PERCEPTUALLY LOSSLESS IMAGE COMPRESSION

Peter J. Hahn and V. John Mathews
Department of Electrical Engineering
University of Utah
Salt Lake City, Utah 84112, USA

ABSTRACT

This paper presents an algorithm for perceptually lossless image compression. A compressed image is said to be perceptually lossless for a specified viewing distance if the reconstructed image and the original image appear identical to human observers when viewed from the specified distance. Our approach utilizes properties of the human visual system in the form of a perceptual threshold function model to determine the amount of distortion that can be introduced at each location of the image. Constraining all quantization errors to be below the perceptual threshold function results in perceptually lossless image compression. The compression system employs a modified form of the embedded zerotree wavelet coding algorithm to limit the quantization errors below the levels specified by the model of the threshold function. Experimental results demonstrate perceptually lossless compression of monochrome images at bit rates ranging from 0.4 to 1.2 bits per pixel at a viewing distance of six times the image height. These results were obtained using a simple, empirical model of the perceptual threshold function which included threshold elevations for the local brightness and local energy in neighboring frequency bands.

1. INTRODUCTION

The motivation behind the use of image compression is evident to most people. Storing the rapidly expanding amounts of digitally represented images requires some form of compression. Medical imaging, satellite imaging, and television are just a few of the different types of media that can or do use digital images. In many instances, human observers judge 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. Perceptually lossless image compression attempts to eliminate all subjective distortions from the coded images. Thus, the reconstructed image *looks* identical to the original image although there may be large numerical differences between the pixel values of the original and reconstructed images. It is important to recognize that the perceptual quality of an image is a function of the viewing distance. For example, a distorted image may look identical to the original image when viewed from a sufficient distance, but the distortions may become visible when the observer moves closer. Consequently, the notion of perceptually lossless compression is also a function of the viewing distance.

The objective of this paper is to present an algorithm for image compression that employs a perceptual threshold function model and produces coded images with distortions that are below the thresholds defined by the model at all locations in the image. Our approach employs a wavelet transform decomposition of the input images. Consequently, it requires a perceptual threshold function model that makes use of the same image decomposition. The algorithm uses the embedded zerotree wavelet (EZW) algorithm that has been modified to permit the termination of the coding process for a wavelet transform coefficient when the quantization error is below the perceptual threshold for the particular coefficient.

The rest of this paper is organized as follows. The next section describes some known properties of the human visual system, some coders that use properties of the human visual system, and some models of the perceptual threshold function. The modified embedded zerotree wavelet algorithm is described in Section 3. Several experimental results are presented in Section 4. Finally, the concluding remarks are made in Section 5.

2. BACKGROUND

The human visual system (HVS) is extremely complex and still not completely understood. We provide a brief review of some properties of the human visual system that are relevant to the work described in this paper. The human visual system exhibits highly-varying responses to single-frequency sinusoidal stimuli. The modulation transfer function (MTF) [1, 2, 3] characterizes the response of the HVS to different frequency components. The MTF at a given frequency is a scaled version of the reciprocal of the minimum amplitude of a sinusoidal stimulus that is visible to the human visual system. One commonly used model of the MTF is [3]

$$H(f) = a[b + cf] \exp(-(cf)^d), \tag{1}$$

where a, b, c, and d are constant parameters and f is the radial frequency given by $f = \sqrt{f_x^2 + f_y^2}$, measured in cycles per degree. The variables f_x and f_y represent the spatial frequency in the x and y coordinates, respectively.

The sensitivity of the HVS to a stimulus depends on the average level of illumination of the scene. A simple characterization of this dependence is given by Weber's law [1, 2], although this condition does not hold for all levels of illumination. Weber's law states that the ratio of the minimum amount of visually detectable change ΔI in the light intensity in a uniform background with intensity I is a constant, *i.e.,*

$$\frac{\Delta I}{I} = k, \tag{2}$$

where k is the constant of proportionality.

The concept of the spatial masking of a stimulus by a complex background scene is important from the perspective of coding. Spatial masking defines the effect where the visibility or detectability of a stimulus changes in the presence of another visual stimulus known as the masking stimulus [4]. Since the masked stimulus can accept more quantization noise without visual distortions, knowledge of the spatial masking property in an image is important for coding. These properties of the human visual system as well as others have formed the basis for the development of image compression systems that attempt to reduce or eliminate visual distortions during the coding process.

We briefly describe several approaches to perceptually-tuned image compression below. One class of perceptually-tuned coders transforms the images into the "visual domain" using a HVS model. The transformed images are then compressed using a lossy compression scheme that minimizes some quantitative distortion measure. During reconstruction the images are transformed back to the original domain [5]. Assuming the HVS model is accurate, it may be argued that the transformed image is what the "eye sees." Therefore, by minimizing the quantitative distortion measure in the visual domain, one minimizes the perceptual distortions introduced by the coder.

Another class of algorithms attempts to reduce the perceptual distortions by weighting different components in an image based on the response of the human visual system to such components [6, 7, 8]. One example of such a system is described in [6] in which the distortions of blocks with edge patterns are weighted differently from the distortions of blocks with shade patterns or more uniform patterns. A second example weights the distortions in the signal components of the different bands of a subband or transform coder on the basis of the response of the HVS to different frequency components [7]. A third example weights the distortions in different blocks of an image on the basis of the activity in each block [8].

The last class of perceptually-tuned coders discussed in this paper uses a model of the perceptual threshold function (PTF) [9, 10, 11, 12, 13]. A perceptual threshold function predicts whether a particular component of the image is visible. If this component is visible, the PTF further predicts the amount of quantization noise that can be introduced to the component before the distortion is

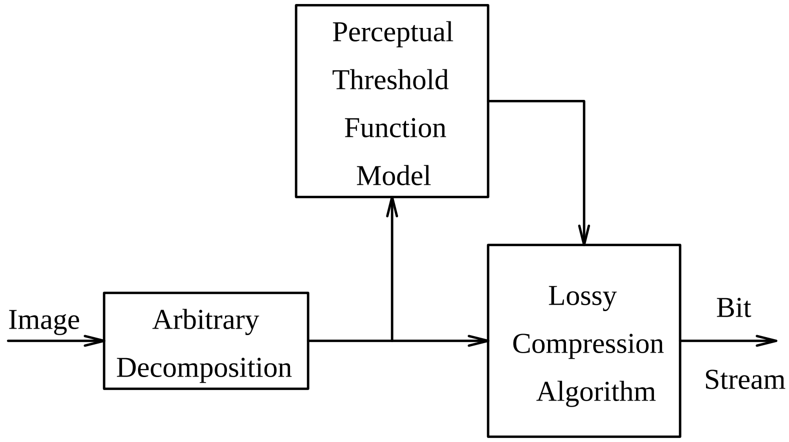

Figure 1: Block diagram of the coder with a perceptual threshold function.

visible. These components can be pixels of the image or coefficients of an appropriate transform or decomposition that completely characterizes the image.

An accurate model of the perceptual threshold function can be employed to guide perceptually lossless image compression systems. A block diagram of such a system is shown in Figure 1. The key is to constrain the quantization errors below the levels predicted by the model of the perceptual threshold function. When the image compression system meets this constraint, the distortions introduced into the image will not be visible to the observer if the PTF model is accurate. The perceptual threshold function model depends on the viewing distance, and thus the viewing distance is an important parameter in the design of perceptually lossless image compression systems.

A number of different models of the perceptual threshold function are available in the literature [11, 12, 13]. Most models of the PTF perform a multichannel decomposition of the image and then estimate a base perceptual threshold value for each channel. This threshold value predicts the distortions that can be introduced into flat areas of the input images with no local activity. The base perceptual threshold function is modified using threshold elevation functions that depends on the local brightness of the image, measures of the local activity in the image, and other spatial masking effects. The particular perceptual threshold function model used in this work is based on the work by Safranek and Johnston [12, 13]. In this model, the base perceptual threshold function was determined by empirical experiments. The model employed a subband decomposition of the image. The base threshold value for the kth subband was measured by adding uniformly distributed noise to an area in the center of the kth subband. The corrupted image was reconstructed, and observers seated at a preselected distance were asked to determine if the distortions were visible. The maximum distortion of the noise was adjusted until the distortion was just below the threshold of visual detection. This amplitude was taken to be the base perceptual threshold for the subband. This process was repeated for each subband. The base perceptual threshold function was modified by multiplicative corrections for brightness and texture energy which is a measure of the local activity in the image. The modification for brightness was found empirically using experiments similar to the ones described above except that the mean level of the image was changed. The texture energy uses a measure of the variability in the image along with the modulation transfer function to increase the threshold in the areas of high activity.

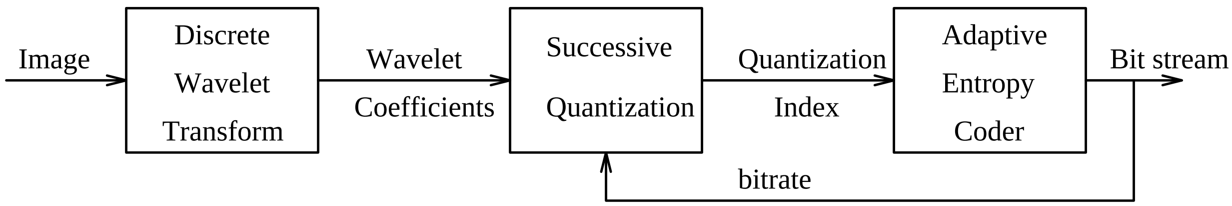

Figure 2: Block diagram of the EZW algorithm.

3. THE PERCEPTUALLY-TUNED EZW ALGORITHM

In this work, the embedded zerotree wavelet (EZW) coder [14] is modified 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 embedded zerotree wavelet algorithm compresses the input images using scalar quantization of the wavelet transform coefficients, but it does the coding such that the images at lower bit rates are "embedded" in the bit stream. The algorithm first computes the wavelet transform of the input image. The coding of the coefficients occurs next with the resulting symbols going to an adaptive arithmetic coder for further compression. The concept of a zerotree consists of a "tree" structure of insignificant wavelet coefficients that are related to each other through their position in the original image. Figure 2 shows the block diagram of the original embedded zerotree wavelet algorithm.

The coding of the wavelet coefficients takes place in a number of iterations. Each iteration consists of two "passes" for the coding of the wavelet transform coefficients. The first or "dominant" pass tags the coefficients that are currently significant and quantizes these coefficients. During an iteration, a working threshold is used to determine which coefficients are significant. The coefficients with magnitudes larger than the threshold are the ones tagged as significant. In the dominant pass, the wavelet transform coefficients can be coded as one of four symbols: *pos*, *neg*, *zero*, or *zerotree*. When the magnitude of the coefficient is larger than the working threshold, the *pos* or *neg* symbols are used depending on whether the coefficient is positive or negative respectively. Otherwise, the *zero* or *zerotree* symbol is used. The zerotree constists of a tree structure where all wavelet coefficients have magnitudes less than the working threshold. If the zerotree does not exist, the *zero* symbol is sent. The quantization levels of the tagged wavelet coefficients are from the set $\{Q, 0, -Q\}$ where $Q = \frac{3}{2}T$ where T is the working threshold. The initial threshold value is selected to be slightly larger than half the maximum magnitude of all the wavelet coefficients. At the end of an iteration the threshold is reduced by a factor of two. Once a coefficient is tagged, it is ignored in subsequent dominant passes.

The second or "subordinate" pass continues the quantization process on all the coefficients tagged in the dominant pass from this and all previous iterations. During the subordinate pass, an additional bit is added to the quantization of all coefficients that had previously been tagged. The quantization levels are changed by an additive factor of $Q_s = \pm\frac{T}{4}$ depending on the quantization error. The entire sequence of symbols from both the dominant and subordinate passes are sent to the adaptive arithmetic coder. The coding process is stopped when a predetermined bit rate is achieved. The algorithm is described in more detail in [14].

In order to use a perceptual threshold function with the EZW algorithm, several modifications must be made. First, all wavelet transform coefficients that are below the quantization error levels suggested by the perceptual detection threshold can be ignored. Second, all wavelet transform coefficients coded with a quantization error magnitude less than the perceptual threshold do not need to be coded any further. The new algorithm stops coding when all the wavelet transform coefficients are either below the perceptual detection threshold function or have a quantization error with magnitude less than the perceptual threshold function. The following sequence of operations describes the perceptually-tuned EZW algorithm. Note that this algorithm uses only one "pass" during each iteration.

1. Compute the two-dimensional wavelet transform of the input image.
2. Generate the perceptual threshold function using the transformed image and an appropriate model of the PTF.
3. Place the transformed image in the workspace of the coder.
4. Remove all wavelet coefficients with magnitudes below the perceptual detection threshold by placing zeros in the workspace for those coefficients.
5. Choose the initial working threshold T to be slightly larger than one-third of the maximum magnitude of the wavelet transform coefficients.
6. Using the working threshold, select the coefficients with magnitudes larger than the threshold and send the appropriate symbol *pos*, *neg*, *zero*, or *zerotree* for the coefficient to the adaptive arithmetic coder. The quantization levels are from the set $\{Q_p, 0, -Q_p\}$ where $Q_p = 2T$.
7. Replace the wavelet coefficient with the residual error in the workspace.
8. If the magnitude of the wavelet coefficient is smaller than the perceptual threshold function, replace the residual error with a zero in the workspace.
9. Reduce the threshold by a factor of three.
10. Repeat steps 6 through 9 until all workspace values are zero.

4. EXPERIMENTAL RESULTS

The experimental results were obtained using a compression algorithm that employs a six-level wavelet transform decomposition. The wavelet transform used filter banks with the filters of nine and seven coefficients found in [15]. For the purpose of testing the algorithm, a simple, empirically derived perceptual threshold function model was used. This perceptual threshold function model used a base perceptual threshold value for each subband in the six-level wavelet transform decomposition, a brightness correction for the average luminance in the local area, and a threshold elevation due to the masking of one subband by the local energy in other subbands.

The perceptual threshold function was estimated using psychophysical experiments with the "forced choice" paradigm. The observer is given a sequence of choices where a choice consists of displaying two images on a monitor simultaneously and asking the observer to choose the better image. The image positions are randomly switched between choices. The objective of the experiment is to determine if the images are perceptually identical. If one image is chosen significantly more frequently, the images are considered visually different, while 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 the images displayed on a standard Sun SPARC 2 computer monitor. Reflections on the monitor were minimized by illuminating the wall behind the monitor and by using black cloth draped over surfaces that reflected light onto the monitor. The observers were trained and knew the objective of the experiment.

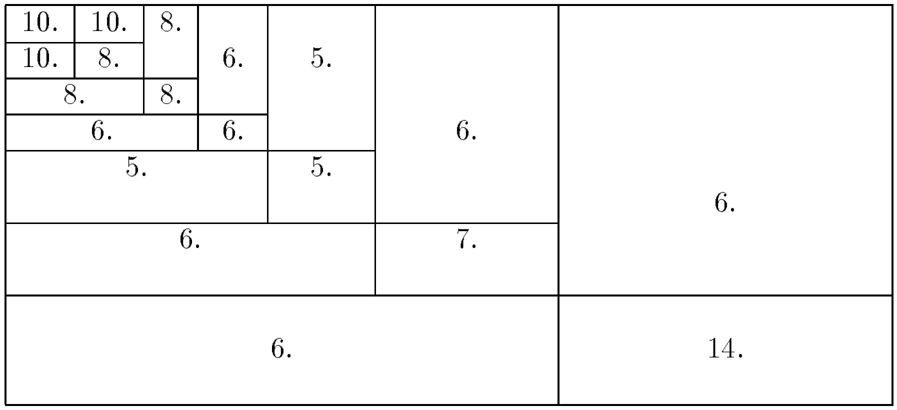

Figure 3: The base perceptual threshold values for a viewing distance of six times the image height.

The base perceptual threshold values were estimated for each subband by comparing a flat gray image of value 127 with a corrupted image obtained by adding uniformly-distributed white noise in the range $[-\delta_k, \delta_k]$ in the subband under consideration. If the two images on the monitor were determined to be perceptually different, the value of the parameter δ_k was reduced until the two original and the corrupted images were visually identical. If the observer determined that the two images were visually identical, the value δ_k was increased until the differences between the two images was barely noticeable. The estimate of the base perceptual detection threshold value for the subband of interest is this value of δ_k. The base perceptual threshold values were measured for a viewing distance of six times the image height. These values are shown in Figure 3 for each band of the two-dimensional wavelet transform.

For the brightness correction, similar experiments were performed for different values of the mean gray level of the images. These experiments were performed on the highest frequency subband. A threshold elevation curve was developed from these experiments for the perceptual threshold function, and this threshold elevation curve is shown in Figure 4.

The threshold elevation due to energy in one subband masking the coefficients in another subband was also found using forced choice experiments. In this case, the original flat image was corrupted with uniformly-distributed white noise with a specified variance level injected in the masking subband. The second image was generated by adding uniformly distributed noise in the range $[-\eta_{kl}, \eta_{kl}]$ to the masked subband. 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 masking subband. Figure 4 shows an example of the threshold elevation due to the masking energy in one subband on the detection threshold of a stimulus in another subband. It was also experimentally determined that the threshold elevation was greatest due to the energy in the subbands in the surrounding levels of the wavelet transform decomposition. Furthermore, it was found that elevation of the perceptual thresholds in a subband due to the energy in other subbands was determined primarily by the subband that causes the most threshold elevation individually.

Combining all the results described above, the model of the perceptual threshold function for the location (x_s, y_s) in the current subband s_c is given by the equation

$$pthresh(x_s, y_s) = base(s_c) bright(x_s, y_s) \max_{s \neq s_c} \{energycorr(x_s, y_s, s)\} \quad (3)$$

where $base(s_c)$ represents the base perceptual threshold for the subband s_c, $bright(x_s, y_s)$ represents

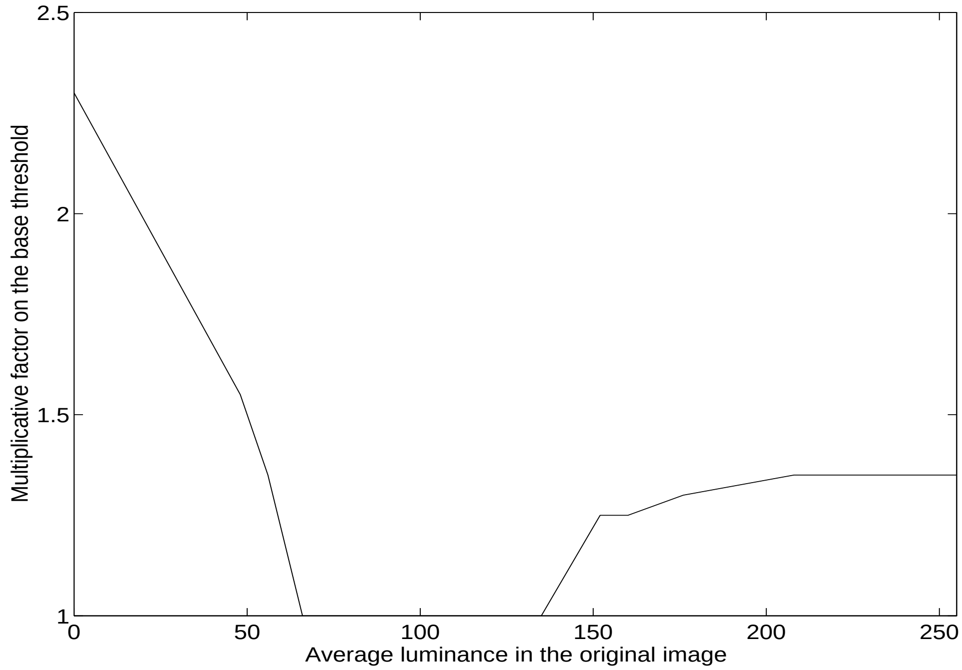

Figure 4: An example of the multiplicative threshold elevation due to the local average luminance in the input image.

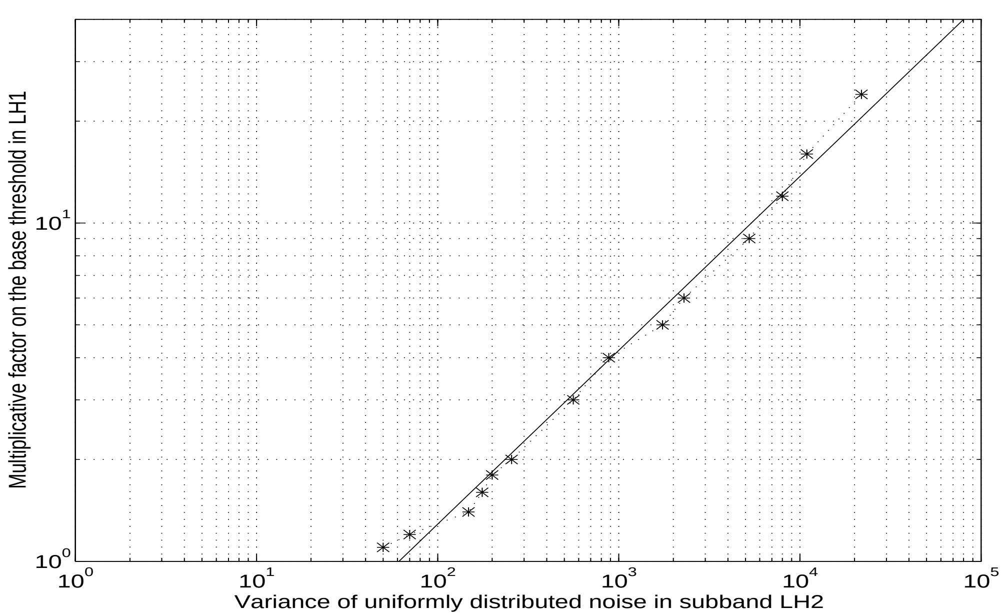

Figure 5: An example of the multiplicative threshold elevation due to energy in one subband masking another subband.

Figure 6: The original image.

the threshold elevation due to the average luminance in the original image corresponding to the (x_s, y_s) location, and $energycorr(x_s, y_s, s)$ represents the threshold elevation due to the energy at the location corresponding to (x_s, y_s) in the subband s. The average luminance is estimated from a 2×2 element block in the original image for the subbands in the first level of the wavelet transform decomposition. This block size doubles for each increase in the level. The local energy is measured as the variance on a 2×2 element block in each subband.

We now present the results of an experiment in compressing monochrome images. All the images used in the experiment contained 512×512 pixels with eight bits per pixel gray scale resolution. The images were all corrected for the monitor nonlinearities before compression. The gamma of the monitor was 2.5. Figure 6 shows the original "Lena" image. Figure 7 shows the compressed image for a viewing distance of six times the image height. The compression ratio is 0.67 bits per pixel. Figure 8 shows the perceptual threshold function for the original image at a distance of six times the image height. The distortions that may still be visible in the printed images from the specified viewing distance are caused primarily by the errors introduced during the printing process.

5. CONCLUSION

This paper presented a perceptually lossless image compression system. The compression system used a modification of the embedded zerotree wavelet algorithm and a model of the perceptual threshold function to guide the compression. The new compression system obtained compressed images in which all quantization errors were constrained below levels suggested by model of the perceptual thershold function at each location of the image. Experimental results demonstrate perceptually lossless compression of monochrome images at bit rates ranging from 0.4 to 1.2 bits per pixel at a viewing distance of six times the image height. These results were obtained using a simple, empirical model of the perceptual threshold function. The authors are presently working on additional refinements to the algorithm, including the development of a perceptual threshold function model that is useful for all viewing distances and image decompositions.

Figure 7: The perceptually lossless compressed image for a viewing distance of six times the image height. The bit rate is 0.67 bits per pixel.

Figure 8: The perceptual threshold function model for the Lena image at a viewing distance of six times the image height.

6. REFERENCES

[1] R. De Valois and L. De Valois, *Spatial Vision*, New York: Oxford University Press, 1988.

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

[3] 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.

[4] H. R. Wilson, D. K. McFarlane, and G. C. Phillips, "Spatial frequency tuning of orientation selective units estimated by oblique masking," *Vision Research*, vol. 23, no. 9, pp. 873-881, Sept. 1983.

[5] T. Stockham Jr. "Image processing in the context of a visual model," *Proc. IEEE*, vol. 60, no. 7, pp. 828-842, July 1972.

[6] B. Ramamurthi and A. Gersho, "Classified vector quantization of images," *IEEE Trans. Communications*, vol. 34, no. 11, pp. 1105-1115, Nov. 1986.

[7] B. Mahesh and W. A. Pearlman, "Image coding on a hexagonal pyramid with noise spectrum shaping," *Proc. SPIE Conf. Visual Communications and Image Proc. IV*, vol. 1199, pp. 764-774, Nov. 1989.

[8] A. N. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proc. IEEE*, vol. 65, no. 4, pp. 536-548, Apr. 1977.

[9] R. Baseri and V. J. Mathews, "Vector quantization of images using visual masking functions," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 365-368, San Fransisco, CA, Mar. 1992.

[10] 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.

[11] 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.

[12] R. J. Safranek and J. D. Johnston, "A perceptually tuned subband image coder with image dependent quantization and post-quantization data compression," *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, pp. 1945-1948, Glasgow, Scotland, 1989.

[13] R. J. Safranek, J. D. Johnston, and R. E. Rosenholtz, "A perceptually tuned subband image coder," *Proc. SPIE Symp. Human Vision and Electronic Imaging: Models, Methods and Applications*, pp. 284-293, Santa Clara, California, Feb. 1990.

[14] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3445-3462, Dec. 1993.

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