An Improved Detection Model for DCT Coefficient Quantization

Heidi A. Peterson
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598

Albert J. Ahumada, Jr. and Andrew B. Watson
NASA Ames Research Center, Human Interface Research Branch
Moffett Field, California 94035-1000

ABSTRACT

A detection model is developed to predict visibility thresholds for discrete cosine transform coefficient quantization error, based on the luminance and chrominance of the error. The model is an extension of a previously proposed luminance-based model, and is based on new experimental data. In addition to the luminance-only predictions of the previous model, the new model predicts the detectability of quantization error in color space directions in which chrominance error plays a major role. This more complete model allows DCT coefficient quantization matrices to be designed for display conditions other than those of the experimental measurements: other display luminances, other veiling luminances, other spatial frequencies (different pixel sizes, viewing distances, and aspect ratios), and other color directions.

1. INTRODUCTION

1.1 Discrete cosine transform-based image compression

The discrete cosine transform (DCT) has become a standard method of image compression.[1,2,3] Typically the image is divided into 8×8-pixel blocks, which are each transformed into 64 transform coefficients. The DCT transform coefficients $I_{m,n}$, of an $N \times N$ block of image pixels $i_{j,k}$, are given by

$$I_{m,n} = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} i_{j,k}\, c_{j,m}\, c_{k,n}, \quad m,n = 0, \ldots, N-1, \tag{1a}$$

where

$$c_{j,m} = \alpha_m \cos\left(\frac{\pi m}{2N}[2j+1]\right), \quad \text{and} \quad \alpha_m = \begin{cases} \sqrt{1/N}, & m = 0 \\ \sqrt{2/N}, & m > 0 \end{cases}. \tag{1b}$$

The block of image pixels is reconstructed by the inverse transform:

$$i_{j,k} = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} I_{m,n}\, c_{j,m}\, c_{k,n}, \quad j,k = 1, \ldots, N-1, \tag{2}$$

which for this normalization is the same as the forward transform. Quantization of the DCT coefficients achieves image compression, but also causes distortion in the decompressed image. Specifically, quantization of coefficient $I_{m,n}$ induces an error image which is simply the associated basis function, with amplitude equal to the coefficient quantization error (neglecting the DCT normalization).

1.2 The Quantization Matrix

The JPEG compression standard[1,2] requires that uniform quantizers be used for all the DCT coefficients. The quantizer step size used for each coefficient is determined by the user. A matrix is used to specify the quantization of the DCT coefficients, where the m,nth entry, $Q_{m,n}$, in the matrix gives the quantizer step size for coefficient $I_{m,n}$. Two example quantization matrices have been included in the JPEG standard. These

matrices are given in Tables K.1 and K.2 of reference(2) and in Table 5 of reference(4). One of these matrices is commonly used for graylevel images, and for the luminance component image of color images; the other matrix is used for chrominance images. These matrices were designed for a particular compression/viewing scenario, and it is not clear how they should be changed when used under different viewing conditions, or especially for compression in a different color space. In this paper we propose a quantization matrix design technique that can be applied under a wide variety of conditions: different display luminances, veiling luminances, spatial frequencies, and color spaces.

2. DETECTION MODELS

2.1 Luminance-only Detection Model

Peterson, Peng, Morgan, and Pennebaker[4] developed quantization matrices for compressing images in the RGB color space (a different matrix is used for each of the R, G, and B component images). The matrices were derived from measured detection thresholds for small patches of replicated DCT basis functions, produced on a monitor using an individual R, G, or B gun on a black background. With minor adjustments, the measured thresholds were converted to quantization matrices which performed well in informal tests.

Ahumada and Peterson[5] proposed that the threshold measurements of Peterson *et al.*[4] could be predicted by a luminance-only detection model. The theoretical basis of their model is the assumption that the detectability of distortion in the decompressed RGB image can be predicted from the luminance contrast of the error image caused in a color component image by quantization of an individual DCT coefficient for a single block. That is, if the quantization error images associated with all the quantized DCT coefficients in all image blocks in all three color component images have amplitudes below their respective visibility thresholds, then no distortion will be visible in the decompressed image.

The Ahumada/Peterson luminance-only detection model approximates the log of the contrast sensitivity function (the dependence of the inverse threshold contrast on spatial frequency) by a parabola in log spatial frequency. The predicted log luminance threshold of the m,nth DCT basis function is

$$\log T_{L,m,n} = \log \frac{s\ b_L}{r_L + (1-r_L)\cos^2\theta_{m,n}} + k_L(\log f_{m,n} - \log f_L)^2, \qquad m,n = 0, \ldots, N-1. \tag{3}$$

The minimum luminance threshold, $s\ b_L$, occurs at spatial frequency f_L, and k_L determines the steepness of the parabola. The parameter $0.0 < s < 1.0$ is to account for visual system summation of quantization errors over a spatial neighborhood. Such spatial summation causes a decrease in threshold. The spatial frequency, $f_{m,n}$, associated with the m,nth basis function, is given by

$$f_{m,n} = \frac{1}{2N}\sqrt{(m/W_x)^2 + (n/W_y)^2}, \tag{4}$$

where W_x is the horizontal and W_y the vertical size of a pixel in degrees of visual angle. The model includes a factor ($r_L + (1-r_L)\cos^2\theta_{m,n}$) which accounts for the imperfect summation of the two Fourier components present in basis functions having two cosine components (m and $n \neq 0$), and also accounts for the reduced sensitivity due to the obliqueness of these Fourier components. The magnitude of the summation/obliqueness effect is determined by $0.0 < r_L < 1.0$, and the angular parameter $\theta_{m,n}$ is given by

$$\theta_{m,n} = \arcsin\frac{2 f_{m,0} f_{0,n}}{f_{m,n}^2}. \tag{5}$$

Based on a fourth power summation rule for the two Fourier components[5], r_L is set to 0.6. The oblique effect can be included by decreasing the value of r_L.

Ahumada and Peterson[5] fit this model to the Peterson *et al.*[4] threshold data, and then used the grating detection data of Van Nes and Bouman[6] to derive luminance dependencies for b_L, f_L, and k_L, thus enabling the to model be used for a range of viewing conditions affecting luminance, contrast, and spatial frequency of the quantization errors. Since the single gun measurements of Peterson *et al.*[4] mainly varied the intensity of the

spatial modulation (chrominance remained relatively constant), the ability of the luminance-only model to predict visibility thresholds for modulations in combined luminance and chrominance directions was not adequately tested. Also, the replicated DCT basis functions used by Peterson *et al.*[4] have Fourier transforms possibly more like those of grating studies than those of single basis functions[7,8]. To address these issues, Peterson[9] made new threshold measurements of single basis function, single monitor-gun test images superimposed on a white background (1931 CIE coordinates: $X_0 = 37.27$, $Y_0 = 41.19$, $Z_0 = 29.65$). This configuration gives test stimuli having more significant chrominance modulation. Figure 1 shows the new measured thresholds for basis functions where m or $n = 0$.

A parabola representing a version of the luminance-only model is also shown in Figure 1. This model does a fair job of predicting the measured thresholds independent of color direction, except for the DC (m and $n = 0$) thresholds, which are obviously different for the three color guns. We propose that the lower thresholds for the R and B gun DC basis functions are the result of chromatic detection mechanisms having greater sensitivity than the luminance mechanism. Thus, even for quantization in the RGB color space, a luminance-only model is not quite sufficient. Color mechanisms must be taken into account to determine appropriate quantization levels for the DC coefficients. More importantly, for images compressed using isoluminant color directions, a complete color space discrimination model for the DCT basis functions is clearly needed.

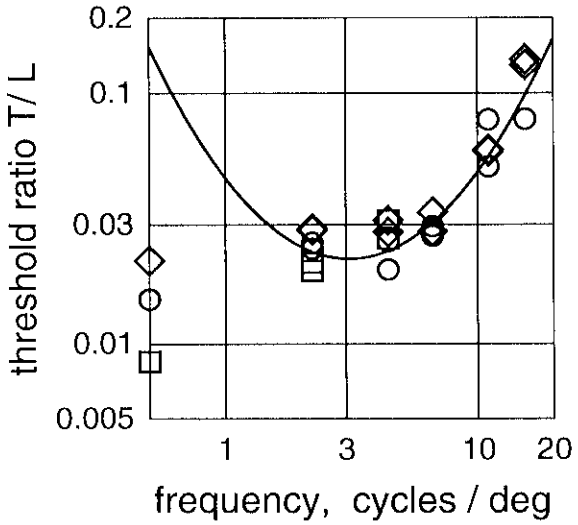

Figure 1: Visibility threshold contrast ratio measurements from Peterson[9] of single basis function, single monitor-gun test images superimposed on a white background, for basis functions where m or $n = 0$. Circles indicate R gun thresholds, diamonds indicate G gun thresholds, and squares indicate B gun thresholds. The points plotted at the far left of the graph are DC basis function (m and $n = 0$) thresholds. The parabola-shaped curve represents a version of the luminance-only model of Equation (3).

2.2 The Luminance/Chrominance Detection Model

To account for the DC sensitivities in the data of Figure 1, we add two chromatic channels to the luminance-only model. A large number of different color spaces have been proposed as appropriate bases for chromatic discriminations. We have selected for our chrominance channels those favored by Boynton[10]: a red-green opponent channel and a blue channel. The relation between these chromatic channels and the CIE 1931 XYZ color space is straightforward. The blue channel is just Z, and the red-green opponent channel O is given by $O = 0.47X - 0.37Y - 0.10Z$. This opponent channel is Boynton's[10] (Red-cone) − 2(Green-cone) channel, with the Red and Green XYZ cone responses taken from MacLeod and Boynton[10]. (We ignore the small correction developed by Vos[14] for going from the 1931 standard CIE values to the scientifically favored 1951 Judd CIE values used by MacLeod and Boynton.) Expressed in matrix form, the transformation from XYZ to

our YOZ opponent color space is

$$[\ \mathrm{YOZ}\ ] = [\ \mathrm{XYZ}\ ]\ {}_{\mathrm{XYZ}}\mathrm{M}_{\mathrm{YOZ}} = [\ \mathrm{XYZ}\ ] \begin{bmatrix} 0 & 0.47 & 0 \\ 1 & -0.37 & 0 \\ 0 & -0.10 & 1 \end{bmatrix} . \tag{6}$$

We model the frequency response of the Y channel with the luminance-only model described above. To reflect this, we subsequently refer to threshold $T_{L,m,n}$ as $T_{Y,m,n}$. The parameters in the luminance channel model will subsequently be referred to with a similar change of subscript (L → Y). To complete our luminance/chrominance model, we must also specify the shape of the frequency responses of the O and Z channels. Measurements of the spatial frequency responses of isoluminant chromatic modulations have typically found the chromatic sensitivity functions (the dependence of the inverse threshold contrasts on spatial frequency) to be low-pass in the frequency range of our basis functions and to be less sensitive at high spatial frequencies than the luminance channel.[11,12] We therefore model each of the O and Z log chromatic thresholds as a parabola, modified by setting it equal to its minimum value for all spatial frequencies to the left of the minimum. Since the data of Peterson[9] are too sparse to estimate two separate chromatic channels in close proximity, we make the simplifying assumption, supported by the results of Mullen[11], that both O and Z have the same shape spatial frequency response. The O and Z log chromatic thresholds for the m,nth DCT basis function can then be written:

$$\log \mathrm{T}_{\mathrm{O},m,n} = \begin{cases} \log \dfrac{s\ b_{\mathrm{O}}}{r_{\mathrm{OZ}} + (1-r_{\mathrm{OZ}})\cos^2\theta_{m,n}}, & \text{if } f_{m,n} \le f_{\mathrm{OZ}} \\ \log \dfrac{s\ b_{\mathrm{O}}}{r_{\mathrm{OZ}} + (1-r_{\mathrm{OZ}})\cos^2\theta_{m,n}} + k_{\mathrm{OZ}}(\log f_{m,n} - \log f_{\mathrm{OZ}})^2, & \text{if } f_{m,n} > f_{\mathrm{OZ}} \end{cases} . \tag{7a}$$

and

$$\log \mathrm{T}_{\mathrm{Z},m,n} = \begin{cases} \log \dfrac{s\ b_{\mathrm{Z}}}{r_{\mathrm{OZ}} + (1-r_{\mathrm{OZ}})\cos^2\theta_{m,n}}, & \text{if } f_{m,n} \le f_{\mathrm{OZ}} \\ \log \dfrac{s\ b_{\mathrm{Z}}}{r_{\mathrm{OZ}} + (1-r_{\mathrm{OZ}})\cos^2\theta_{m,n}} + k_{\mathrm{OZ}}(\log f_{m,n} - \log f_{\mathrm{OZ}})^2, & \text{if } f_{m,n} > f_{\mathrm{OZ}} \end{cases} . \tag{7b}$$

Note that Equations (7a) and (7b) are identical, except for the parameters b_{O} and b_{Z}; $\mathrm{T}_{\mathrm{O},m,n}$ and $\mathrm{T}_{\mathrm{Z},m,n}$ share the parameters s, k_{OZ}, f_{OZ}, and r_{OZ}. To obtain the overall model threshold $\mathrm{T}_{m,n}$ from the three channel thresholds, we use the "minimum of" combination rule:

$$\mathrm{T}_{m,n} = \min\{\ \mathrm{T}_{\mathrm{Y},m,n},\ \mathrm{T}_{\mathrm{O},m,n},\ \mathrm{T}_{\mathrm{Z},m,n}\ \} . \tag{8}$$

In order to estimate the parameters in the model described above, we fit the model to the data of Peterson[9] shown in Figure 1. Recall that the Peterson[9] thresholds were measured for single basis functions. To reflect the absence of a spatial summation effect in this data, we fixed $s = 1.0$ during the fitting process. This fit resulted in the parameter values shown in Table 1 for k_{Y}, f_{Y}, k_{OZ}, and f_{OZ}. We chose $r_{\mathrm{OZ}} = 0.6$, the same as r_{Y}.

Boynton[10] claims that at moderately high intensities, the Z channel's minimum threshold ($s\ b_{\mathrm{Z}}$ in our model) is approximately proportional to the background activity of the Z channel, and the minimum thresholds for the Y and O channels ($s\ b_{\mathrm{Y}}$ and $s\ b_{\mathrm{O}}$ in our model) are approximately proportional to the background Y. Based on the fit of our model to the Figure 1 data, we set the constants of proportionality to be: $b_{\mathrm{Y}} = 0.0219\,\mathrm{Y}_0$, $b_{\mathrm{O}} = 0.0080\,\mathrm{Y}_0$, and $b_{\mathrm{Z}} = 0.0647\,\mathrm{Z}_0$, where Y_0 and Z_0 are the CIE values of average white. To determine a value for s, we compared the thresholds measured in Peterson[9] to those measured by Van Nes and Bouman[6] for large test pattern sinusoidal gratings. The Peterson[9] thresholds are consistently higher than the Van Nes and Bouman[6] thresholds, a result attributable to spatial summation. Multiplication of the Peterson[9] data by 0.25 brings them into approximate agreement with the Van Nes and Bouman[6] data. We therefore chose $s = 0.25$. These results are summarized in Table 1.

Table 1. Parameter values estimated for the model of Equation (8).

model channel	parameter values				
	s	r	f	k	b
Y	0.25	0.6	3.1	1.34	$0.0219\,Y_0$
O	0.25	0.6	1.0	3.00	$0.0080\,Y_0$
Z	0.25	0.6	1.0	3.00	$0.0647\,Z_0$

As part of the model fitting, we also tried the Euclidean distance combination rule:

$$T_{m,n}{}^{-2} = T_{Y,m,n}{}^{-2} + T_{O,m,n}{}^{-2} + T_{Z,m,n}{}^{-2} . \tag{9}$$

However, when the data of Figure 1 were fit using this rule, in order to prevent contributions from the chromatic channels at low spatial frequencies, f_{OZ} was forced to be unrealistically low, and/or k_{OZ} was forced to be unrealistically high. This led to our selection of the "minimum of" rule for $T_{m,n}$.

3. QUANTIZATION MATRIX DESIGN

Quantization errors in an arbitrary color space are interpreted in the following way. Suppose we wish to compress a color image whose pixels are computed as a linear combination of XYZ values,

$$[DEF] = [XYZ]\,{}_{XYZ}M_{DEF} . \tag{10}$$

That is, the DCT is to be performed on an image in color space DEF, and ${}_{XYZ}M_{DEF}$ is the transformation from XYZ to DEF. The image in DEF space can be thought of as being transformed to XYZ space, and then converted by the visual system to YOZ space for discrimination. We need to determine limits on the sizes of errors in each of the D, E, and F color space dimensions, in order for the resulting errors in the Y, O, and Z channels to all be below the thresholds established by our model. These DEF thresholds determine the quantization matrices. For example, a unit error in the amplitude of a DCT coefficient in dimension D induces errors whose amplitudes in the Y, O, and Z channels are given by the first row of ${}_{DEF}M_{YOZ}$:

$${}_{DEF}M_{YOZ} = {}_{DEF}M_{XYZ} \times {}_{XYZ}M_{YOZ} = \begin{bmatrix} M_{1,1} & M_{1,2} & M_{1,3} \\ M_{2,1} & M_{2,2} & M_{2,3} \\ M_{3,1} & M_{3,2} & M_{3,3} \end{bmatrix} , \tag{11}$$

where ${}_{DEF}M_{XYZ}$ is the inverse of ${}_{XYZ}M_{DEF}$.

We now describe in detail the procedure to calculate $Q_{D,m,n}$, $Q_{E,m,n}$, and $Q_{F,m,n}$, the quantization matrix entries for DCT coefficient $I_{m,n}$ in the D, E, and F component images. First, using Equations (3) and (7), the display parameters W_x and W_y, and the model parameters given in Table 1, the model channel thresholds, $T_{Y,m,n}$, $T_{O,m,n}$, and $T_{Z,m,n}$, for the m,nth DCT basis function are calculated. Now let ${}_{Y}T_{D,m,n}$, ${}_{O}T_{D,m,n}$, and ${}_{Z}T_{D,m,n}$ indicate the thresholds imposed on the quantization error in the D component by the model's thresholds for the Y, O, and Z channels, respectively. Each of the Y, O, and Z model channel thresholds are converted to a D threshold as follows:

$${}_{Y}T_{D,m,n} = \frac{T_{Y,m,n}}{|M_{1,1}|} , \quad {}_{O}T_{D,m,n} = \frac{T_{O,m,n}}{|M_{1,2}|} , \text{ and } \quad {}_{Z}T_{D,m,n} = \frac{T_{Z,m,n}}{|M_{1,3}|} . \tag{12a}$$

Similarly for E and F:

$${}_{Y}T_{E,m,n} = \frac{T_{Y,m,n}}{|M_{2,1}|} , \quad {}_{O}T_{E,m,n} = \frac{T_{O,m,n}}{|M_{2,2}|} , \quad {}_{Z}T_{E,m,n} = \frac{T_{Z,m,n}}{|M_{2,3}|} , \tag{12b}$$

$${}_{Y}T_{F,m,n} = \frac{T_{Y,m,n}}{|M_{3,1}|} , \quad {}_{O}T_{F,m,n} = \frac{T_{O,m,n}}{|M_{3,2}|} , \quad {}_{Z}T_{F,m,n} = \frac{T_{Z,m,n}}{|M_{3,3}|} . \tag{12c}$$

Then the combination rule is used to determine the D, E, and F thresholds. We use the "minimum of" rule:

$$T_{D,m,n} = \min\{ {}_{Y}T_{D,m,n},\ {}_{O}T_{D,m,n},\ {}_{Z}T_{D,m,n} \}, \tag{13a}$$

$$T_{E,m,n} = \min\{ {}_{Y}T_{E,m,n},\ {}_{O}T_{E,m,n},\ {}_{Z}T_{E,m,n} \}, \tag{13b}$$

$$T_{F,m,n} = \min\{ {}_{Y}T_{F,m,n},\ {}_{O}T_{F,m,n},\ {}_{Z}T_{F,m,n} \}. \tag{13c}$$

Finally, the D, E, and F quantization matrix entries are obtained by dividing the thresholds above by the DCT normalization constants (α_m in Equation (1b)):

$$Q_{D,m,n} = 2\frac{T_{D,m,n}}{\alpha_m \alpha_n}, \qquad Q_{E,m,n} = 2\frac{T_{E,m,n}}{\alpha_m \alpha_n}, \qquad Q_{F,m,n} = 2\frac{T_{F,m,n}}{\alpha_m \alpha_n}. \tag{14}$$

The factor 2 results from the maximum quantization error being half the quantizer step size.

3.1 Quantization in RGB space

For quantization in monitor-RGB space, we require the matrix to transform from RGB to XYZ space, ${}_{RGB}M_{XYZ}$. Assuming that R, G, and B take on values between 0 and 1, ${}_{RGB}M_{XYZ}$ is the monitor calibration matrix giving the XYZ values for unit changes in each of the RGB signals. For our monitor,

$${}_{RGB}M_{XYZ} = \begin{bmatrix} 26.1 & 13.3 & 2.3 \\ 25.2 & 48.9 & 10.2 \\ 9.3 & 4.7 & 35.7 \end{bmatrix}. \tag{15}$$

This matrix is post-multiplied by ${}_{XYZ}M_{YOZ}$ to obtain ${}_{RGB}M_{YOZ}$:

$$[\ YOZ\ ] = [\ RGB\ ]\ {}_{RGB}M_{YOZ} = [\ RGB\ ] \begin{bmatrix} 13.3 & 7.1 & 2.3 \\ 48.9 & -7.3 & 10.2 \\ 4.7 & -0.9 & 35.7 \end{bmatrix}. \tag{16}$$

The matrix ${}_{RGB}M_{YOZ}$ gives the amplitude of the YOZ errors resulting from unit errors in RGB. These values indicate the sensitivity of the discrimination model YOZ channels to RGB errors. For example, a unit error in the R component leads to an error of 7.1 in the O channel of the model.

We can calculate the R, G, and B coordinate increments which induce a minimum threshold step in each of the Y, O, and Z channels. These are the the entries of ${}_{RGB}M_{YOZ}$ divided into the appropriate minimum threshold: $s\,b_Y$, $s\,b_O$, or $s\,b_Z$, calculated using the expressions in Table 1 and the Y_0 and Z_0 values of our average white. For example, letting $({}_{RGB}M_{YOZ})_{1,1}$, signify the upper left corner entry in ${}_{RGB}M_{YOZ}$, the increment in R which results in a minimum threshold change in Y is $(s\,b_Y) / ({}_{RGB}M_{YOZ})_{1,1}$. RGB minimum threshold increments calculated in this way are given in Table 2 for YOZ. Note that the minimum threshold for G is determined by the Y channel (0.0046 versus 0.0113 and 0.0469). That is, the Y channel imposes the strictest limit on G in order for a G change to not induce "too large" a change in YOZ-space. Similarly, the minimum threshold for R comes from the O channel (0.0116), and for B comes from the Z channel (0.0134). Following the procedure described above, using ${}_{RGB}M_{YOZ}$, the model parameters in Table 1, and the Y_0 and Z_0 values for our monitor, we obtain the quantization matrices shown in Table 3 for our RGB color space.

Table 2. Minimum thresholds imposed on R, G, and B quantization errors by the Y, O, and Z model minimum thresholds.

	Y	O	Z
R	0.0170	0.0116	0.2091
G	0.0046	-0.0113	0.0469
B	0.0483	-0.0881	0.0134

Table 3. RGB quantization matrices. The values in these matrices are obtained following the procedure described in Section 3. The $Q_{0,0}$ value is located in the upper left corner of each quantization matrix. As specified in the JPEG standard, the values have been rounded to the nearest integer. JPEG also requires that values in the quantization matrix be ≤ 255.

R quantization matrix	47	52	53	69	94	127	170	224
	52	57	53	60	75	98	128	167
	53	53	77	89	103	124	154	192
	69	60	89	119	142	166	197	236
	94	75	103	142	181	217	254	297
	127	98	124	166	217	269	320	373
	170	128	154	197	254	320	388	457
	224	167	192	236	297	373	457	544
G quantization matrix	19	14	14	19	26	35	46	61
	14	16	14	16	21	27	35	45
	14	14	21	24	28	34	42	52
	19	16	24	32	39	45	54	64
	26	21	28	39	49	59	69	81
	35	27	34	45	59	73	87	102
	46	35	42	54	69	87	106	124
	61	45	52	64	81	102	124	148
B quantization matrix	55	94	151	197	268	363	486	641
	94	164	151	171	216	281	367	477
	151	151	221	254	294	355	440	550
	197	171	254	340	406	475	562	675
	268	216	294	406	519	621	727	851
	363	281	355	475	621	770	915	1066
	486	367	440	562	727	915	1109	1306
	641	477	550	675	851	1066	1306	1556

Figure 2 plots all the measured R, G, and B gun, single basis function thresholds from Peterson[9] (including those for the dual frequency (m and $n \neq 0$) basis functions), after correction by the summation/obliqueness factors of Equations (3) and (7). Figure 2 also shows the curves for the model threshold predictions $T_{Y,m,n}$, $T_{O,m,n}$, and $T_{Z,m,n}$, using the parameters in Table 1, except with $s = 1.0$. This value for s was used to reflect the absence of a spatial summation effect in the single basis function data. In addition, the $T_{O,m,n}$ and $T_{Z,m,n}$ threshold prediction curves have been converted to luminance units, since all the threshold data plotted are in luminance units. This is accomplished by multiplying the $T_{O,m,n}$ threshold predictions by 13.3 / 7.1, and the $T_{Z,m,n}$ threshold predictions by 4.7 / 35.7. These factors are obtained from the ${}_{RGB}M_{YOZ}$ matrix. Figure 2 shows that for the B component, the DC and lowest spatial frequency thresholds are determined by the Z channel, and for the R component, the DC threshold is determined by the O channel. All the G thresholds are assumed to be determined by the Y channel. Note that the DC threshold for the Y channel (which we assume to be the DC threshold measured for G) is not predicted on a theoretical basis. The dot-dashed line in Figure 2 demonstrates that the measured DC threshold for G, and hence our DC threshold for Y, was found to be approximately equal to the minimum threshold of the Y channel.

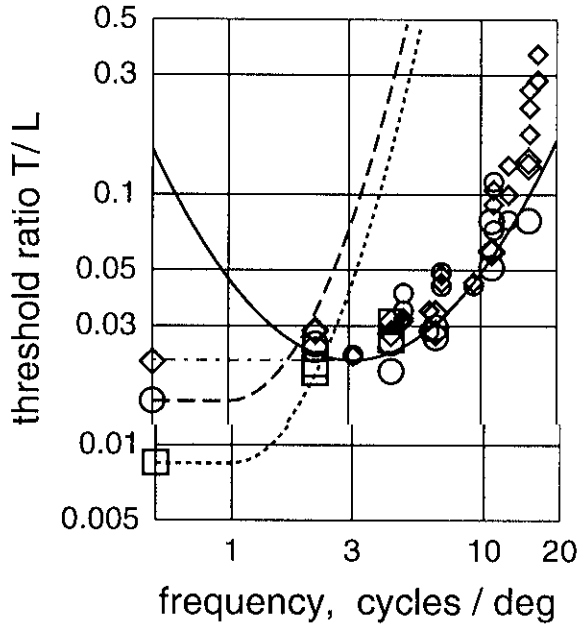

Figure 2: Visibility threshold contrast ratio measurements from Peterson[9] of single basis function, single monitor-gun test images superimposed on a white background, for all basis functions ($0 \le m, n < N$). Circles indicate R gun thresholds, diamond-shaped points indicate G gun thresholds, and squares indicate B gun thresholds. Large symbols indicate thresholds for basis functions for which m or $n = 0$; thresholds for basis functions for which m and $n \neq 0$ are indicated with small symbols. The points plotted at the far left of the graph are DC basis function (m and $n = 0$) thresholds. All thresholds are plotted after correction by the summation/obliqueness factors of Equations (3 & 7), with $r_Y = r_{OZ} = 0.6$. The solid, dashed, and dotted curves show the channel model threshold predictions for $T_{Y,m,n}$ (Equation (3)), $T_{O,m,n}$ (Equation (7a)), and $T_{Z,m,n}$ (Equation (7b)), respectively, using the parameters and conversion factors as described in Section 3.1. The dot-dashed line shows that the measured G DC threshold, which is the DC threshold for the Y channel, is approximately equal to the minimum threshold of the Y channel.

3.2 Quantization in YC_rC_b space

The visual system is known to be much more sensitive to high spatial frequency information in luminance, compared to chrominance. In an attempt to put all luminance information in a single channel which can then be compressed with maximum fidelity, color images are often represented in the YC_rC_b color space for image compression. Pennebaker and Mitchell[2] give the transformation from RGB to YC_rC_b as

$$Y' = 0.3\,R + 0.6\,G + 0.1\,B\ , \tag{17a}$$

$$C_r = (R - Y')/1.6 + 0.5\ , \tag{17b}$$

$$C_b = (B - Y')/2 + 0.5\ . \tag{17c}$$

The additive constants keep the coefficients between 0 and 1, and can be ignored for our purposes. The matrix ${}_{YC_rC_b}M_{YOZ}$ is obtained from ${}_{YC_rC_b}M_{RGB}$ (the inverse of the matrix of coefficients above) as follows:

$${}_{YC_rC_b}M_{YOZ} = {}_{YC_rC_b}M_{RGB} \times {}_{RGB}M_{XYZ} \times {}_{XYZ}M_{YOZ} = \begin{bmatrix} 1.0 & 1.0 & 1.0 \\ 1.6 & -0.8 & 0.0 \\ 0.0 & -0.3 & 2.0 \end{bmatrix} {}_{RGB}M_{YOZ} \tag{18}$$

$$= \begin{bmatrix} 66.9 & -1.1 & 48.2 \\ -17.8 & 17.1 & -4.5 \\ -7.0 & 0.6 & 67.9 \end{bmatrix}.$$

The matrix ${}_{YC_rC_b}M_{YOZ}$ gives the amplitude of the YOZ errors resulting from unit errors in YC_rC_b. These values indicate the sensitivity of the discrimination model YOZ channels to YC_rC_b errors.

In a fashion analogous to Table 2 for RGB, Table 4 shows the increments in Y', C_r, and C_b which induce minimum threshold changes in the Y, O, and Z channels. These are calculated in the same way as for RGB, but using ${}_{YC_rC_b}M_{YOZ}$ instead of ${}_{RGB}M_{YOZ}$. From Table 4, we see that the minimum thresholds for the chromatic channels C_r and C_b are determined by the model color discrimination channels O and Z, respectively. In addition, the ratio of these thresholds to the Y channel thresholds are relatively low, so that the "minimum of" combination rule results in the O and Z channels determining $T_{Y',m,n}$, $T_{C_r,m,n}$, and $T_{C_b,m,n}$ for a wider range of spatial frequencies than O and Z did for $T_{R,m,n}$, $T_{G,m,n}$, and $T_{B,m,n}$. Table 5 shows the quantization matrices for YC_rC_b color space.

Table 4. Minimum thresholds imposed on Y', C_r, and C_b quantization errors by the Y, O, and Z model minimum thresholds.

	Y	O	Z
Y'	0.0034	-0.0733	0.0099
C_r	-0.0127	0.0048	-0.1066
C_b	-0.0323	0.1473	0.0071

Table 5. YC_rC_b quantization matrices. The values are obtained as described in Section 3 and formatted as in Table 3.

Y' quantization matrix	14	10	11	14	19	25	34	45
	10	11	11	12	15	20	26	33
	11	11	15	18	21	25	31	38
	14	12	18	24	28	33	39	47
	19	15	21	28	36	43	51	59
	25	20	25	33	43	54	64	74
	34	26	31	39	51	64	77	91
	45	33	38	47	59	74	91	108
C_r quantization matrix	20	34	39	52	70	95	127	168
	34	43	40	45	57	74	96	125
	39	40	58	67	77	93	115	144
	52	45	67	89	107	125	147	177
	70	57	77	107	136	163	191	223
	95	74	93	125	163	202	240	280
	127	96	115	147	191	240	291	342
	168	125	144	177	223	280	342	408
C_b quantization matrix	29	49	101	132	179	243	325	428
	49	110	101	114	144	188	245	319
	101	101	148	170	197	237	294	367
	132	114	170	227	272	318	376	451
	179	144	197	272	347	415	486	569
	243	188	237	318	415	514	611	713
	325	245	294	376	486	611	741	873
	428	319	367	451	569	713	873	1040

4. CONCLUSIONS

We have presented a model for predicting visibility thresholds for DCT coefficient quantization error, from which quantization matrices for use in DCT-based compression can be designed. We estimated values for the parameters of our model based on experimentally measured visibility thresholds. The frequency parameters we estimated, f_Y and f_{OZ}, agree fairly well with results others have reported for similar parameters. The values we have estimated for k_Y and k_{OZ} are similar to those estimated by others, however we have found these parameters to vary for the different experimentally measured thresholds. The value we have proposed for the obliqueness/summation parameters, r_Y and r_{OZ}, only reflects summation and does not reflect an effect due to obliqueness. More data may be needed to more determine values for k_Y, k_{OZ}, r_Y, and r_{OZ} more reliably; though those we propose here are reasonable and result in quantization matrices which perform well in preliminary tests. The value for s we have proposed is based on a limited amount of data. Further experiments are needed to determine the spatial extent over which summation occurs among DCT quantization errors, in order to estimate s more accurately.

The quantization matrices computed by the techniques described above take no account of image content. A promising extension of this model may be to optimize the quantization matrices for individual images or a class of images. That is, use an image-dependent approach to quantization matrix design. Watson[15] has shown how this may be done for grayscale images, by taking into account local light adaptation, local contrast masking, and error pooling. Watson's technique can be extended to the case of color images by adopting rules governing masking and adaptation within the O and Z channels.

5. ACKNOWLEDGMENTS

We appreciate the help of Jeffrey B. Mulligan. 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. A. Peterson, H. Peng, J. H. Morgan, W. B. Pennebaker, "Quantization of color image components in the DCT domain", in B. E. Rogowitz, M. H. Brill, J. P. Allebach, eds., *Human Vision, Visual Processing, and Digital Display II*, Proc. SPIE, vol. 1453, pp. 210-222, 1991.
5. A. J. Ahumada, Jr., H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," in B. E. Rogowitz, ed., *Human Vision, Visual Processing, and Digital Display III*, Proc. SPIE, vol. 1666, pp. 365-374, 1992.
6. F. L. van Nes, M. A. Bouman, "Spatial modulation transfer in the human eye", *Journal of the Optical Society of America*, vol. 57, pp. 401-406, 1967.
7. R. J. Clarke, "Spectral response of the discrete cosine and Walsh-Hadamard transforms," *IEEE Proceedings*, vol. 130, pp. 309-313, 1983.
8. S. A. Klein, A. D. Silverstein, T. Carney, "Relevance of human vision to JPEG-DCT compression," in B. E. Rogowitz, ed., *Human Vision, Visual Processing, and Digital Display III*, Proc. SPIE, vol. 1666, pp. 200-215, 1992.
9. 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.

10. R. M. Boynton, *Human Color Vision*, Holt, Rinehart, and Winston, New York, 1979
11. K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," *Journal of Physiology*, vol. 359, pp. 381-400, 1985.
12. D. H. Kelly, "Spatio-temporal frequency characteristics of color-vision mechanisms," *Journal of the Optical Society of America*, vol. 64, pp. 55-72, 1974.
13. D. I. A. MacLeod, R. M. Boynton, "Chromaticity diagram showing cone excitation by stimuli of equal luminance," *Journal of the Optical Society of America*, vol. 69, pp. 1183-1186, 1979.
14. J. J. Vos, "Colorimetric and photometric properties of a 2° fundamental observer," *Color Research and Application*, vol. 3, pp. 125-128, 1978.
15. A. B. Watson, "DCT' quantization matrices visually optimized for individual images," in B. E. Rogowitz, J. P. Allebach, eds., *Human Vision, Visual Processing, and Digital Display IV*, (SPIE, Bellingham, WA, 1993).