Proceedings of the IEEE International Conference on Image Processing, Austin, TX, Nov., 1994.

PERCEPTUAL OPTIMIZATION OF DCT COLOR QUANTIZATION MATRICES

Andrew B. Watson

MS 262-2 NASA Ames Research Center
Moffett Field CA 94035-1000
beau@vision.arc.nasa.gov

ABSTRACT

The JPEG baseline standard for image compression employs a block Discrete Cosine Transform (DCT) and uniform quantization. For a monochrome image, a single quantization matrix is allowed, while for a color image, distinct matrices are allowed for each color channel.. Here we describe a method, called DCTune, for design of color quantization matrices that is based on a model of the visibility of quantization artifacts. The model describes artifact visibility as a function of DCT frequency, color channel, and display resolution and brightness. The model also describes summation of artifacts over space and frequency, and masking of artifacts by the image itself. The DCTune matrices are different from the de facto JPEG matrices, and appear to provide superior visual quality at equal bit-rates.

1. INTRODUCTION

Many image compression schemes employ a block Discrete Cosine Transform (DCT) followed by uniform quantization. Acceptable rate/distortion performance depends upon proper design of the quantization matrix. In previous work, we showed how to use a model of the visibility of DCT basis functions to design quantization matrices for arbitrary display resolutions and color spaces [1, 2, 3]. In recent research we elaborated this model to incorporate the effects of display resolution [4], frequency summation [5], spatial summation [6], and contrast masking [7]. Subsequently, I showed how to optimize greyscale quantization matrices for individual images, for optimal rate/perceptual distortion performance [8]. I dubbed this technique "DCTune." Here I describe extensions of the DCTune algorithm to color images.

2. DESCRIPTION OF THE ALGORITHM

2.1. Quantization Error

The theoretical basis for this algorithm is provided in [8]. Consider an image defined in a particular color space. We identify the space and index the three color channels by the symbol θ. We will call θ the quantization color space, because it is the space in which the quantization is performed. As an example, many current applications use YCbCr as a quantization color space [9]. We write the blocked DCT of this image as $c_{u,v,b,\theta}$, where u,v are the indices of the DCT frequency, which each range from 0 to 7, and b is the block index. Quantization is done by division of the blocked coefficients by the corresponding quantization matrix $q_{u,v,\theta}$. The quantized DCT coefficients are therefore given by

$$k_{u,v,b,\theta} = \mathrm{Round}\left[c_{u,v,b,\theta}/q_{u,v,\theta}\right] \quad . \tag{1}$$

The quantization error is then

$$e_{u,v,b,\theta} = c_{u,v,b,\theta} - k_{u,v,b,\theta}\, q_{u,v,\theta} \quad . \tag{2}$$

2.2. Visibility Model

The model by which we compute the visibility of quantization error is described in [2]. That model contains a function that returns the visibility threshold for DCT basis functions as a function of frequency, mean luminance, and perceptual color channel. The three perceptual color channels employed are called Y (luminance), O (opponent, or red-green), and Z (blue). We identify and index this space by the symbol ϕ. The threshold function can be written symbolically as

$$t_{u,v,\phi} = V\left[u, v, \phi, Y, p_x, p_y, \ldots\right] \quad . \tag{3}$$

In the present algorithm, the value of Y (luminance) is set to the mean luminance of the display, and p_X, p_Y are set to the horizontal and vertical dimensions of a single pixel in degrees of visual angle. Values are computed for the 8x8 frequencies in the three color channels.

2.3. Transforming the DCT coefficients into the perceptual space.

The DCT coefficients in the quantization color space θ can be transformed into the perceptual color space ϕ by an appropriate transformation. If the color spaces are linearly related, this transformation is multiplication by a 3x3 matrix. The transformed DCT coefficients are written $c_{u,v,b,\phi}$.

2.4. Luminance Masking

Human visual sensitivity to luminance patterns is reduced as the mean local adapting luminance is increased. To implement this effect, the thresholds are adjusted by a power function of the DC coefficient of each block in the luminance channel ($c_{0,0,b,Y}$) relative to the DC coefficient corresponding to the display mean luminance ($\bar{c}_{0,0,Y}$ = 1024 for 8 bit images)

$$a_{u,v,b,\phi} = t_{u,v,\phi} \left(c_{0,0,b,Y} / \bar{c}_{0,0,Y} \right)^{a_T} , \qquad (4)$$

where a_T is a parameter. Note that all three perceptual color channels are adjusted by the *luminance* channel coefficients. This treatment of the two color channels (O and Z) is somewhat speculative.

2.5. Contrast Masking

Next the thresholds are adjusted for contrast masking. The adjustment factor is a power function, of the ratio of the coefficient magnitude to the luminance-adjusted threshold

$$m_{u,v,b,\phi} = a_{u,v,b,\phi} \operatorname{Max}\left[1 , \left| \frac{c_{u,v,b,\phi}}{a_{u,v,b,\phi}} \right|^{w_{u,v,\phi}} \right] . \qquad (5)$$

This factor has a floor of 1, so that only suprathreshold contrasts produce any masking. The exponent $w_{u,v,\phi}$ is allowed to vary with frequency and perceptual color channel, but we have typically used a constant value of 0.7.

This masking model is derived from classical results on masking of sinusoidal luminance gratings[10, 11]. We have recently provided psychophysical evidence that it is also a reasonable model for masking of DCT luminance basis functions[7].. The limited information on masking within the chromatic channels [12]. suggests essentially similar behavior.

This model of masking explicitly assumes that masking occurs only within a single block, frequency, and color channel. This assumption is undoubtedly wrong on all three counts [7].[12] , but the errors thereby introduced were thought to be small and worth the computational simplification. A possible exception is the rather substantial masking of luminance signals by chromatic masks[12] which may be worth incorporating into a future version of the masking model.

2.6. Transforming Quantization Error into the Perceptual Color Space

The quantization errors in the quantization color space θ can be transformed into the perceptual color space ϕ by an appropriate transformation. If the color spaces are linearly related, this transformation is multiplication by a 3x3 matrix consisting of partial derivatives relating the two spaces. An example of such a matrix, from YC_bC_r to the particular perceptual color space YOZ, is given in [2]. The transformed quantization errors are written $e_{u,v,b,\phi}$.

2.7. Just-Noticeable-Differences

The quantization errors, in the perceptual color space, are then divided by the luminance and contrast adjusted thresholds to yield just-noticeable-differences (jnd's):

$$j_{u,v,b,\phi} = e_{u,v,b,\phi} / m_{u,v,b,\phi} . \qquad (6)$$

These jnd's are then pooled over blocks, to yield a perceptual error matrix for each frequency and perceptual color channel. The pooling employs a Minkowski summation with an exponent of β_s :

$$p_{u,v,\phi} = \left(\sum_{b} \left| j_{u,v,b,\phi} \right|^{\beta_s} \right)^{1/\beta_s} . \qquad (7)$$

The perceptual error matrices are then pooled over frequency, with an exponent of β_f, to yield a total perceptual error for each perceptual color channel:

$$P_{\phi} = \left(\sum_{u,v} p_{u,v,\phi}{}^{\beta_f} \right)^{1/\beta_f} . \qquad (8)$$

Finally, these perceptual errors are pooled over color channel, with exponent β_c, to yield a total perceptual error:

$$P = \left(\sum_{\phi} P_{\phi}^{\beta_c} \right)^{1/\beta_c} \quad . \tag{9}$$

3. OPTIMIZATION

It can be shown that when β_f and β_C are infinite, the minimum perceptual error for any particular bit-rate is obtained when

$$p_{u,v,\phi} = \psi \quad . \tag{10}$$

While we believe that β_f and β_C are probably equal to β_S, and around 4, we believe that the algorithmic simplification afforded by this assumption greatly outweighs the possible errors introduced.

A direct strategy is therefore to set ψ to some value, and for each frequency u,v, to vary the three numbers $q_{u,v,\theta}$ until $p_{u,v,\phi} = \psi$ for all three indices of ϕ (Y,O, and Z). Since the quantization matrix entries are integers between 1 and 255 for 8-bit images, the search space contains 255^3 locations. Standard gradient techniques cannot be used because the error functions are not always monotonic.

3.1. Hierarchical Direct Search

We have experimented with a hierarchical direct search algorithm in which the 3-space of $q_{u,v,\theta}$ is first sampled very coarsely (three samples in each dimension) and then progressively refined through bisection of the sampling intervals in the vicinity of the local optimum. We first define a new scalar error function that describes the proximity of the perceptual error matrix entries in each color channel to the target value:

$$E_{u,v} = \left(\sum_{\phi} \left| p_{u,v,\phi} - \psi \right|^{\beta_E} \right)^{1/\beta_E} \quad . \tag{11}$$

In some cases, even the maximum quantization ($q_{u,v,\theta}$= {255,255,255}) yields a value $p_{u,v,\phi}$ that is less than the target, in which case we substitute this maximum for the target. This search procedure is practical, but still requires considerable computation.

3.2. Optimization in Quantization Color Space

Since it is unlikely that most applications will expend the considerable computation required by the preceding method, we have also considered an alternative approach in which the existing quantization channels (eg YC_bC_r) are taken as approximations to the perceptual channels. In effect, we set $\phi = \theta$. This means that the transformation of quantization errors from quantization color space to perceptual color space described in section 2.5 becomes an identity transformation. This in turn means that variation of a particular quantization matrix entry affects only the perceptual error in the corresponding frequency and quantization color channel. Thus the 3x64=192 quantization matrix entries may each be optimized independently, in separate one-dimensional optimizations. This is the approach we have pursued most extensively.

4. DOWN-SAMPLING OF CHROMATIC CHANNELS

The JPEG standard permits the chromatic channels to be down-sampled horizontally and/or vertically before compression. This greatly complicates the optimization in perceptual color space, since the transformation of quantization errors from quantization to color spaces involves spatial as well as color transformations, and we have not attempted this approach. When the optimization is conducted in the quantization color space, each color channel is optimized separately and down-sampling introduces no great complexities. It is only necessary to correctly compute the pixel sizes for the down-sampled chromatic channels in Eq. (3). This simplicity provides an additional argument for optimization in the quantization color space.

5. RESULTS

Figure 1 shows a grayscale miniature of an image for which we designed an optimized matrix. The image was acquired from a Kodak PhotoCD demonstration disc at a resolution of 512x796 pixels. It was converted to RGB (24 bits/pixel) in Abobe Photoshop

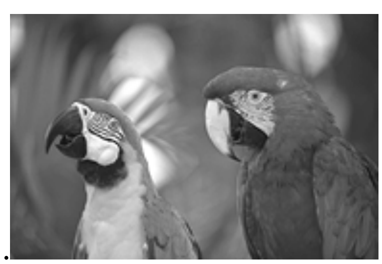

Figure 1. Grayscale version of color image used to design the matrices shown in Fig. 2.

Figure 2 shows color quantization matrices optimized in YCbCr space for a bit rate of 0.25

bits/pixel. The target display resolution and luminance were set to 64 pixels/deg and 33.45 cd/m^2. The image was transformed to YCbCr, and the two color channels Cb and Cr were down-sampled by two in both dimensions before compression. With this matrix, the reconstructed image was nearly perceptually lossless at the specified visual resolution. In this printed paper we cannot achieve the necessary color or resolution to show the reconstructed image.

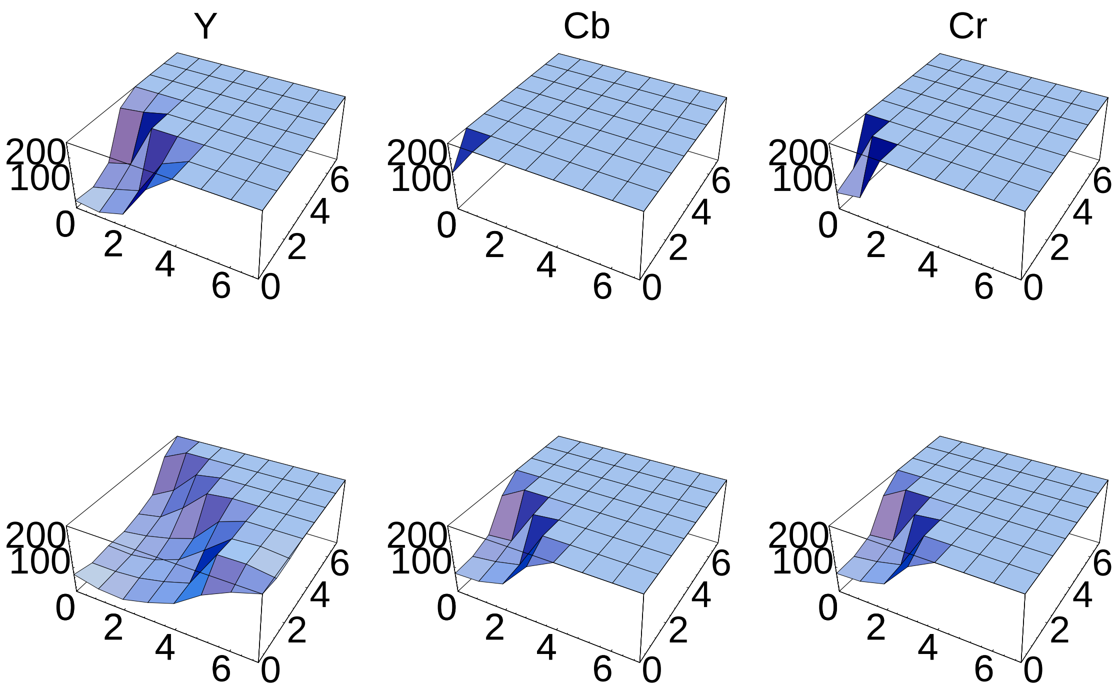

Figure 2. YCbCr Color quantization matrices produced by the DCTune method (top) and by scalar multiplication of the de facto JPEG quantization matrices (bottom).

The JPEG standards documents provide an example set of quantization matrices [9].. While not part of the standard, they have become almost universally used, and for this reason I call them the "de facto" matrices. Variations in bit-rate and quality are typically obtained by scalar multiplication of these matrices.

Figure 2 also shows for comparison the matrices that result from the scalar multiplication of the "de facto" JPEG quantization matrices that yields 0.25 bits/pixel for the same image. Several differences between the two sets of matrices should be noted. First, in the luminance (Y) matrices (leftmost column of Fig. 2) the DCTune matrix, viewed as a surface, is narrower but deeper. This means that high spatial frequencies are more severely compressed, but that low frequencies, and especially the DC coefficient, are less severely quantized. In particular, the DC quantization coefficient is 31 for DCTune, and 70 for de facto JPEG.

A second difference is that the DCTune chromatic matrices (Cb and Cr) are both shallower and narrower than the comparable JPEG de facto matrices. DCTune compresses the color more severely. A third difference is that while there is only a single de facto matrix for color (Cb and Cr use the same matrix), DCTune creates different matrices for the two chromatic channels. The Cb channel is narrower and shallower than Cr. This is a consequence of the color model, in which the "blue" perceptual channel (Z) is lower resolution than the opponent channel (O).

It should be noted that the matrix design of Fig. 2 was undertaken for a particular display resolution of

64 pixels/deg. This corresponds to 72 pixels/inch (typical of desktop computer displays) viewed at 51 inches. This is a higher resolution (longer viewing distance) than that typically encountered in desktop environments, but is much less than some high-end printing resolutions. A virtue of our technique is that it automatically adapts to the specified resolution, while the JPEG de facto matrices are invariant with display resolution.

We have recently completed psychophysical tests on grayscale images which showed that for a set of 10 test images, compression with JPEG de facto matrices required on average 39% more bits to achieve perceptual losslessness than did DCTune optimized matrices [13]. We are now conducting similar tests for color images.

6. CONCLUSIONS

I have described an algorithm for design of color quantization matrices for DCT-based image compression such the JPEG compression standard. The method attempts to discover the matrix that yields a minimum perceptual error for a specified bit-rate, or minimum bit-rate for a specified perceptual error. Perceptual error is computed from a model of visual sensitivity to DCT basis functions. Initial results indicate that DCTune matrices are superior to de facto JPEG matrices.

7. REFERENCES

1. Ahumada, A. J., Jr. and A. B. Watson. A visual detection model for DCT coefficient quantization. Technology 2003. **2**: 404-415, 1993.
2. Ahumada, A. J., Jr. and H. A. Peterson. A visual detection model for DCT coefficient quantization. Computing in Aerospace 9. 314-318, 1993.
3. Peterson, H., A. J. Ahumada Jr. and A. Watson. "An Improved Detection Model for DCT Coefficient Quantization." Human Vision, Visual Processing, and Digital Display IV. Allebach ed. 1993 SPIE. Bellingham, WA.
4. Watson, A. B., A. Gale, J. A. Solomon and A. J. Ahumada Jr. Visibility of DCT Quantization Noise: Effects of Display Resolution. Society for Information Display. 697-700, 1994.
5. Peterson, H. A., J. A. J. Ahumada and A. B. Watson. "Visibility of DCT Quantization Noise: Spatial Frequency Summation." Society for Information Display International Symposium Digest of Technical Papers,. Morreale ed. 1994 SID. Santa Ana, CA.
6. Peterson, H. A., J. A. Ahumada and A. B. Watson. "The Visibility of DCT Quantization Noise." Society for Information Display International Symposium Digest of Technical Papers,. Morreale ed. 1993 SID. Playa del Rey, CA.
7. Solomon, J. A., A. B. Watson and J. A. J. Ahumada. "Visibility of DCT Quantization Noise: Contrast Masking." Society for Information Display International Symposium Digest of Technical Papers. Morreale ed. 1994 Society for Information Display. Santa Ana, CA.
8. Watson, A. B. "DCT quantization matrices visually optimized for individual images." Human Vision, Visual Processing, and Digital Display IV. Rogowitz ed. 1993 SPIE. Bellingham, WA.
9. Pennebaker, W. B. and J. L. Mitchell. "JPEG Still image data compression standard." 1993 Van Nostrand Reinhold. New York.
10. Legge, G. E. A power law for contrast discrimination. Vision Research. **21**: 457-467, 1981.
11. Legge, G. E. and J. M. Foley. Contrast masking in human vision. Journal of the Optical Society of America. **70**(12): 1458-1471, 1980.
12. De Valois, K. K. and E. Switkes. Simultaneous masking interactions between chromatic and luminance gratings. Journal of the Optical Society of America A. **73**: 11-18, 1983.
13. Watson, A. B. and J. A. Solomon. Perceptaul evaluation of optimized JPEG quantization matrices. in preparation. : 1994.