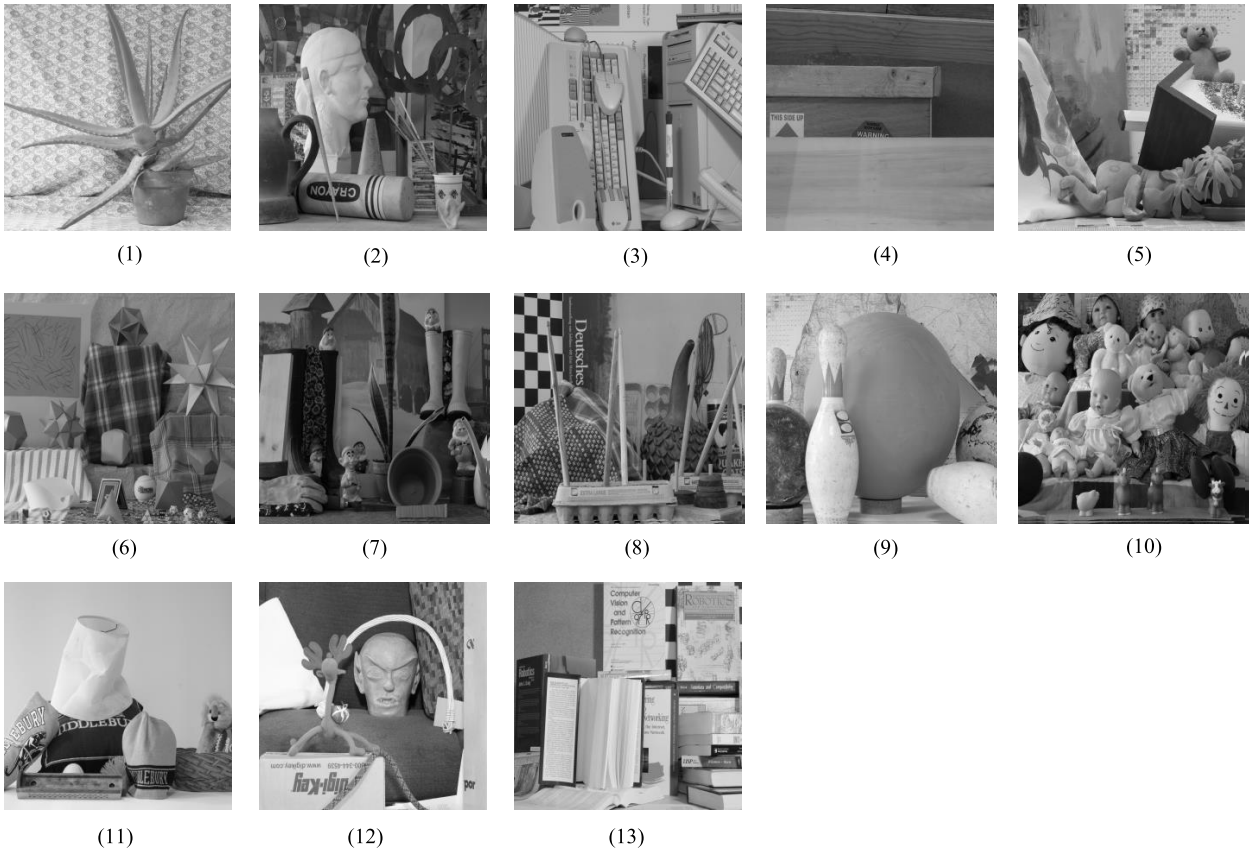


Fig. 6. The thirteen monochrome stereo images used in the experiments (only the left image of each stereo pair is shown). (1) Aloe, (2) Art, (3) Computer, (4) Wood2, (5) Teddy, (6) Moebius, (7) Dwarves, (8) Drumsticks, (9) Bowling2, (10) Dolls, (11) Midd1, (12) Reindeer, and (13) Books.

TABLE IV

BITRATES FOR (A) INFORMATION LOSSLESS JPEG2000, (B) VISUALLY LOSSLESS CODING FOR 2D IMAGES [12], AND (C) THE PROPOSED VISUALLY LOSSLESS METHOD FOR STEREOSCOPIC 3D IMAGES

Image	Dimension (W×H)	(a) Left	(a) Right	(b) Left	(b) Right	(c) Left	(c) Right
1. Aloe	1282×1110	3.826	3.792	1.315	1.304	2.976	2.964
2. Art	1390×1110	2.539	2.494	0.635	0.619	1.763	1.711
3. Computer	1330×1110	1.837	1.866	0.484	0.515	1.398	1.425
4. Wood2	1306×1110	1.79	1.739	0.306	0.294	1.053	0.964
5. Teddy	1800×1500	2.938	2.865	0.762	0.734	2.072	2.02
6. Moebius	1390×1110	2.842	2.889	0.737	0.752	1.858	1.879
7. Dwarves	1390×1110	2.069	2.056	0.457	0.461	1.367	1.321
8. Drumsticks	1390×1110	2.746	2.594	0.818	0.734	1.821	1.663
9. Bowling2	1330×1110	2.324	2.175	0.625	0.575	1.842	1.695
10. Dolls	1390×1110	2.705	2.656	0.76	0.742	1.939	1.862
11. Midd1	1396×1110	2.022	1.942	0.545	0.518	1.713	1.621
12. Reindeer	1328×1050	3.545	3.292	1.084	0.987	2.544	2.387
13. Books	1390×1110	2.649	2.58	0.81	0.795	2.229	2.208
Average	-	2.603	2.534	0.718	0.695	1.89	1.825

cropping or subsampling. Thus, such techniques should be avoided during perceptual testing. For this reason, *spatial* 3AFC testing was found to be inappropriate for this work. *Sequential* 3AFC testing has been used previously in subjective *sound* quality evaluations [41]. In this method, three audio clips were presented sequentially to subjects. Two of the clips were original and one of the clips had been compressed. The subjects were forced to choose the clip that was different from the other two.

In this work, we employed sequential 3AFC testing for the validation of compressed monochrome stereo pairs. Fifteen subjects with normal or corrected-to-normal vision participated in the validation. Each subject viewed the full sequence of thirteen stereo pairs a total of six times. The order of the thirteen stereo pairs was randomized in each of the six sequences. The total number of subjective trials was then $15 \times 13 \times 6 = 1170$. In each such trial, one compressed copy and two original copies of a stereo image were

TABLE V
THE RESULTING SAMPLE PROPORTION OF CORRECT CHOICES FOR STEREOSCOPIC 3D IMAGES

	Image												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Number of Correct Choices n_C	27	27	28	32	31	27	30	28	27	30	30	29	31
Sample Proportion $\hat{p}$	0.3	0.3	0.311	0.356	0.344	0.3	0.333	0.311	0.3	0.333	0.333	0.322	0.344
95% Confidence Interval	[0.205, 0.395]	[0.205, 0.395]	[0.215, 0.407]	[0.257, 0.454]	[0.246, 0.443]	[0.205, 0.395]	[0.236, 0.431]	[0.215, 0.407]	[0.205, 0.395]	[0.236, 0.431]	[0.236, 0.431]	[0.226, 0.419]	[0.246, 0.443]

displayed in stereo mode, with the subjects wearing 3D shutter glasses. These three copies were displayed sequentially in random order. The subjects could move backward/forward between the three copies of the image via the left/right keys on the keyboard. The subjects were allowed to observe each copy of the image as many times and for however long they wanted. To avoid the phenomenon of persistence of vision, a black screen was displayed for 0.5 seconds while switching between copies. Despite the fact that our VTs were designed at a typical viewing distance $\upsilon = 60$ cm between subject and display, the subjects were allowed to approach the screen as closely as they wanted during the validation. According to a number shown in the upper left corner of the screen, the subjects were asked to use the keyboard to indicate which copy was different from the other two. No guidance was provided to the subjects regarding the types of differences that they might encounter, e.g., image quality, depth, naturalness, or discomfort. Additionally, no feedback was provided to the subjects to indicate whether their choices were correct.

It is possible that the 3AFC test detailed above may bias the observer by implying that one image is different. However, it is worth noting that any resulting bias would be *against* the hypothesis that the images are visually lossless, and thus would not invalidate a finding that the images were indeed visually lossless.

For a given compressed stereo pair, the resulting sample proportion of correct choices obtained from the subjective validations is computed as

$$\hat{p} = n_C / n, \tag{7}$$

where n_C is the number of correct choices out of the $n = 15 \times 6 = 90$ trials for that stereo pair. Under the assumption that the images are visually lossless, the population proportion of correct choices would be 1/3. A confidence interval for the population percentage can be calculated [42] via

$$\hat{p} \pm Z^* \times \sqrt{\hat{p}\left(1-\hat{p}\right)/n}, \tag{8}$$

where the value of Z^* is tabulated in [42] according to the desired confidence level. For a 95% confidence interval, the value of Z^* is 1.96. The number of correct responses for each image, together with the resulting values of $\hat{p}$ and 95% confidence intervals are provided in Table V. Figure 7 presents the same information in graphical form. Specifically, the height of the blue bars represent the values of $\hat{p}$, while the green

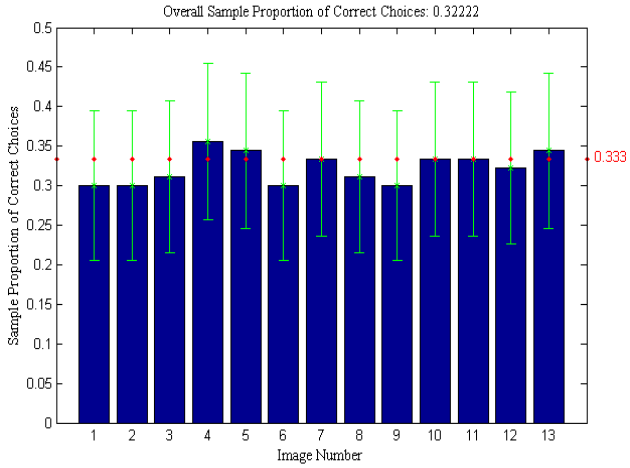


Fig. 7. Validation results. The blue rectangular bars represent the individual sample proportion of correct choices for each image individually. The green vertical bars denote the 95% confidence interval for each image separately. The assumed population proportion of 1/3 is represented by the red dotted line.

error bars indicate the 95% confidence intervals. The value of the hypothesized population fraction of correct responses (i.e., 1/3) is shown by the red dots. As can be seen in the figure and the table, the confidence intervals for individual images are rather loose (approximately ± 0.1). On the other hand, taken in aggregate (over all 13 images), $n_C = 377$ which results in $\hat{p} = 377/1170 = 0.322$, and a 95% confidence interval with lower and upper bounds of 0.349 and 0.295, respectively (0.322 ± 0.027). Since the hypothesized population mean of 1/3 is well within this 95% confidence interval, it is claimed that the compressed monochrome stereo pairs are visually lossless.

It should be emphasized that these results are for the specific 3D stereoscopic display system and lighting conditions described in Section III. On the other hand, results for other display systems and/or lighting conditions may be obtained via the proposed methodology. Although only images from the Middlebury stereo datasets [35]–[38] were employed in the formal validation studies, results are similar for other stereo images. In particular, 3D images of natural scenes were obtained by extracting the first frame from video sequences in the RMIT3DV database [34], [43], [44]. Informal validation studies verify that visually lossless quality is also obtained for these images. This is not surprising since no actual images were employed in the design of VTs.

Indeed, only pseudo-randomly generated stimulus images were used in this regard. Thus, while the compressor depends on the particular stereoscopic display system and viewing conditions employed, it does not depend on any particular image source.

VI. Conclusions

A methodology for visually lossless compression of monochrome stereoscopic 3D images was provided. Since the crosstalk effect is an inherent perceivable problem in current 3D display technologies, the measurement of VTs in this work considered not only the factors characterized for 2D images in [12], but also the various combinations of luminance values in both the left and right channels of stereoscopic images. To ensure that neither eye can perceive quantization errors when crosstalk occurs, the final VT for the left image is taken as the minimum between the VTs of the left eye and the right eye for distortion introduced in only the left image. The same considerations apply to the right image, simply reversing the roles of left and right.

It is prohibitive to measure VTs for all possible combinations of coefficient variance and left/right image gray levels. Thus, a separable model for VTs was proposed. The VTs obtained via the proposed model were then employed in the development of a visually lossless coding scheme for monochrome stereoscopic images. Compressed codestreams created via the proposed encoder can be decompressed using any JPEG2000 compliant decoder.

The bitrate required for visually lossless compression of 3D monochrome stereo pairs is larger than that required for visually lossless 2D compression of the individual left and right images. However, the resulting left and right images obtained via the proposed method are visually lossless in both 2D and 3D mode, while the images compressed individually are not visually lossless in 3D mode. Additionally, the proposed method results in a significantly lower bit rate than required for information lossless encoding, while still resulting in visually lossless quality. Sequential 3AFC testing was conducted to validate that compressed monochrome stereo pairs obtained from the proposed system are visually lossless.

References

[1] A. P. Dal Poz, R. A. B. Gallis, J. F. C. da Silva, and E. F. O. Martins, "Object-space road extraction in rural areas using stereoscopic aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 654–658, Jul. 2012.

[2] D. P. Noonan, P. Mountney, D. S. Elson, A. Darzi, and G.-Z. Yang, "A stereoscopic fibroscope for camera motion and 3D depth recovery during minimally invasive surgery," in *Proc. IEEE Int. Conf. Robot. Autom.*, Kobe, Japan, May 2009, pp. 4463–4468.

[3] L. Lipton, "The stereoscopic cinema: From film to digital projection," *SMPTE J.*, pp. 586–593, Sep. 2001.

[4] D. A. Bowman, *3D User Interfaces: Theory and Practice*. Boston, MA, USA: Addison-Wesley, 2005.

[5] S. Pastoor and M. Wöpking, "3D displays: A review of current technologies," *Displays*, vol. 17, no. 2, pp. 100–110, Apr. 1997.

[6] H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the art in stereoscopic and autostereoscopic displays," *Proc. IEEE*, vol. 99, no. 4, pp. 540–555, Apr. 2011.

[7] M. Barkowsky, S. Tourancheau, K. Brunnström, K. Wang, and B. Andrén, "Crosstalk measurements of shutter glasses 3D displays," in *Proc. SID Int. Symp.*, 2011, pp. 812–815.

[8] S. Shestak, D. Kim, and S. Hwang, "Measuring of gray-to-gray crosstalk in a LCD based time-sequential stereoscopic display," in *SID Dig.*, Seattle, WA, USA, May 2010, pp. 132–135.

[9] C.-C. Pan, Y.-R. Lee, K.-F. Huang, and T.-C. Huang, "Cross-talk evaluation of shutter-type stereoscopic 3D display," in *SID Dig.*, Seattle, WA, USA, May 2010, pp. 128–131.

[10] J. D. Yun, Y. Kwak, and S. Yang, "Evaluation of perceptual resolution and crosstalk in stereoscopic displays," *J. Display Technol.*, vol. 9, no. 2, pp. 106–111, Feb. 2013.

[11] P. Boher, T. Leroux, V. C. Patton, and T. Bignon, "Optical characterization of shutter glasses stereoscopic 3D displays," *Proc. SPIE*, vol. 7863, p. 786312, Feb. 2011.

[12] H. Oh, A. Bilgin, and M. W. Marcellin, "Visually lossless encoding for JPEG2000," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 189–201, Jan. 2013.

[13] S.-M. Jung *et al.*, "Improvement of 3D crosstalk with over-driving method for the active retarder 3D displays," in *SID Symp. Dig. Tech. Papers*, Seattle, WA, USA, May 2010, pp. 1264–1267.

[14] J. Konrad, B. Lacotte, and E. Dubois, "Cancellation of image crosstalk in time-sequential displays of stereoscopic video," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 897–908, May 2000.

[15] S. Pastoor, "Human factors of 3D imaging: Results of recent research at Heinrich-Hertz-Institute Berlin," in *Proc. 2nd Int. Display Workshop*, 1995, pp. 69–72.

[16] L. Wang *et al.*, "Crosstalk evaluation in stereoscopic displays," *J. Display Technol.*, vol. 7, no. 4, pp. 208–214, Apr. 2011.

[17] H. C. Feng, M. W. Marcellin, and A. Bilgin, "Measurement of visibility thresholds for compression of stereo images," in *Proc. Int. Telemetering Conf.*, San Diego, CA, USA, Oct. 2012.

[18] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiol.*, vol. 197, pp. 551–566, Aug. 1968.

[19] S. J. Daly, "Visible differences predictor: An algorithm for the assessment of image fidelity," *Proc. SPIE*, vol. 1666, pp. 2–15, Aug. 1992.

[20] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1164–1175, Aug. 1997.

[21] H. Oh, A. Bilgin, and M. W. Marcellin, "Multi-resolution visually lossless image coding using JPEG2000," in *Proc. IEEE Int. Conf. Image Process.*, Hong Kong, Sep. 2010, pp. 2581–2584.

[22] H.-C. Feng, M. W. Marcellin, and A. Bilgin, "Visually lossless compression of stereo images," in *Proc. IEEE Data Compress. Conf.*, Snowbird, UT, USA, Mar. 2013, p. 490.

[23] H. C. Feng, M. W. Marcellin, and A. Bilgin, "Validation for visually lossless compression of stereo images," in *Proc. Int. Telemetering Conf.*, Las Vegas, NV, USA, Oct. 2013.

[24] M. Menozzi, U. Näpflin, and H. Krueger, "CRT versus LCD: A pilot study on visual performance and suitability of two display technologies for use in office work," *Displays*, vol. 20, no. 1, pp. 3–10, Feb. 1999.

[25] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R Rec. BT.500-12, Sep. 2009.

[26] *Subjective Assessment Methods for Image Quality in High-Definition Television*, document ITU-R Rec. BT.710-4, Nov. 1998.

[27] A. J. Ahumada, Jr., and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," *Proc. SPIE*, vol. 1666, pp. 365–374, Aug. 1992.

[28] C.-H. Chou and C.-W. Chen, "A perceptually optimized 3D subband codec for video communication over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 143–156, Apr. 1996.

[29] C.-H. Chou and Y.-C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467–476, Dec. 1995.

[30] D. H. Brainard, "The psychophysics toolbox," *Spatial Vis.*, vol. 10, no. 4, pp. 433–436, 1997.

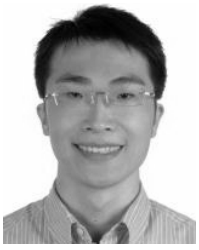
[31] A. B. Watson and D. G. Pelli, "QUEST: A Bayesian adaptive psychometric method," *Perception Psychophys.*, vol. 33, no. 2, pp. 113–120, 1983.

[32] D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Boston, MA, USA: Kluwer, 2002.

[33] G. Saygili, C. G. Gürler, and A. M. Tekalp, "Quality assessment of asymmetric stereo video coding," in *Proc. 17th IEEE ICIP*, Sep. 2010, pp. 4009–4012.

[34] H. R. Wu, A. R. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual visual signal compression and transmission," *Proc. IEEE*, vol. 101, no. 9, pp. 2025–2043, Aug. 2013.

[35] D. Scharstein and R. Szeliski. *Middlebury Stereo Datasets*. [Online]. Available: http://vision.middlebury.edu/stereo/, accessed Dec. 6, 2011.
[36] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Madison, WI, USA, Jun. 2003, pp. 195–202.
[37] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.
[38] H. Hirschmüller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.
[39] Kakadu software. [Online]. Available: http://www.kakadusoftware.com, accessed Oct. 7, 2010.
[40] D. M. Chandler, N. L. Dykes, and S. S. Hemami, "Visually lossless compression of digitized radiographs based on contrast sensitivity and visual masking," *Proc. SPIE*, vol. 5749, pp. 359–372, Apr. 2005.
[41] M. Frank, A. Sontacchi, T. Lindenbauer, and M. Opitz, "Subjective sound quality evaluation of a codec for digital wireless transmission," in *Proc. Audio Eng. Soc. 132nd Conv.*, Budapest, Hungary, Apr. 2012.
[42] D. J. Rumsey, *Statistics Workbook for Dummies*. Hoboken, NJ, USA: Wiley, 2005.
[43] *RMIT3DV: An Uncompressed Stereoscopic 3D HD Video Library*. [Online]. Available: http://www.rmit3dv.com/index.php, accessed Jun. 5, 2014.
[44] E. Cheng, P. Burton, J. Burton, A. Joseski, and I. Burnett, "RMIT3DV: Pre-announcement of a creative commons uncompressed HD 3D video database," in *Proc. 4th Int. Workshop Quality Multimedia Exper. (QoMEX)*, Jul. 2012, pp. 212–217.



Hsin-Chang Feng received the B.S. degree in electronics engineering from the Oriental Institute of Technology, Taipei, Taiwan, in 2002, and the M.S. degree in electrical engineering from Tatung University, Taipei, in 2008. He is currently pursuing the Ph.D. degree in electrical and computer engineering from the University of Arizona, Tucson, AZ, USA. His current research interests include stereoscopic 3D image, perceptual image processing, and data compression.

Michael W. Marcellin (S'81–M'87–SM'93–F'02) was born in Bishop, CA, USA, in 1959. He received the B.S. (*summa cum laude*) degree in electrical engineering from San Diego State University, San Diego, CA, USA, in 1983, where he was recognized as the most outstanding student in the College of Engineering, and the M.S. and Ph.D. degrees in electrical engineering from Texas A&M University, College Station, TX, USA, in 1985 and 1987, respectively.

He has been with the University of Arizona, Tucson, AZ, USA, since 1988, where he is currently a Regents' Professor of Electrical and Computer Engineering. He has authored or co-authored over 250 papers in international journals and conferences, and has co-authored a book entitled *JPEG2000: Image Compression Fundamentals, Standards and Practice* (Boston, MA, USA: Kluwer Academic Publishers, 2002). This book is a graduate level textbook on image compression fundamentals and the definitive reference on JPEG2000. His current research interests include digital communication and data storage systems, data compression, and signal processing.

Dr. Marcellin is a major contributor to JPEG2000, the second-generation standard for image compression. Throughout the standardization process, he was the Chair of the JPEG2000 Verification Model Ad Hoc Group, which was responsible for the software implementation and documentation of the JPEG2000 algorithm. He was a Consultant to Digital Cinema Initiatives, a consortium of Hollywood studios, on the development of the JPEG2000 profiles for digital cinema. He is a member of the Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi. He was a recipient of the National Science Foundation Young Investigator Award in 1992, the IEEE Signal Processing Society Senior (Best Paper) Award in 1993, the University of Arizona Technology Innovation Award in 2006, and Teaching Awards from Nanyang Technological University in 1990 and 2001, the IEEE/Eta Kappa Nu Student Section in 1997, and the University of Arizona College of Engineering in 2000, 2010, 2013. He was named as the San Diego State University Distinguished Engineering Alumnus in 2003. From 2001 to 2006, he was the Litton Industries John M. Leonis Distinguished Professor of Engineering. He is currently the International Foundation for Telemetering Chaired Professor of Electrical and Computer Engineering with the University of Arizona.

Ali Bilgin (S'94–M'03–SM'08) received the B.S. degree in electronics and telecommunications engineering from Istanbul Technical University, Istanbul, Turkey, in 1992, the M.S. degree in electrical engineering from San Diego State University, San Diego, CA, USA, in 1995, and the Ph.D. degree in electrical engineering from the University of Arizona, Tucson, AZ, USA, in 2002. He is currently an Assistant Professor with the Department of Biomedical Engineering and the Department of Electrical and Computer Engineering, University of Arizona. He has authored or co-authored over 175 research papers in international journals and conferences. He holds nine granted and several pending patents. Dr. Bilgin was on the Organizing Committees of many conferences, and was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 2010 to 2012 and the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2010 to 2014. He is currently an Associate Editor of the IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING. His current research interests include signal and image processing, image and video coding, data compression, and magnetic resonance imaging.