

A NEW MODEL OF PERCEPTUAL THRESHOLD FUNCTIONS FOR APPLICATION IN IMAGE COMPRESSION SYSTEMS

K. S. Prashant, V. John Mathews, and Peter J. Hahn
Department of Electrical Engineering
University of Utah
Salt Lake City, Utah 84112
E-mail: mathews@ee.utah.edu

Abstract

This paper discusses the development of a perceptual threshold model for the human visual system. The perceptual threshold functions describe the levels of distortions present at each location in an image that human observers can not detect. Models of perceptual threshold functions are useful in image compression problems because an image compression system that constrains the distortion in the coded images below the levels suggested by the perceptual threshold function performs *perceptually lossless* compression. Our model involves the decomposition of an input image into its Fourier components and spatially localized Gabor elementary functions. Data from psychophysical masking experiments are then used to calculate the perceptual detection threshold for each Gabor transform coefficient in the presence of sinusoidal masks. The result of one experiment involving distorting an image using additive noise of magnitudes as suggested by the threshold model is also included in this paper.

1 Introduction

In many applications involving lossy image compression, the human viewer is often the ultimate judge of the quality of the compressed images. Since subjective assessment of image quality depends largely on the the properties of the human eye, many researchers have attempted to develop image compression systems that utilize an understanding of the behavior of the eye [1]-[5].

One method for designing perceptually-tuned image compression systems involves predicting the amount of distortion that can be introduced into images without being perceived or detected by the human viewer. We will refer to the maximum amount of distortion that can be present at various locations of an image as the perceptual threshold function for the image. Note that the maximum distortion levels will vary with the viewing distance. An image compression system that constrains all the distortions below the *perceptual threshold function* for a specific viewing distance will be *perceptually lossless* for that viewing distance. Examples of data compression systems utilizing this approach include [3], [4], [5].

In this paper, we present a method for calculating the visual thresholds at which the distortions are barely visible in a given image from a specified viewing distance. Our model appears to predict larger threshold values than previously available models and therefore, the image compression systems utilizing it can produce perceptually lossless image compression at lower bit rates than previously possible.

2 An Overview of Results in Vision Research Relevant to Perceptual Threshold Modeling

Over the past three decades, a large number of efforts have been devoted in psychophysics and physiology to analyze the response of the visual system. This section presents a summary of experimental results of human vision research.

The receptor cells in the visual system exhibit nonlinear responses to stimuli. This nonlinearity has often been approximated using logarithmic input-output relationships [8]. Assuming a logarithmic nonlinearity to be the first component of the human visual system, Campbell and Green [9] and other researchers attempted to model the eye as a static nonlinearity followed by a linear filter. Stockham's homomorphic model [1] belongs to the above class of models of the human visual system. Although the above model of the eye is overly simplified, it shows qualitatively the sensitivity of the human system to stimuli of different frequencies.

Several researchers [10] made the suggestion that the visual system might contain groups of independent, linear channels, each of which was more narrowly tuned to smaller ranges of spatial frequencies than the overall contrast sensitivity function (CSF). The CSF would then reflect not the sensitivity of a single visual channel, but some envelope of the sensitivities of all these multiple channels. The psychophysical basis for such claims were experiments based on *adaptation*, [10] and *masking* [11]. If a stimulus is shown to an observer for a long time, the visual sensitivity for the same kind of stimuli decreases. This behavior is called the adaptation process. In other words, if the subject adapts to a spatial frequency f_0 , the post-adaptation CSF will show a bandlimited attenuation centered at the adaptation frequency. This effect was found to be limited to frequencies that were approximately an octave (a factor of 2) on either side of the adaptation frequency [10].

Masking studies measure the detectability of a particular pattern alone and then in the presence of another masking pattern. Experiments were conducted to measure the values of raised visual thresholds due to masking effects [11]. The results of the experiments showed that the presence of a sine wave of frequency f_0 resulted in the elevation of the detection threshold of test frequencies in the neighborhood of f_0 . Sachs, Nachmias, and Robson [12] carried out experiments with compound gratings consisting of two frequency components. Whenever the second component differed in frequency from the first component by more than an octave, the data were consistent with the two frequencies being detected independently.

Findings from physiological studies on the cells in the visual cortex also support

the basic multiple channel notion. Using experiments on cats' visual cortex, Hubel and Wiesel [13] first discovered a class of cells, whose response depends upon the frequency and orientation of the visual stimuli. DeValois, *et. al.* [14] measured the spatial frequency contrast sensitivity of cells at two different positions in the primate striate cortex of a macaque. It was observed that many striate cells have quite narrow spatial bandwidths and that the distribution of the peak frequency covers a wide range of frequencies.

Masking experiments [15] testing the effect of the presence of a pattern of one orientation on a pattern of a different orientation have indicated the presence of separate orientation channels. It was observed that a masking pattern sufficiently different in orientation from the test pattern did not interfere with its detection, although patterns from nearby orientations did. The orientation bandwidth was found to be close to 30° . Subsequent adaptation experiments by Blakemore and Nachmias [16] and physiological experiments [13], [18], have also confirmed the presence of orientation selective cells in the eye.

Psychophysical masking experiments described so far used vertical cosine gratings as both mask and test patterns. These gratings cover a large portion of the visual field, and they consider the detection of the overall pattern. It is known that the receptive fields of striate cells are spatially localized. Stimuli used to measure characteristics of such cells must be tuned to such sharply tuned receptive fields. Signals that are spatially localized serve this purpose. In addition, the limited spatial extent of such signals minimizes any effects due to the spatial inhomogeneity of the visual system. Wilson [17] used higher derivatives of Gaussian functions and Difference of Gaussian (DOG) functions as test stimuli for his psychophysical experiments. Threshold elevation data for DOG stimuli masked by sinusoidal patterns have been published in [17]. From an analysis of thresholds for spatially localized aperiodic patterns, it was shown that we may be able to model the behavior of the eye with as few as four to six different tuned mechanisms at each location in the visual system.

3 The Perceptual Threshold Model

The fundamental idea behind the model discussed in this section is that we may be able to model the behavior of the eye using a finite number of channels. Each of these channels is selectively tuned to a particular frequency and orientation. Even though inadequate evidence exists at this time to suggest that there are only a small number of distinct mechanisms which determine the properties of the eye, the results of several physiological and psychophysical experiments indicate that the behavior of the early portions of the eye may be adequately approximated using as few as six distinct mechanisms. This is very important from the point of view of building threshold models for image compression, since ease of computation is of importance in such applications.

The development of the perceptual threshold model relies heavily on Wilson' experiments [17], [19]. These experiments provide threshold elevation data for spatially

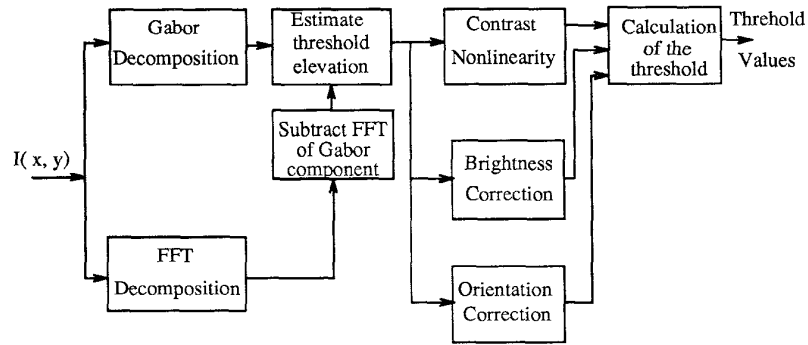


Figure 1: The perceptual threshold model

localized functions in the presence of sinusoidal masks. Our model utilizes this data, but incorporates several additional features based on information that is available in the literature as well as that gained by our own experiments. The block diagram of the model is shown in Figure 1. We now describe the various components of the model.

Gabor and Fourier Decompositions: Even though Wilson's experiments employed difference of Gaussian stimuli, we have used Gabor elementary functions (GEFs) [20] as a basic building block for our model. There are two main reasons for this: (a) Several efficient algorithms for decomposing signals into Gabor elementary functions are available [25], and (b) there is considerable physiological evidence in the literature [22] that suggests that two-dimensional Gabor elementary functions (GEFs) are fundamental to visual processing in several mammalian species. Tests of the responses of cells to gratings of a wide range of spatial frequencies and orientations [24] show that the actual striate cells behave much as would be predicted if the receptive fields approximated a two-dimensional Gabor function. Gabor fits to psychophysical data obtained by Wilson [17] are quite close to corresponding DOG fits [23].

In our model, the image is first decomposed into a set of Gabor transform coefficients. These coefficients are grouped together according to their orientations. Recall that Wilson's experiments measured the threshold elevation for detecting spatially localized stimuli when they were masked by sinusoidal signals. In order to utilize the results of Wilson's experiments, we also need to decompose the input images into sinusoidal components. This is accomplished by computing the discrete Fourier transform of a block of pixels of the image centered around the mid-point of each Gabor elementary function employed in the Gabor decomposition of the image. The size of the block varies with viewing distance, and is chosen such that it matches the average size of the receptive field. The size of a receptive field of

cells with an average bandwidth of about 1 octave is close to 0.4° [7]. The block size is chosen such that it matches the average size of the receptive field. In our work, we assumed an aspect ratio of 1 while determining the block sizes. Square blocks of size 8, 16, and 32 pixels per side correspond reasonably closely to the size of the receptive field for viewing distances of 3, 6, and 12 times the image height, respectively.

Estimation of the Threshold Elevation: Once the Fourier coefficients have been computed, the effect of the sinusoidal components on the detection of the GEF's can be calculated from measured data that are available in the literature. The threshold elevation of a Gabor function in the presence of another sinusoidal function can be estimated directly from the data in [17]. The threshold elevation documented in [17] assumes a mask contrast of 40% and a mask orientation of 14.5° . The threshold elevation data of [17] has to be corrected for differences in contrast and orientation of the sinusoidal components of the image as well as differences in the local brightnesses of the image from the conditions of the original experiments. While calculating the threshold elevation for a particular Gabor component, the Fourier transform of that component is subtracted from the Fourier transform of the corresponding block. This ensures that the masking sinusoidal components do not have any contribution from the stimulus whose threshold elevation is to be estimated.

Correction for Contrast Nonlinearity: It is well known that the threshold elevation is a nonlinear function of the mask contrast. Experiments using varying mask contrasts were conducted by Wilson [17], and it was found that the mask contrast and the threshold elevation had a power-law relationship. The nonlinearity can be modeled as

$$\Delta C = K \cdot C_m^\epsilon. \quad (1)$$

In the above equation, ϵ is the power law exponent, and K is a sensitivity parameter that needs to be empirically determined. This type of relationship has also been suggested by Legge [26] who suggested a value of 0.60 for ϵ . Our model uses this value.

Correction for Orientation of the Masking Sinusoid: Our model assumes that the threshold elevation decays linearly from the maximum value at 0° orientation to zero elevation at 30° orientation of the masking sinusoid. The orientations of the frequencies of the masking components are computed relative to the orientation of the center frequency of the Gabor elementary function under consideration.

Brightness Correction: The threshold elevation is also corrected for dependence on local brightness. The value of the local brightness is estimated from the DC value of the corresponding block. The brightness correction factor is similar to the one used by Safranek and Johnston [4] and is defined as,

$$\text{BrightCorr} = \begin{cases} \exp[k_1 \times (mid - DC)]; & \text{if } DC \leq mid \\ \exp[k_2 \times (DC - mid)]; & \text{if } DC > mid, \end{cases} \quad (2)$$

where, *mid* denotes the middle value of the intensities. The values of k_1 and k_2 were empirically selected as -0.03 and -0.004, respectively, in our model. The value of the threshold elevation is multiplied with the brightness correction factor to give the final value of the threshold elevation of the Gabor coefficient.

Calculation of the Masked Threshold Contrast: The masked threshold contrast C_m is defined as the contrast at which a stimulus can be just noticed in the presence of masking. At this stage of the model, the value of the threshold elevation of the Gabor functions constituting the image are available for each sinusoidal component masking it. The overall effect of all the masking signals is evaluated in the model as a weighted sum of the threshold elevations due to each component. We have found through our experiments that accounting for as few as five of the most dominant components will provide a reasonably accurate model of the overall threshold elevation. Let ΔC_{th} denote the threshold elevation for a Gabor elementary function under consideration due to all the sinusoidal components that mask it. This quantity can now be expressed as

$$\Delta C_{th} = \frac{C_m}{C_u}, \quad (3)$$

where C_m is the value of the masked threshold contrast of the Gabor elementary function and C_u is the value of the unmasked threshold contrast and can be obtained from any standard contrast sensitivity function curve [8]. Knowing the value of the threshold elevation ΔC_{th} from the earlier steps, the value of C_m can be calculated as

$$C_m = C_u \cdot \Delta C. \quad (4)$$

This value is an indicator of the perceptual threshold contrast of the Gabor elementary function in the presence of masking. A value below the perceptual threshold is called subthreshold and above the threshold is called suprathreshold. GEFs whose contrasts are subthreshold cannot be seen.

Very little data is available on masking effects of suprathreshold stimuli. Weber's law [8] and experiments by Legge [26] suggest that the unmasked just noticeable difference (JND) contrast at suprathreshold contrasts ΔC_u^{st} is related to the background contrast C through an exponential relationship. Therefore, we model the JND contrast as

$$\Delta C_u^{st} = kC^N, \quad (5)$$

where k is a constant and N is a constant less than 1. The value of N is chosen to be 0.8 in our model. This value is purely an empirical selection. In order to incorporate the effects of masking at suprathreshold contrasts, this value of ΔC_u^{st} is now used as the unmasked just noticeable difference. The masked JND contrast ΔC_m^{st} is calculated using the threshold elevation data ΔC_{th} computed earlier as

$$\Delta C_m^{st} = \Delta C_u^{st} \cdot \Delta C_{th}. \quad (6)$$

Empirical Calibration of the Model: The model contains certain parameters whose values are determined empirically. The calibration of the model was done using a set of 8 monochrome images of size 512 x 512 pixels and 256 amplitude levels per pixel. The calibration was an iterative process. The parameters to be determined were set to certain initial values. At each iteration, the perceptual threshold function was calculated for one image using the current set of parameter values for a specific viewing distance. The original image was then corrupted with an independent, identically distributed additive noise sequence. The amplitude of the noise at each location of the image was determined by the threshold value at that location as predicted by our model. The two images were then simultaneously displayed on a Sun monitor after correcting for the nonlinearities in the monitor characteristics. The display had a gamma factor of 2.5, and was set to accommodate a dynamic range of approximately 100. The ambient light levels were such that only 0.2 ft-lamberts of light was reflected of a blank monitor. This value is just about 3% of the maximum intensity value that could be displayed on the monitor. The observers, who were mostly untrained, but aware of the objectives of the experiment, were asked to view the images from a specified distance and determine if the two images were identical or different. Whenever the consensus opinion of the observers was that the images were identical, one of the parameter values was changed in such a way that the threshold values predicted by the model would increase. If the images were determined to be different from each other, the parameter value was changed so that the threshold values would decrease. This process was repeated till any change in the parameter value caused a change in the perceived quality of the images. The process was repeated for all the parameters and all the images several times till convergence was obtained for all the parameters. At this point, the differences between the two sets of images – the original and the corrupted ones – would be barely visible to the observers.

4 Experimental Results

We conducted a large number of experiments using human observers to determine the usefulness of the perceptual threshold model we developed and also to compare our model with previously available models [3], [4]. In each case, we found that our model was superior to previously available models because it either allowed for larger unperceivable distortions than other models, or when the average threshold levels were comparable for two models, the threshold levels predicted by other models produced distinctly noticeable distortions. As an example, when the commonly used test image “Lena” (monochrome image of size 512 x 512 pixels and 256 amplitude levels) was distorted according to the thresholds suggested by our model, the untrained observers could not distinguish the original from the distorted image from the specified viewing distances. However, the mean-square distortions in the noisy images were as much as 18, 38 and 67, respectively for viewing distances of 3, 6, and 12 times the image height. These values correspond to peak-to-peak signal-to-noise ratios of 39, 32 and 30 dB, respectively. Figures 2 and 3 display the original image and the corrupted image corresponding to a viewing distance of 6 times the image



Figure 2: The original image



Figure 3: The image corrupted with noise at levels predicted by the perceptual threshold model

height, respectively. We can see that the two images appear reasonably close to each other from the appropriate viewing distance. The slight differences that are visible in the photographs are partly due to the inaccuracies in the compensation for nonlinearities in the film recording and printing process.

5 Concluding Remarks

We have developed a model for the perceptual threshold function of the human visual system. We have incorporated a large number of currently known properties of the eye in our model, and therefore, we believe that the model we developed is more accurate than previously available ones. Experimental results are in agreement with this assessment. Additional work is underway to further refine this model and to use it in developing perceptually-tuned image compression systems.

6 Acknowledgment

The work described in this paper was supported in parts by the National Science Foundation under grand MIP-9016331, by NASA under grant NAG 5-2200, and by an IBM Departmental Grant.

References

- [1] T. Stockham Jr, "Image processing in the context of a visual model," *Proc. of the IEEE*, Vol. 60, No-7, July 1972, pp. 828-842.
- [2] J. N. Bradley, T. G. Stockham and V. J. Mathews, "An optimal design procedure for subband vector quantization," to appear in the *IEEE Trans. Communications*, February 1995.
- [3] A.N. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proc. IEEE*, Vol. 65, April 1977, pp. 536-548.
- [4] R.J. Safranek and J.D. Johnston, "A perceptually tuned subband image coder with image dependent quantization and post-quantization data compression," *Proc. ICASSP*, Glasgow, Scotland, May 1989, pp. 1945-1948.
- [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, October 1993, pp. 1385-1424.
- [6] J. O. Limb, "Distortion criteria of the human viewer," *IEEE Trans. Systems, Man and Cybernetics*, Vol. SMC-9, December 1979.
- [7] R. De Valois and K. De Valois, "Spatial Vision," *Oxford Science Publications*, 1988, pp. 241-242.
- [8] T. Cornsweet, "Visual Perception," Academic Press, New York, 1970.
- [9] F. Campbell and D. Green, "Optical and retinal factors affecting vision resolution," *J. Physiology*, Vol. 181, 1965, pp. 576-593.

- [10] F.W. Campbell and J.G. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiology*, London, Vol. 197, 1968, pp. 551-566.
- [11] C. Stromeyer and B. Julesz, "Spatial frequency masking in vision: Critical bands and spread of masking," *J. Opt. Soc. Am.*, Vol. 62, October 1972, pp. 1221-1231.
- [12] M. Sachs, J. Nachmias, and J. Robson, "Spatial frequency channels in human vision," *J. Opt. Soc. Am.*, Vol. 61, pp. 1176-1186.
- [13] D. Hubel and T. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cats visual cortex," *J. Physiology*, London, Vol. 60, 1962, pp. 551-566.
- [14] R.L. De Valois, D.G. Albrecht, and L.G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vision Research*, Vol. 22, 1982, pp. 545-559.
- [15] F.W. Campbell and J. Kulikowski, "Orientation selectivity of the human vision system," *J. Physiology*, London, Vol. 197, 1966, pp. 431-441.
- [16] C. Blakemore and J. Nachmias, "The orientation specificity of two visual after effects," *J. Physiology*, London, Vol. 213, pp. 157-174.
- [17] H. Wilson, D. McFarlane, and G. Phillips, "Spatial frequency tuning of orientation selective units estimated by oblique masking," *Vision Research*, Vol. 23, pp. 873-882.
- [18] R. De Valois, W. Yund and N. Helper, "The orientation and direction selectivity of cells in macaque visual cortex," *Vision Research*, Vol. 22, 1982, pp. 531-544.
- [19] G. Phillips and H. Wilson, "Orientation bandwidths of spatial mechanisms measured by masking," *J. Opt. Soc. Am.*, Vol. 1, 1984, pp. 226-232.
- [20] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng.*, Vol. 93, 1946, pp. 429-459.
- [21] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Am.*, Vol. 70, 1980, pp. 1297-1300.
- [22] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Am.*, Vol. 2, 1985, pp. 1160-1169.
- [23] D. Regan "Spatial Vision," *Macmillan Press, Boca Raton, FL*, Vol. 10, 1991.
- [24] M. Webster and R. DeValois "Relationship between spatial-frequency and orientation tuning of striate cortex cells," *J. Opt. Soc. Am.*, Vol. 2, 1985, pp. 1124-1132.
- [25] M. Bastiaans, "Gabor's expansion of a signal into Gaussian elementary signals," *Proc of the IEEE*, Vol. 68, No. 4, April 1980, pp. 538-539.
- [26] G. Legge, "A power law for contrast discrimination," *Vision Research*, Vol. 21, March 1982, 1980, pp. 457-467.