A Fast DCT Block Smoothing Algorithm

Rensheng Horng
Stanford University
horng@leland.stanford.edu
P.O.Box 9881, Stanford, CA 94309, USA

Albert J. Ahumada, Jr.
Human and Systems Technologies Branch, NASA Ames Research Center
al@vision.arc.nasa.gov
NASA Ames Research Center, MS 262-2, Moffett Field CA 94035, U.S.A.

ABSTRACT

Image compression based on quantizing the image in the discrete cosine transform (DCT) domain can generate blocky artifacts in the output image. It is possible to reduce these artifacts and RMS error by adjusting measures of block edginess and image roughness, while restricting the DCT coefficient values to values that would have been quantized to those of the compressed image. This paper presents a fast algorithm to replace our gradient search method for RMS error reduction and image smoothing after adjustment of DCT coefficient amplitude.

Keywords:
discrete cosine transform, image compression, artifact reduction, blocking artifacts, image smoothing.

INTRODUCTION

The discrete cosine transform (DCT) is currently used in the MPEG and JPEG standards,[1–3] and it also appears in proposed HDTV standards.[4] We have been developing algorithms for improving the quality of images that have been compressed by partitioning the image into blocks, converting each block to DCT coefficients, and then quantizing these coefficients.[5, 6]

When the inverse DCT is applied to the quantized coefficients, unpleasant blocky artifacts appear at the block boundaries. Our goal has been to reduce the blockiness while improving reconstruction accuracy. We have found a simplification of our previous methods that performs nearly as well as those methods and that can be computed much more rapidly and locally.

When errors are clearly visible, the blockiness of some artifacts distinguishes them from the original image content, suggesting a way of reducing these artifacts. Presumably, the human viewer identifies the artifacts by assuming the image does not have 8-pixel by 8-pixel block features. Block edge variance, the sum of squared differences between adjacent block edge pixels, is one simple measure of blockiness. A similar measure offset from the block edge provides an estimate of what this measure would be if the image were not blocky. In previous work,[5, 6] we found that using a global algorithm to lower the block edge variance to this estimate can reduce both apparent blockiness and RMS error. Figures 1 through 5 illustrate the performance of a new version of our algorithm for 5 64x64 pixel images having a range of image content.

The basic steps of the algorithm are similar to those we have proposed before.[5, 6]
1) Adjust the DCT coefficients to minimize RMS error.
2) Compute the block edge variance of the image and the estimate of what it should be.
3) Lower the block edge variance to the estimate.
4) Ensure that all DCT coefficients quantize to those of the compressed image.

In the rest of the paper, we make this description more precise, describe some quantitative results for these images, and relate our work to that of others. We conclude that if the quantization is strong enough to generate significant block artifacts, this faster method gives moderate de-blocking and a small decrease in the RMS image error.

THEORY

The DCT image transform

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

$$I_{u,v} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} i_{x,y}\ c_{x,u}\ c_{y,v}\ , \quad u,v = 0, N-1, \tag{1a}$$

where

$$c_{x,u} = \alpha_u \cos\left(\frac{\pi u}{2N}[2x+1]\right), \tag{1b}$$

and

$$\alpha_u = \begin{cases} \sqrt{1/N}\ , & u = 0 \\ \sqrt{2/N}\ , & u > 0 \end{cases}. \tag{1c}$$

DCT coefficient quantization

In JPEG quantization[1–3] a coefficient is quantized by the operation

$$S_{u,v} = \mathrm{Round}\left(\frac{I_{u,v}}{Q_{u,v}}\right). \tag{2}$$

The compressed image contains both the $S_{u,v}$ for all the blocks and the $Q_{u,v}$. To retrieve the image, first the DCT coefficients are restored (with their quantization error) by

$$\hat{I}_{u,v} = S_{u,v}\ Q_{u,v}\ , \tag{3}$$

where $Q_{u,v}$ denotes the quantizer step size used for coefficient $I_{u,v}$. The blocks of image pixels are reconstructed by the inverse transform:

$$\hat{i}_{x,y} = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \hat{I}_{u,v}\ c_{x,u}\ c_{y,v}\ , \tag{4}$$

which for this normalization is the same as the forward transform. Our goal is to find better estimates of these coefficients. A constant quantization matrix was used in the work reported here because the intended application was astronomical images, where the high spatial frequency information might be of higher value than usual.

DCT coefficient amplitude adjustment

The standard method of restoring the coefficient, using Equation 3, is equivalent to replacing each coefficient by the center of the quantization interval in which the original coefficient falls. The distribution of the non-DC coefficients for a given u,v peaks at zero and decreases monotonically. For quantization intervals not including zero, the distribution of the original coefficients is denser at the end of the interval closer to zero. The mean of the distribution is the minimum mean squared error reconstructor. For simplicity, we model the distribution of absolute amplitudes as exponential with mean μ. We estimate μ by the mean of the $|S_{u,v}|$. We then replace $S_{u,v}$ by

$$S_{u,v} - 0.5 + \mu - \frac{e^{-1/\mu}}{1-e^{-1/\mu}}\ , \quad \text{if } S_{u,v} > 0\ ,$$

$$S_{u,v}\,,\ \ \text{if } S_{u,v} = 0\,, \tag{5}$$

$$S_{u,v} + 0.5 - \mu + \frac{e^{-1/\mu}}{1 - e^{-1/\mu}}\,,\ \ \text{if } S_{u,v} < 0\,.$$

Figure 6 shows the RMS error improvement in dB, 20 times the logarithm to the base 10 of the ratio of the RMS difference between the original and the standard restoration to the RMS difference between the original and the amplitude adjusted restoration, for levels of quantization ranging from 5 to 100 in steps of 5. The image was that of Figure 1. A constant quantization matrix was used. For moderately high levels of quantization, the amplitude adjustment was not as effective. Comparing the predicted means of the interval distribution with the actual means, we find that Equation 5 overestimates the desired correction when the mean of the $|S_{u,v}|$ is a small fraction. This is probably caused by the poor fit of the exponential near zero, where the actual distribution is flat.

We originally developed the amplitude adjustment because we were afraid that the improvement of RMS error from smoothing might just be the result of the smoothing algorithm lowering the amplitudes of the coefficients to values closer to the mean. We have kept this step in our algorithm because the improvement in RMS error has almost always been increased by including it.

A measure of blockiness

Suppose i_1 and i_2 are the image values of two pixels that are next to each other in the same row or column, but are in different blocks. We assume that the blockiness of the compressed image is related to the fact that before compression, the values of i_1 and i_2 were usually similar, but they have been made more different by the quantization. We define the edge variance E to be sum of the squared differences for all such pixel pairs.

$$E = \sum (i_1 - i_2)^2, \tag{6}$$

The block edge variance E is our measure of image blockiness.

We estimate the desired value of the edge variance by computing the same measure for the pixels just inside the edge on either side and taking the average. If this estimate is less than the edge variance, we reduce the edge variance to this value.

A fast smoothing algorithm

Previously, we computed the block edge variance and its estimate over the entire image and then adjusted the DCT coefficients in the direction of the gradient of block edge variance.[5,6] This changes only the edge pixels, moving each value towards the average of the its value and that of the adjacent edge pixel. Following the method of Yang, Galatsanos, and Katsaggelos,[7] we now average the edge pixels to obtain a desired amount of block edge variance. Adjacent edge pixels i_1 and i_2 are replaced by i_1' and i_2', where

$$i_1' = \alpha\, i_1 + (1 - \alpha)\, i_2\,,$$
$$i_2' = \alpha\, i_2 + (1 - \alpha)\, i_1\,. \tag{7}$$

If E_d is the desired block edge variance and E_c is the current block edge variance, the appropriate α is given by

$$\alpha = \frac{1}{2} + \frac{1}{2}\sqrt{E_d / E_c}\,. \tag{8}$$

Smoothing is performed only when E_d is less than E_c.

The resulting smoothing is very similar to that of the gradient method, yet can be fast enough for real-time applications. In most cases, we find that using the estimate as the target value works better than using the block edge variance of the original image as was done by Yang, *et al.*[7]

Smoothing results

Section (d) in Figures 1 through 5 shows the RMS error improvement in dB as in Figure 6 for the full algorithm. quantization level, the ratio of the RMS error in the smoothed picture to that of the compressed image. It shows that the smoothing usually improves the accuracy of image restoration. The RMS error is slightly lower except when the quantization is very low and the next-to-edge estimate is also low. Fortunately, in this case, the image will only be slightly changed and there would be no apparent need for de-blocking. The improvement is better for images with smooth surfaces than for those with rough surfaces.

Since the smoothing only changes the pixels on the block boundaries, one would expect the success of this algorithm to be affected by the size of the the blocks. Figure 7 shows the improvement in RMS error as a function of quantization level for the image of Figure 1 with the block size as a parameter with values of 5x5, 8x8, and 11x11. The RMS error reduction was best at the block size of 8x8. Block sizes of 3x3 and 4x4 were also tested and gave results worse than those of 5x5.

Since the smoothing algorithm changes the edge pixels only, one can imagine that the inter-block smoothing reduces within-block smoothness. Previously we proposed checking the value of a with-in block smoothness measure and lowering that measure by within block smoothing if it was increased by the inter-block smoothing. Counting the number of times the routine was invoked, showed that within-block smoothing is rarely necessary. It only contributes to a small improvement of the RMS error and no noticeable visual improvement in the smoothing. We also checked to see how often the quantization constraint needs to be enforced. We find that the quantization constraint is very rarely enforced when the amplitude adjustment step is included. Out of all 5 images and levels of quantization, there were only two instances of quantization constraint enforcement in the image of Figure 1 at quantization level 5. We now propose that when speed is important, amplitude adjustment be done in the DCT domain, followed only by the rapid smoothing method in the image domain.

DISCUSSION

Our problem is a special case of the optimal decoding problem discussed by Wu and Gersho,[8] finding the image that minimizes the average value of a distortion function. They applied this strategy in the derivation of an optimal additive correction to a block for each possible level of each DCT coefficient .[9] Their (NLI) decoder gave a 0.7 dB improvement in mean square error on a diverse 23 image training set and about 0.5 dB improvement on new images. They report apparent reduction in blockiness, but since the method was restricted to within blocks, it does not directly attack the problem. Our amplitude adjustment is an application of this strategy to the single DCT coefficient amplitudes. The use of the exponential distribution model removes the dependence on amplitude and allows the effects of coefficient indices and quantization level to be represented by a single parameter easily estimated from the quantized data.

Our methods and results are similar to the iterative projection method of Yang, Galatsanos, and Katsaggelos[7], They also use edge variance and the quantization constraint. They compute separate horizontal and vertical edge variances and force them to their correct values in the original image by a weighted averaging of edge pixels. They iterate these two constraints in conjunction with the quantization constraint and range constraints in both the space and DCT domains. Since the constraints are projections onto convex sets, iterating them is guaranteed to terminate, since the original image is a solution. They report a 1 dB improvement in RMS error of reconstruction and strong apparent reduction in the blockiness for the 256×256 Lena image when the PSNR for the original reconstruction was 27.9 dB. Our method differs from their method mainly in the addition of the amplitude adjustment step (which seems to make the range constraints unimportant), and the estimation of the edge variance (which would be necessary if one begins with a JPEG compressed image).

SUMMARY

We have presented a method of estimating DCT coefficients from their quantized values and the quantization matrix, which are both included in the JPEG[2] standard compressed image file. This method of image reconstruction can reduce blockiness and RMS error in DCT quantized images. It includes a simple method of DCT coefficient amplitude adjustment that reduces the RMS error itself. Compared to our gradient method, the new method is about 15 times faster on a SPARC 2 when the image size is 512x512.

We conclude that if the quantization is strong enough to generate significant block artifacts, this method provides moderate to good de-blocking (by our informal personal evaluation), with RMS error reduction of up to 2 dB. Its intrinsic regularity and scalability allow this fast smoothing algorithm to be implemented easily on digital signal processors or in ASIC's.

ACKNOWLEDGMENTS

We appreciate the assistance of A. Watson and J. Solomon. This work was supported by NASA RTOP Nos. 199-06-31 and 505-64-53 and NASA Cooperative Agreement NAC 2-307 with Stanford University.

REFERENCES

1. G. Wallace, "The JPEG still picture compression standard", *Communications of the ACM*, vol. 34, no. 4, pp. 31-44, 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. P. J. Hearty, "Achieving and confirming optimum image quality," A. B. Watson, ed., *Digital Images and Human Vision*, MIT Press, Cambridge, MA, pp. 149-162, 1993.
5. A. J. Ahumada, Jr., R. Horng, "De-blocking DCT compressed images", B. Rogowitz and J. Allebach, eds., *Human Vision, Visual Processing, and Digital Display V*, Proceedings vol. 2179, SPIE, Bellingham, WA, pp. 109-116, 1994.
6. A. J. Ahumada, Jr., R. Horng, "Smoothing DCT compression artifacts", J. Morreale, ed., *Society for Information Display International Symposium V*, Digest of Technical Papers, vol. 25, pp. 708-711, 1994.
7. Y. Yang, N. Galatsanos, A. Katsaggelos, "Iterative Projection algorithms for removing the blocking artifacts of block-DCT compressed images", IEEE ISCASSP Proceedings, vol. V, pp. 405-408, 1993.
8. S. Wu, A. Gersho, "Enhanced video compression with standardized bit stream syntax", IEEE ISCASSP Proceedings, vol. I, pp. 103-106, 1993.
9. S. Wu, A. Gersho, "Enhancement of transform coding by nonlinear interpolation", Visual Communications and Image Processing '91: Visual Communication vol. 1605, SPIE, Bellingham, WA, pp. 487-498, 1991.