AN ANALYTICAL MODEL OF THE PERCEPTUAL THRESHOLD FUNCTION FOR MULTICHANNEL IMAGE COMPRESSION

Peter J. Hahn and V. John Mathews

Department of Electrical Engineering
University of Utah
Salt Lake City, Utah 84112, USA

ABSTRACT

The human observer is often the final judge of the quality of compressed images. One way to design a compression system that attempts to reduce or eliminate subjective distortions in the coded images is to incorporate a perceptual threshold function model into the compression system. The perceptual threshold function describes the amount of quantization error that can be introduced into a particular component of the image without introducing any visual distortions. This paper describes an analytical approach for the determination of the perceptual threshold values for use in an arbitrary multichannel image compression system. Experimental results obtained from a compression system that incorporates the perceptual threshold function are also included in the paper.

1. INTRODUCTION

In many applications of image compression, the human observer is the final judge of the quality of the compressed images. In such situations, it is important to design image compression systems that attempt to reduce or eliminate subjective distortions in the coded images. This can be accomplished by using models of the human visual system in the compression scheme [3, 5, 11]. One approach to developing perceptually-tuned image compression systems is to define a perceptual threshold associated with each component of the image, and then to design a system that constrains the quantization errors to fall below the levels suggested by the thresholds. If the thresholds are defined such that distortions with magnitudes smaller than the thresholds are not visible to human viewers, the system performs perceptually lossless image compression.

This paper describes an analytical approach for determining the perceptual threshold function (PTF) for use in an arbitrary multichannel image compression system such as subband, wavelet transform, and discrete cosine transform coders. In typical models of PTFs [3, 5, 8, 11] a base perceptual threshold value is estimated for each channel. The base perceptual threshold for a given channel is the minimum strength of the components in that channel before human viewers can detect them. Additional corrections are then made for the threshold elevation caused by the presence of signals in the channels, as well as for the differences in the detection thresholds due to variations of the local brightness values in the images. In previously available methods [3, 5, 11], the threshold values were determined by extensive empirical experimentation with human subjects. This is often time consuming and tedious for the human observers and the system designer. Our earlier work [8] derived an analytical model of the PTF for a particular type of image decomposition. In this paper, we derive an analytical method for estimating the base threshold values and the threshold elevation curves for arbitrary multichannel image decomposition and arbitrary viewing distances.

2. THE PTF MODEL

Our model estimates the perceptual threshold function at the location (x_k, y_k) in the kth channel as

$$T_k(x_k, y_k) = \tau_k E_B(x_k, y_k) E_M(x_k, y_k), \qquad (1)$$

where τ_k represents the base perceptual threshold for the kth channel, $E_B(x_k, y_k)$ represents the threshold elevation due to the average luminance in the original image in a local neighborhood of the location (x_k, y_k), and $E_M(x_k, y_k)$ represents the threshold elevation due to the energy in a neighborhood of the location (x_k, y_k) in the kth channel.

2.1. Base Threshold Values

We find a single threshold for a each channel by projecting a model of the modulation transfer function (MTF) of the human visual system onto that channel. The modulation transfer function at a given frequency $\mathbf{f}_p$

0-8186-8821-1/98 $10.00

is the reciprocal of the perceptual detection threshold of a sinusoidal stimulus with frequency $\mathbf{f}_p$. Let $X_k(\mathbf{f})$ represent the two-dimensional Fourier transform of the signal in the kth channel of the image decomposition. Here $\mathbf{f}$ is a two-dimensional spatial frequency vector. Also, let $H_k(\mathbf{f})$ denote the frequency response of the kth filter in the synthesis bank of the multichannel decomposition. The input $X_k(\mathbf{f})$ is processed by the filter $H_k(\mathbf{f})$, and the resulting output is summed with the output of the other channels to reconstruct the image.

The base threshold values for the channels are estimated by first finding the coefficients w_k that minimizes the cost function

$$\int \left| \sum_k X_k(\mathbf{f}) H_k(\mathbf{f}) (M(\mathbf{f}) - w_k) \right|^2 d\mathbf{f} \qquad (2)$$

over all frequencies, where $M(\mathbf{f})$ denotes the model of the MTF. This optimization problem seeks to find the coefficient w_k that represents the best constant approximation to the MTF in the kth channel in the image decomposition. The base threshold value τ_k for the kth channel is then estimated as

$$\tau_k = \frac{1}{w_k}. \qquad (3)$$

Let $\mathbf{W} = [w_1, w_2, \ldots, w_N]^T$, where $(\cdot)^T$ is the matrix transpose operation. It is straightforward to show that

$$\mathbf{W} = \mathbf{R}^{-1}\mathbf{P}, \qquad (4)$$

where the ith element of the vector $\mathbf{P}$ is given by

$$\begin{aligned} \mathbf{P}_i &= \sum_{k=1}^{N} \int (\Re\{X_i(\mathbf{f}) H_i(\mathbf{f})\} \Re\{X_k(\mathbf{f}) H_k(\mathbf{f})\} \\ &+ \Im\{X_i(\mathbf{f}) H_i(\mathbf{f})\} \Im\{X_k(\mathbf{f}) H_k(\mathbf{f})\}) M(\mathbf{f}) d\mathbf{f}, \end{aligned} \qquad (5)$$

and the (i,j)th element of the matrix $\mathbf{R}$ is given by

$$\begin{aligned} \mathbf{R}_{ij} &= \int (\Re\{X_i(\mathbf{f}) H_i(\mathbf{f})\} \Re\{X_j(\mathbf{f}) H_j(\mathbf{f})\} \\ &+ \Im\{X_i(\mathbf{f}) H_i(\mathbf{f})\} \Im\{X_j(\mathbf{f}) H_j(\mathbf{f})\}) d\mathbf{f}. \end{aligned} \qquad (6)$$

In the above expressions $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real part and imaginary part, respectively, of $\{\cdot\}$.

2.2. Brightness Correction

The brightness correction factor makes the assumption that the contrast for detection C is constant at all brightness levels of interest [2]. In other words,

$$C = \frac{\Delta I}{I_{ave}} = k_w \qquad (7)$$

where I_{ave} is the average intensity value in a local region, ΔI is the increase in intensity required for a just noticeable difference, and k_w is a constant value. This relationship is often referred to as Weber's ratio. It is known that this relationship does not hold for all values of I_{ave}, but for the case of images displayed on a monitor, this is a reasonable approximation. The brightness correction used in our model is

$$E_B(x_k, y_k) = \frac{k_w I(x_k, y_k)}{k_w I_{th}} = \frac{I(x_k, y_k)}{I_{th}} \qquad (8)$$

where $I(x_k, y_k)$ is the average intensity in a local neighborhood of the location (x_k, y_k), and I_{th} is the average intensity at which the base thresholds were found.

2.3. Masking Correction

The presence of other frequency components in an image will change the detection threshold of the frequency of interest. We will refer to the frequency of interest as the target frequency and the frequency of the interfering signal components as the masking frequency. The masking component must have significant contrast and be similar in frequency and orientation to the target frequency for it to increase the perceptual threshold of the target frequency. The threshold elevation is a function of the target frequency $\mathbf{f}_T$, the masking frequency $\mathbf{f}_m$, and the contrast of the masking frequency component C_m [6]. We have ignored the small decrease in the threshold, a phenomenon known as facilitation [6], that has been observed when the masking contrast is relatively small.

The threshold elevation is modeled as

$$T_e = K C_m^d W_o(\mathbf{f}_m, \mathbf{f}_T) W_m(\mathbf{f}_m, \mathbf{f}_T) \qquad (9)$$

where K and d are constants. The correction factor W_o accounts for the dependence of the masking effect on the relative orientation of the frequency $\mathbf{f}_m$ with respect to the frequency $\mathbf{f}_T$. Similarly, W_m accounts for the dependence of the masking effect on the difference in the magnitudes of the two frequency components. The correction for the orientation difference between the two frequency vectors is modeled as a triangular function between a distance of $\pm 30°$ [6] and is given by

$$W_o(\mathbf{f}_m, T) = \begin{cases} 1 - \frac{\Delta(\mathbf{f}_m, \mathbf{f}_T)}{30°}; & \text{for } \Delta(\mathbf{f}_m, \mathbf{f}_T) < 30°, \\ 0; & \text{otherwise,} \end{cases} \qquad (10)$$

with $\Delta(\mathbf{f}_m, \mathbf{f}_T) = |\angle(\mathbf{f}_m) - \angle(\mathbf{f}_T)|$ and $\angle(\mathbf{f}) = \tan^{-1}(\frac{f_y}{f_x})$. The correction for the distance between the frequency vectors is also modeled as a triangular function within

± 1 octave, and is given by

$$W_m(\mathbf{f}_m, \mathbf{f}_T) = \begin{cases} 1 - \frac{\log(\frac{|\mathbf{f}_m|}{|\mathbf{f}_T|})}{\log(2)}; & \text{for } \log(\frac{|\mathbf{f}_m|}{|\mathbf{f}_T|}) < \log(2), \\ 0; & \text{otherwise.} \end{cases} \tag{11}$$

Equation (9) shows how the signal strength at a specific frequency effects the visibility of a signal at another frequency. Since the signal in any channel contains numerous frequency components, the model attempts to find a single value for the threshold elevation caused by all the components in a particular channel. This may be done in a manner similar to the method employed for finding the base threshold values by projecting the function defined by (9) onto the channel of interest. This is accomplished by minimizing the cost function

$$\int\int |X_k(\mathbf{f}_T)H_k(\mathbf{f}_T)|^2 \, |(T_e(C_m(\mathbf{f}_m), \mathbf{f}_m, \mathbf{f}_T) - E_M(x_k, y_k)|^2 d\mathbf{f}_T d\mathbf{f}_m \tag{12}$$

with respect to $E_M(x_k, y_k)$. The contrast $C_m(\mathbf{f}_m)$ is estimated from the local energy around (x_k, y_k) in the channel as

$$C_m(\mathbf{f}_m) = \frac{|X_k(\mathbf{f}_m)H_k(\mathbf{f}_m)|}{I_{ave}}, \tag{13}$$

where I_{ave} is the local average intensity. For a given image and decomposition or an assumed statistical model, we can find $X_k(\mathbf{f}_m)H_k(\mathbf{f}_m)$. Minimizing the cost function (12) gives

$$E_M(x_k, y_k) = \frac{\int\int |X_k(\mathbf{f}_T)H_k(\mathbf{f}_T)|^2 T_e d\mathbf{f}_m d\mathbf{f}_T}{\int\int |X_k(\mathbf{f}_T)H_k(\mathbf{f}_T)|^2 d\mathbf{f}_m d\mathbf{f}_T}. \tag{14}$$

In this model, we assumed the masking effects from other channels on the kth channel were negligible. This model can be extended in a straightforward manner to include these effects.

3. MODEL VERIFICATION

The model described in the previous section was verified using two different approaches. In the first approach, we compared the analytically obtained threshold values with empirically measured values. In the second approach, we incorporated the analytically derived perceptual threshold function into a perceptually lossless compression system to check if the quantization errors can be detected by human viewers. In both cases we employed a multichannel decomposition of the input images using a five-level wavelet transform and the 9-7 tap filters described in [1] for a viewing distance of six times the image height.

Band Num	Level Num	Band Type	Analytical threshold	Empirical threshold
1	5	LL	2.5	2.5
2	5	LH	2.5	2.5
3	5	HL	2.5	2.5
4	5	HH	2.5	2.3
5	4	LH	2.4	2.1
6	4	HL	2.4	2.1
7	4	HH	2.3	2.1
8	3	LH	2.3	2.1
9	3	HL	2.3	2.1
10	3	HH	2.1	2.1
11	2	LH	2.4	2.5
12	2	HL	2.4	2.5
13	2	HH	2.5	3.2
14	1	LH	4.2	3.8
15	1	HL	4.2	3.8
16	1	HH	7.3	7.6

Table 1: The thresholds for a viewing distance of six times the image height by the theoretical and empirical methods.

3.1. Analytical Versus Empirical Thresholds

The perceptual threshold values were measured using psychophysical experiments. These experiments employed the "forced choice" paradigm. In these experiments, the observers are shown two images displayed side by side on a monitor. The observers have to pick between the same two images several times in an experiment. The image positions are randomly switched for each choice. If one image is chosen significantly more frequently than the other, the images are considered visually different. On the other hand, if neither image is chosen significantly more frequently, the images are considered perceptually identical. The psychophysical experiments were performed in a dimly lit room with reflections on the monitor minimized. In the experiments to measure the base perceptual threshold, one of the images has constant value at all locations and the other image is a corrupted version of that image obtained by adding uniformly-distributed white noise in the range $[-\delta_k, \delta_k]$ to the kth channel. The parameter δ_k for which the difference between the two images was barely visible is the estimated threshold for the channel.

Table 1 compares the theoretical values of the base thresholds obtained from our model and the experimental results. To derive the thresholds using our model, the input $X_k(\mathbf{f})$ for each channel was modeled as white and uniformly distributed noise. The numeri-

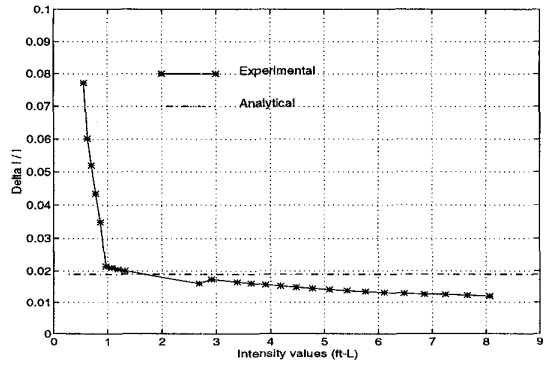

Figure 1: Comparison between the analytical and empirically derived change in intensity over intensity versus intensity for brightness for subband HH1.

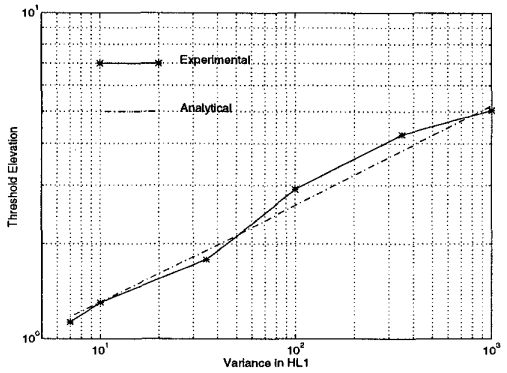

Figure 2: Comparison between the analytical threshold elevation and the empirically derived for subband HL1.

cal integration technique used was the 4-5 Runge-Kutta method. The modulation transfer function used in the computations is related to the the closed form expression given by [7]

$$M(f_p) = 2.6[0.0192 + 0.114 f_p] \exp(-(0.114 f_p)^{1.1}). \quad (15)$$

where f_p is the polar frequency $f_p = \sqrt{f_x^2 + f_y^2}$ and f_x and f_y are the frequency variables in cycles per degree in the x and y directions. In order to simplify the calculations, we used a separable approximation of (15) in our estimates. We can see from the results of the table that there is reasonably good agreement between the analytically obtained values and the experimentally measured values of the base perceptual thresholds for the channels.

The brightness correction factor was measured using similar experiments performed for different values of the mean gray level of the images. Due to time constraints, these experiments were performed on the highest frequency subband $HH1$ and extrapolated to the other bands. The comparison between the experimental and analytical brightness correction factors are shown in Figure 1. The graph shows the change in intensity divided by the average intensity value versus the average intensity value. For low average intensity values, the graph is much higher than predicted because of a saturation effect of the monitor used in the experiments. The experimental values for the rest of the graph are relatively close to the expected Weber's ratio of approximately 0.019 [9].

The threshold elevation due to energy in one channel masking the coefficients in the same channel was also found using forced choice experiments. In this case, the original image was corrupted with uniformly distributed white noise with a specified variance level injected in the channel. The second image was generated by adding uniformly distributed noise in the range $[-\eta_k, \eta_k]$ to the channel. The experimental results showed that the threshold elevation curves exhibited an exponential relationship between the multiplicative factor of the threshold elevation and the energy in the subband. The comparison for one subband is shown in Figure 2. The analytical obtained values of the threshold elevation are a close match to the threshold elevation found by the empirical forced choice technique.

3.2. Perceptually Lossless Image Compression

We have made modifications to the set partitioning in hierarchical trees (SPIHT) algorithm [10] to allow the coding process to be guided by a model of the perceptual threshold function. The modifications allow the algorithm to stop coding when all quantization errors are below the levels suggested by the perceptual threshold function. The modified SPIHT coder that incorporates the perceptual threshold function in the coding process is described in [4]. Both the local average values for the brightness correction and the variances for the masking elevation were found over a region twice the size of the plane of support of the wavelet filter.

The results of compressing the commonly used Lena image is displayed in Figures 3 and 4. Figure 3 shows the test image which contains 512×512 pixels at 8 bits per pixel grayscale resolution. The perceptual threshold model was derived for a viewing distance of six times the image height. The coding process requires 0.56 bits per pixel to bring all the quantization errors to values below those suggested by our model. The compressed image is shown in Figure 4. We can see few if any distortions in this image when viewed from the

Figure 3: Original Lena image.

Figure 4: Lena image compressed to 0.56 bits/pixel using the modified SPIHT algorithm.

appropriate distance implying that our model predicts the perceptual threshold function reasonably well.

4. CONCLUSION

This paper presented an analytical technique for modeling the perceptual threshold function for arbitrary multichannel image decompositions. Deriving an analytical model such as the one described in this paper is important since the perceptual threshold function must otherwise be found using empirical techniques. The experiments conducted to verify the usefulness of the model presented in the paper showed that the thresholds found empirically and those obtained using the analytical techniques matched reasonably well. We have compressed a large number of monochrome still images using the modified SPIHT algorithm. Perceptually lossless compression was achieved at bit rates in the range of 0.4 to 1.1 bits per pixel in our experiments.

5. REFERENCES

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205-220, Apr. 1992.

[2] T. Cornsweet, *Visual Perception*, New York: Academic Press, 1970.

[3] P. J. Hahn and V. J. Mathews, "Perceptually lossless image compression," *1997 Data Compression Industry Workshop*, pp. 77-86, Snowbird, UT, Mar. 1997.

[4] P. J. Hahn and V. J. Mathews, "A perceptually-tuned image compression system," *Eighth IEEE Dig. Sig. Proc. Workshop*, Bryce Canyon, UT Aug. 1998.

[5] N. S. Jayant, J. D. Johnston and R. J. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, Vol. 81, No. 10, pp. 1385-1424, Oct. 1993.

[6] G. E. Legge, "A power law for contrast discrimination," *Vision Research*, vol. 21, pp. 457-467, March 1982.

[7] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Information Theory*, vol. 20, no. 4, pp. 525-536, July 1974.

[8] K. S. Prashant, V. J. Mathews, and P. J. Hahn, "A new model for perceptual threshold functions for application in image compression systems," *Proc. Data Compression Conf.*, pp. 371-380, Snowbird, UT, Mar. 1995.

[9] C. A. Poynton, "Gamma and its disguises: The nonlinear mappings of intensity in perception, CRTs, film, and video," *SMPTE Journal*, pp. 1099-1108, Dec. 1993.

[10] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, pp. 243-250, June 1996.

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