Perceptual adaptive JPEG coding

Ruth Rosenholtz and Andrew B. Watson

MS 262-2 NASA Ames Research Center
Moffett Field, CA 94035-1000
email: rruth@vision.arc.nasa.gov

ABSTRACT

An extension to the JPEG standard (ISO/IEC DIS 10918-3) allows spatial adaptive coding of still images. As with baseline JPEG coding, one quantization matrix applies to an entire image channel, but in addition the user may specify a multiplier for each 8x8 block, which multiplies the quantization matrix, yielding the new matrix for that block. MPEG 1 and 2 use much the same scheme, except there the multiplier changes only on macroblock boundaries.

We propose a method for perceptual optimization of the set of multipliers. We compute the perceptual error for each block based upon DCT quantization error adjusted according to contrast sensitivity, light adaptation, and contrast masking, and pick the set of multipliers which yield maximally flat perceptual error over the blocks of the image. We investigate the bitrate savings due to this adaptive coding scheme and the relative importance of the different sorts of masking on adaptive coding.

1. ADAPTIVE JPEG

The JPEG image compression standard for the compression of both grayscale and color continuous-tone still images is based upon the Discrete Cosine Transform (*DCT*) of 8x8 image blocks, followed by lossy quantization and lossless entropy coding of the results. For clarity, we concentrate on the case of a single-component, grayscale image. In basic JPEG encoding, quantization is accomplished through division and rounding, according to the rule

$$u_{ijk} = \text{Round}\left[c_{ijk} / q_{ij}\right] \tag{1}$$

where c_{ijk} is the (i,j)th coefficient in the kth block (in raster scan order), and q_{ij} are the elements of the *quantization matrix*, Q (integers between 1 and 255). The user specifies a single quantization matrix which is used throughout the image.

To compress the image as much as possible without visible artifacts, one should design Q with the visibility of the quantization error in each of the DCT coefficients in mind, more finely quantizing coefficients with more visible quantization error, and more coarsely quantizing coefficients with less visible error. Ahumada & Peterson [1], and Peterson, Ahumada, & Watson [2] have shown how to compute this "perceptually lossless" Q for a variety of viewing conditions. Watson [3] demonstrated a technique for optimizing Q for a particular image which finds Q for a given image quality or for a given bitrate.

However, the visibility of quantization artifacts varies with (among other things) the local luminance level and the local contrast. In order to fully exploit varying local image characteristics, we need a spatially adaptive quantization scheme.

ISO/IEC DIS 10918-3 specifies an extension to the JPEG standard which allows limited adaptive coding within an image. Specifically, while Q remains fixed throughout the image, the user may specify a multiplier m_k for each block k. The quantized coefficient is then

$$u_{ijk} = \text{Round}\left[c_{ijk} / \left(q_{ij} \cdot m_k\right)\right] \tag{2}$$

This scaling applies only to the AC coefficients; the DC quantization remains unchanged. The user initially chooses one of two prespecified tables (linear or non-linear) of 31 possible multiplier values. For the linear table, used for the examples in this paper, the multipliers are 1/16 through 31/16. A 5 bit index identifies a particular table entry, and a single bit code is used to signal blocks in which the multiplier changes (unchanged blocks inherit the multiplier from the previous block, in raster scan order). Thus there is a cost of 6 bits for each change in the multiplier. We note that MPEG uses a similar scheme, so that studying this form of adaptive coding for still images may provide insight into adaptive MPEG coding.

2. THE PERCEPTUAL ERROR METRIC

For adaptive quantization, we need a measure of the local *perceptual error*, i.e. the local visibility of quantization error. Here we adopt the DCT-based perceptual error metric developed by Watson [3]. In that metric, the quantization errors for each coefficient in each block are scaled by the corresponding visual sensitivities to each DCT basis function in each block. These sensitivities

0-7803-3258-X/96/$5.00

are determined by three factors: contrast sensitivity, luminance masking, and contrast masking.

The *contrast sensitivity* v_{ij} for each DCT basis function has been measured by Ahumada *et al.* [1] who also provided a formula for approximating this sensitivity under a variety of viewing conditions.

Due to light adaptation, the greater the mean luminance of an image region, the greater the amplitude required to see a pattern within that region. Watson [3] has suggested that one may approximate this *luminance masking* by computing block contrast sensitivities

$$v_{ijk} = v_{ij}\left(c_{00k} / \bar{c}_{00}\right)^{-a_T} \quad (3)$$

where c_{00k} is the DC coefficient of block k, $\bar{c}_{00}$ is the mean luminance of the display, and a_T determines the degree of masking (he recommends a value of 0.65).

The visibility of a pattern is reduced by the presence of other components in the image. The effect is strongest when the two components appear at the same location in the image and share the same spatial frequency. We follow Watson [3] and model this *contrast masking* as

$$s_{ijk} = v_{ijk} \cdot \mathrm{Min}\left[1, \left|c_{ijk}\right| v_{ijk}\right]^{-w_{ij}} \quad (4)$$

where s_{ijk} is the *block masked contrast sensitivity* and w_{ij} determines the degree of contrast masking (Watson recommends $w_{00} = 0$, and $w_{ij} = 0.7$ for all other coefficients).

Perceptual error in each frequency of each block is then computed by multiplying the quantization error e_{ijk} by the block masked contrast sensitivity,

$$d_{ijk} = e_{ijk} s_{ijk} \quad (5)$$

To get the total perceptual error, we need to pool the errors over both frequency and space. To simplify our optimization procedure, we take

$$P = \mathrm{Max}\left(d_{ijk}\right) \quad (6)$$

3. OPTIMIZATION

Ideally one should jointly optimize the quantization matrix, Q, and the matrix of multipliers, M. To simplifiy matters, we find the two matrices separately. The optimal Q and M should be reasonably independent, since it seems unlikely that spatially adapting the quantizer will greatly change the shape of the optimal Q, which indicates the relative coarseness with which to code different frequencies. Furthermore, jointly optimizing the two matrices would be unrealistic for MPEG, in which one specifies Q for an entire group of pictures and would thus have to jointly optimize Q and, say, 12 M matrices.

We first optimize Q using the method described by Watson [3], with spatial and frequency error pooling performed by taking the maximum over the blocks and frequencies, respectively.

The procedure for optimizing M for a given total perceptual error, P, is quite parallel to that used to find Q. Since P is equal to the maximum block perceptual error, p_k, the minimum bitrate for a given P is obtained when all $p_k = P$. Therefore the optimization of M reduces to separately adjusting each scale factor, m_k, until the block perceptual error equals the desired total perceptual error. Here we separately optimize the elements of M to get flat perceptual error across the blocks of the image, whereas Watson separately optimized the elements of Q to get flat perceptual error across frequency. The perceptual error in block k is a roughly monotonic function of m_k, so efficient search procedures may be used.

We must spend 6 additional bits for every block in which the quantization multiplier differs from the previous block (in raster-scan order). For large images this number of bits may be small compared to the total number used to code the image. When no constraints are placed on the number of multiplier changes, in several examples we found that the percentage of bits spent on adaptive coding overhead was as high as 6% for a 384x256 image. In the MPEG-2 Test Model 5 encoder, the index into the table of scale factors must differ by at least 3, or else the scale factor remains unchanged. We developed a more principled approach for deciding when to change the scale factors, so as to maintain the same total perceptual error.

Starting with the optimal set of multipliers, we are guaranteed that the perceptual error in each block is less than or equal to some P. Higher scale factors correspond to lower bitrates and higher perceptual error. Therefore, we lower a scale factor from the recommended value to that of the previous scale factor, if the coarser quantization does not save us at least 6 bits (the bits saved can be estimated on a block-by-block basis). We never raise the value of a multiplier, since this would increase the perceptual error above the target amount.

4. RESULTS AND ANALYSIS

Here we show some results of applying our adaptive coding algorithm. Figure 1 shows the orignal image. Figure 2 shows the results of applying Watson's non-adaptive coding method to find a Q which yields a perceptual error of P=1. Assuming that our visibility model is correct, at this P we should be "perceptually lossless" under the prescribed viewing conditions (mean luminance = 65 cd m^{-2}, 32 pixels/deg).

Figure 3 shows the matrix M resulting from our adaptive coding method, again aimed at P=1, where mid-gray represents a multiplier of 1, i.e. no change from non-adaptive coding, and in lighter blocks we more coarsely quantized in the hopes of acheiving a lower bitrate.

Figure 1: original 376x248 image

Figure 2: non-adaptive, P=1, 1.29 bpp

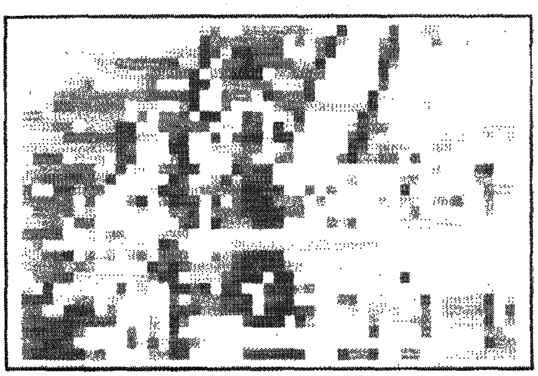

Figure 3: multipliers for Figure 4

Figure 4 shows the adaptively coded image. The bitrate for the non-adaptively coded image was 1.29 bpp, and for the adaptively coded image, 1.01 bpp, where 0.04 bpp were devoted to code the matrix M. This is a bitrate savings of roughly 22%. Note that no blocks have been more finely quantized than in the non-adaptively coded image; each block has either the same perceptual error as before, or greater. If our perceptual model is correct, these added errors will not be visible.

Figure 5 shows the results of non-adaptive coding using Watson's algorithm, at the same bitrate as our adaptive result. Highlighted regions in Figure 6 indicate blocks with P>1; blocks where we expect Figure 5 to have more visible error than the adaptively-coded Figure 4.

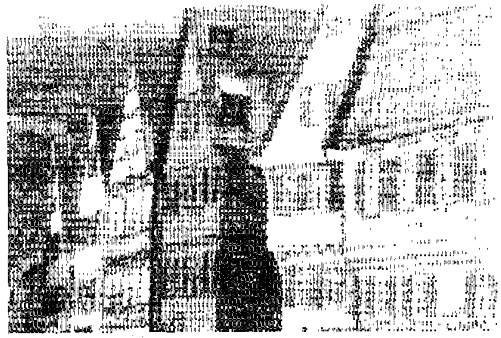

Figure 4: adaptive, P=1, 1.01 bpp

Figure 5: non-adaptive, 1.02 bpp

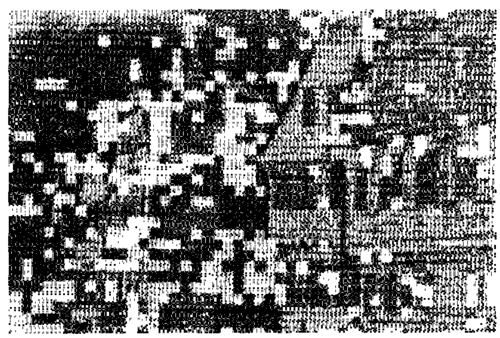

Figure 6: blocks in Figure 5 with $P > 1$

Many currently implemented solutions to finding the quantization scale factors are based upon the heuristic that quantization errors will be more visible in flat regions of roughly constant graylevel than in regions with "texture," and thus one should quantize more finely in flat regions. This resembles the contrast masking part of our perceptual error model. A number of measures of "textureness" have been suggested, and many of them resemble that of Test Model 5 (a suggested MPEG-2 encoder), which calculates the variance of each block and computes a normalized "activity" measure:

$$\begin{aligned} &\text{act}_k = 1 + \sigma_k^2 \\ &\text{norm. activity} = (2 \cdot \text{act}_k + \text{avg\_act}) / (\text{act}_k + 2 \cdot \text{avg\_act}) \end{aligned} \quad (7)$$

where σ_k^2 is the variance of block k, and avg_act is the average value of the local activity, act_k. The quantization multipliers are chosen according to this activity measure and "buffer fullness," which is a measure of how many bits have been spent in the image so far compared to the number which would have been spent if bits were parcelled out uniformly over the image.

Figure 7: multipliers, luminance masking only

It is worth comparing the results using this heuristic for the visibility of error to the results using our more complicated model which also incorporates luminance masking. This is particularly true because luminance masking seems to account for much of the appearance of our optimal M matrix; Figure 7 shows the M matrix resulting from our optimization procedure using only luminance masking (w_{ij} was set to 0), and it looks a great deal like the matrix in Figure 3. For comparison, Figure 8 shows the multiplier matrix resulting from applying the Test Model 5 activity-based adaptive coder (modified to allow a new multiplier in each block), at the same bitrate as the image in Figure 4. This matrix looks quite different. Figure 9 shows the resulting image, and highlighted blocks in Figure 10 are the blocks with $P > 1$.

CONCLUSIONS

We have demonstrated a method for perceptual optimization of the multiplier matrix, M, in JPEG adaptive coding. In the example shown, this method yields a 22% reduction in bitrate over non-adaptive coding at what is, according to our perceptual error model, the same image quality. The resulting multiplier matrix differs considerably from that resulting from an "activity"-based adaptive coder. If our perceptual model is correct, this raises questions about the common use of a "textureness" metric for adaptive coding, and suggests, in particular, the incorporation of luminance masking.

Acknowledgments

This work was funded by an NRC postdoctoral fellowship to Ruth Rosenholtz and NASA Life and Biomedical Sciences and Applications Division Grant 199-06-12-39.

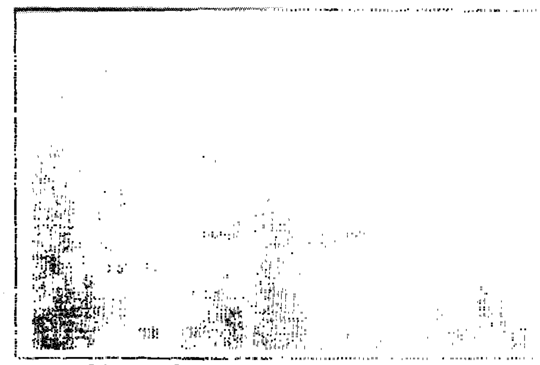

Figure 8: multipliers for Figure 9

Figure 9: activity-based adaptive, 1.02 bpp

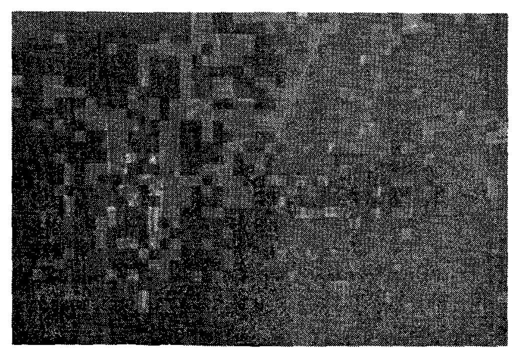

Figure 10: blocks in Figure 9 with $P > 1$

Bibliography

1. A. J. Ahumada Jr., A. B. Watson, & H. A. Peterson, "A visual detection model for DCT quantization," AIAA Computing in Aerospace, pp. 314-318, San Diego, CA, 1993
2. H. A. Peterson, A. J. Ahumada Jr., & A. B. Watson, "An improved detection model for DCT coefficient quantization," Proc. SPIE, 1913:191-201, 1993.
3. A. B. Watson, "DCT quantization matrices visually optimized for individual images," Proc. SPIE, 1913:202-16, 1993.