THE VISIBILITY OF DCT QUANTIZATION NOISE

H. A. Peterson
IBM T. J. Watson Research Center, Yorktown Heights, NY 10598
A. J. Ahumada, Jr., and A. B. Watson
NASA Ames Research Center, Moffett Field, CA 94035-1000

ABSTRACT

Many standard image compression techniques apply the Discrete Cosine Transform (DCT) to the image, then quantize the resulting transform coefficients. For optimal compression, the DCT coefficients should be quantized as coarsely as possible, while allowing minimal visible distortion in the decompressed image. Quantization of a DCT coefficient induces a noise pattern over the image consisting of random amplitude replications of the corresponding basis function. Here we measure the detectability of such noise patterns for three different size test patterns. Implications of the experimental results are discussed. These measurements will facilitate the design of visually optimized DCT coefficient quantization schemes.

1. INTRODUCTION

1.1. DCT-Based Compression

The Discrete Cosine Transform (DCT) has become a standard image compression method.[1,2,3] With this technique the image is divided into two-dimensional (typically 8×8) blocks of pixels. Each block is transformed into 64 DCT coefficients. The DCT coefficients are weights associated with the DCT basis functions, from which the block of image pixels can be reconstructed. The (m,n)'th DCT basis function, $B_{m,n}(j,k)$, of size $N \times N$ and with amplitude a can be written:

$$B_{m,n}(j,k) = a \cos\left(\frac{\pi m}{2N}[2j+1]\right)\cos\left(\frac{\pi n}{2N}[2k+1]\right), \quad (1)$$

$$j,k,m,n = 0, \cdots, N-1 .$$

The DCT coefficients are quantized and encoded to form the compressed version of the image.

A substantial part of the compression gained using a DCT-based scheme derives from quantization of the DCT coefficients. DCT coefficient quantization induces distortion in the decompressed image. To achieve optimal compression, the DCT coefficients should be quantized as coarsely as possible, while allowing minimal visible distortion in the decompressed image. Variations in human visual system contrast sensitivity make different degrees of quantization appropriate for different coefficients.[4−8] Psychophysical measurements of the detectability of DCT quantization noise patterns facilitate the design of visually optimized DCT coefficient quantization schemes.

1.2. DCT Coefficient Quantization Noise

The effect of uniformly quantizing a variable x using a quantization interval of width $2q$ is to generate an error e that is approximately uniformly distributed with range $[-q, q]$. For any unimodal continuous distribution on the original variable x, this approximation improves as q is reduced. In addition, for correlated, appropriately continuously jointly distributed random variables x_1 and x_2, the errors e_1 and e_2 become independent of each other and of x_1 and x_2 as q is reduced. Since images can be fairly well modeled as random processes, it follows that for small q, the error image resulting from quantization of a single DCT coefficient should be fairly well modeled as consisting of replications of that DCT basis function, having random amplitudes uniformly distributed over the interval $[-q, q]$. Figure 1 shows three examples of such a noise model (on a gray background), for the $N = 8$, (m,n) equal to (0, 0), (2, 0), and (2, 4) DCT basis functions.

Measurements are reported here of visibility thresholds for DCT quantization noise patterns like those in Figure 1. We compare thresholds measured for three different size test patterns: 1×1 ($P = 1$), 3×3 ($P = 3$), and 6×6 ($P = 6$) arrays of replicated basis functions.

2. EXPERIMENTS

2.1. Apparatus

The experiments used a 19 inch IBM 6091 color monitor, driven at 60 Hz (non-interlaced) by a 24 bit (8 bits per component) display adapter. The color look-up tables provided a linear relationship between digital image value and measured output luminance. The monitor has 100 pixels per inch in both horizontal and vertical directions, and maximum luminance of 66.4 cd/m^2. Test subjects viewed the monitor from 104 cms. (approximately 41 inches), yielding a display resolution of 72 pixels/degree of visual angle. The test room was dark.

2.2. Test Stimuli

Test stimuli were presented on a midrange gray background (with luminance 33.2 cd/m^2) that filled most of the screen (an 1170×910 pixel window on a 1280×1024 pixel monitor). This background can be thought of as an image with all DCT coefficients equal to zero, except for the DC (m and $n = 0$) coefficient.

Figure 1. Three examples of our DCT coefficient quantization noise model are shown, on a gray background for $N = 8$ and $P = 3$. The left, center, and right patterns are for (m, n) values of (0, 0), (2, 0), and (2, 4), respectively. These also illustrate the type of test stimuli used in our experiments.

A test stimulus was constructed in the following manner. First an 8×8 pixel DCT basis function was generated according to Equation (1) with amplitude equal to q, and pixel replication was used to increase its size to 16×16 pixels. Pixel-replication serves to reduce effects due to bandwidth limitations in the monitor response. For $P = 1$, this basis function formed the test stimulus. For values of $P > 1$, a test stimulus was constructed as a $P \times P$ array of replications of the basis function, with each replication assigned a random amplitude uniformly distributed on $[-q, q]$. A new sample of random values for the basis function amplitudes was drawn for each trial in the experiment.

The $PN \times PN$ test stimulus image was added to the flat gray background to create the test pattern. Three examples of such test patterns for $P = 3$ are shown in Figure 1.

The viewing distance mentioned above is twice the "typical" viewing distance of 1-2 screen heights. The effects of doubling the viewing distance and doubling the effective pixel size cancel each other, having no net effect on the spatial frequency at the eye of the test stimuli. In this configuration, a basis function spans 0.22 degrees of visual angle.

2.3. Experimental Procedure

We describe the experimental procedure for a particular value of P. This procedure was repeated for $P = 1, 3,$ and 6. To reduce the duration of the experiment, data were collected for only 30 of the 64 basis functions (see Figure 2). The forced choice procedure used two temporal intervals and feedback. A test stimulus appeared during one of the temporal intervals, chosen at random. The test stimulus was presented in the center of the display, and between trials four small tic marks indicated the locations of the corners of the stimulus. Following presentation of the two temporal intervals in a trial, the subject indicated which interval contained the stimulus by pressing a key on a keyboard. If the subject's answer was incorrect, a "beep" was sounded. A correct answer resulted in no beep. For each of the 30 stimuli, 64 trials were performed, and all trials for one stimulus were completed before beginning the next. The 30 different stimuli were presented in a series of six sessions, with five stimuli in each session.

X	X	X	X		X		X
X	X	X	X		X		X
X	X	X					
X	X		X		X		X
X	X		X		X		X
X	X		X		X		X

Figure 2. The 30 basis functions for which data were collected are indicated by X's. The m and $n = 0$ basis function is in the upper left corner.

The size q of the amplitude range (or simply the amplitude for $P = 1$) of the test stimulus was varied from trial to trial in order to establish a visibility threshold for that basis function. The QUEST[9] adaptive psychometric method determined the value of q for each trial, and the q values were spaced logarithmically. The stimulus duration was 31 frames (approximately 517 milliseconds). To reduce the effects of discontinuous temporal variation in the stimulus, the amplitude of the stimulus had a gaussian shaped envelope in time. The amplitude scaling factor grew from $e^{-\pi}$ to 1 for the first 16 frames, and fell back to $e^{-\pi}$ over the last 15 frames.

2.4. Threshold Estimation

A separate data set was collected for each basis function and value of P. Each set consisted of percent correct versus log q data. A Weibull function (maximum = 0.99, minimum = 0.5, slope = 4.0) was fit to each data set and a threshold was estimated as the value of q for which the Weibull function was equal to 0.82.[10]

3. RESULTS

Visibility thresholds estimated as described above for $P = 1$, 3, and 6 are plotted in Figure 3. Only thresholds for DC (m and $n = 0$) and purely vertical or horizontal basis functions (m or $n = 0$) are shown. The DC thresholds are plotted at the far left of the graph. The purely vertical or horizontal thresholds are plotted as a function of spatial frequency,

$$f_{0,n} = f_{n,0} = \frac{n}{2NW} , \quad (2)$$

where $NW = 0.22$ degrees, and $N = 8$.

In order to estimate the experimental effect of test pattern size, for each value of P we average the log thresholds over all frequencies. These averages are $t_1 = 1.218$, $t_3 = 1.160$, and $t_6 = 1.085$, for $P = 1$, 3, and 6, respectively. We refer to the difference in an average experimental log threshold measured for two different values of P, say $P = r$ and $P = s$, as $d_{\mathrm{E}}(r,s) = t_r - t_s$. For our data, $d_{\mathrm{E}}(1,3) = 0.058$, and $d_{\mathrm{E}}(3,6) = 0.075$.

If the spatial summation effect in our data follows probability summation[10,11] we would expect t_3, and t_6 to be related in proportion to the fourth root of the areas of the test patterns. Using this rule, and assuming the effect is independent of spatial frequency, the theoretical difference in a log threshold measured for two different values of P is

$$d_{\mathrm{T}}(r,s) = \log\left[\left(\frac{r^2}{s^2}\right)^{1/4}\right] = \frac{1}{2}\log\left(\frac{r}{s}\right) . \quad (3)$$

Using this equation $d_{\mathrm{T}}(1,3) = 0.239$ and $d_{\mathrm{T}}(3,6) = 0.151$.

There is a further effect on the thresholds measured for $P = 1$, due to the non-random nature of this test pattern's amplitude. For a test pattern with amplitude q, for the $P > 1$ case, the average amplitude of the P^2 basis functions comprising the test pattern is $q/2$. On the other hand, for a test pattern with amplitude q and $P = 1$, the amplitude of the single basis function comprising the test pattern is simply q. Since the average basis function amplitude for $P > 1$ in one half the basis function amplitude for $P = 1$, we expect t_1 to be reduced by log(2) = 0.3, in relation to t_3. Combining this effect with the spatial summation effect, we obtain the prediction $d_{\mathrm{T}}'(1,3) = 0.239 - 0.3 = -0.061$.

Comparing the experimental results to the theoretical predictions we see that $d_{\mathrm{E}}(1,3)$ is larger than expected theoretically ($0.058 > -0.061$), and $d_{\mathrm{E}}(3,6)$ is smaller than expected theoretically ($0.075 < 0.151$). The summation effect we have measured appears stronger than probability summation from $P = 1$ to $P = 3$, and weaker than probability summation from $P = 3$ to $P = 6$. The stronger summation supports models with strong local summation,[11,12,13] although multiple channel models can show such effects.[14] To predict summation weaker than probability summation, an appeal to foveal non-homogeneity may be needed.[15]

ACKNOWLEDGEMENTS

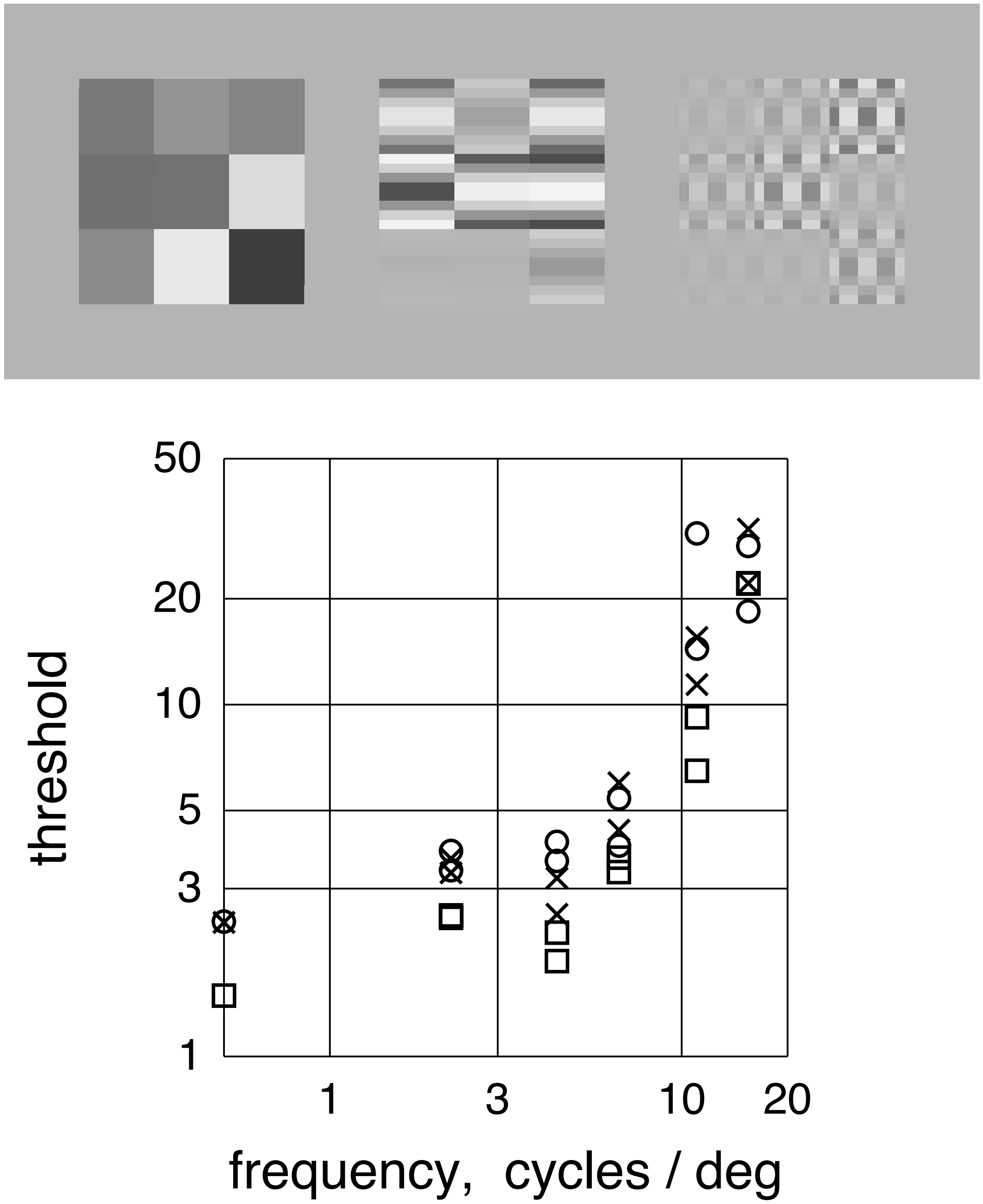
This work was supported in part by the IBM Independent Research and Development Program and by NASA RTOP Nos. 506-59-65 and 505-64-53.

REFERENCES

1. G. Wallace, "The JPEG still picture compression standard", *Communications of the ACM*, vol. 34, no. 4, pp. 30-44, 1991.
2. W. B. Pennebaker, J. L. Mitchell, *JPEG Still Image Data Compression Standard*, van Nostrand Reinhold, New York, 1993.
3. D. LeGall, "MPEG: A video compression standard for multimedia applications", *Communications of the ACM*, vol. 34, no. 4, pp. 46-58, 1991.
4. H. Lohscheller, "A subjectively adapted image communication system", *IEEE Transactions on Communication*, vol. COM-32, pp. 1316-1322, 1984.
5. N. Nill, "A visual model weighted cosine transform for image," *IEEE Transactions on Communication*, vol. COM-33, pp. 551-557, 1985.
6. H. A. Peterson, "DCT basis function visibility thresholds in *RGB* space," in J. Morreale, ed., *1992 SID International Symposium Digest of Technical Papers*, Society for Information Display, Playa del Rey, CA, pp. 677-680, 1992.
7. H. A. Peterson, A. J. Ahumada, Jr., A. B. Watson, "An improved detection model for DCT coefficient quantization", *Human Vision, Visual Processing, and Digital Display IV*, B. E. Rogowitz, J. P. Allebach,

eds., (SPIE, Bellingham, WA, 1993).
8. A. B. Watson, "DCT quantization matrices visually optimized for individual images," *Human Vision, Visual Processing, and Digital Display IV*, B. E. Rogowitz, J. P. Allebach, eds., (SPIE, Bellingham, WA, 1993).
9. A. B. Watson, D. G. Pelli, "QUEST: A Bayesian adaptive psychometric method," *Perception and Psychophysics*, vol. 33(2), pp. 113-120, 1983.
10. A. B. Watson, "Probability summation over time", *Vision Research*, vol. 19, pp. 515-522, 1979.
11. J. G. Robson, N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," *Vision Research*, vol. 21, pp. 409-418, 1981.
12. B. Girod, "The Information Theoretical Significance of Spatial and Temporal Masking," *Human Vision, Visual Processing, and Digital Display IV*, B. E. Rogowitz, ed., Proc. SPIE 1077, pp.178-187, (1989).
13. J. O. Limb, "Distortion Criteria of the Human Viewer," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-9, pp. 778-793, 1979.
14. C. Zetzsche, G. Hauske, "Multiple channel model for the prediction of subjective image quality," *Human Vision, Visual Processing, and Digital Display IV*, B. E. Rogowitz, ed., Proc. SPIE 1077, pp.209-216, (1989).
15. C. A. Curcio, K. R. Sloan, O. Packer, A. E. Hendrickson, R. E. Kalina, "Distribution of cones in human and monkey retina: individual variability and radial asymmetry," *Science*, vol. 236, pp. 579-582, 1987.

Figure 3. Estimated visibility thresholds for three test pattern sizes, $P = 1$ (round symbols), $P = 3$ ("X" symbols), and $P = 6$ (square symbols), are plotted as a function of spatial frequency. DC (m and $n = 0$) thresholds are plotted at the far left.


50
20
10
5
3
1
threshold
1
3
10
20
frequency, cycles / deg