Preprocessing for Improved Performance in Image and Video Coding

V. Ralph Algazi Gary E. Ford Adel I. El-Fallah Robert R. Estes, Jr.

CIPIC, Center for Image Processing and Integrated Computing
University of California, Davis.

ABSTRACT

In previous work, we have reported on the benefits of noise reduction prior to coding of very high quality images. Perceptual transparency can be achieved with a significant improvement in compression as compared to error free codes. In this paper, we examine the benefits of pre-processing when the quality requirements are not very high, and perceptible distortion results. The use of data dependent anisotropic diffusion that maintains image structure, edges, and transitions in luminance or color is beneficial in controlling the spatial distribution of errors introduced by coding. Thus, the merit of pre-processing is for the control of coding errors. In this preliminary study, we only consider preprocessing prior to the use of the standard JPEG and MPEG coding techniques.

Keywords: image coding, video coding, image processing, inhomogeneous diffusion.

1 Introduction

In previous work, we have reported on the benefits of noise reduction prior to coding of very high quality images. Perceptual transparency can be achieved with a significant improvement in compression as compared to error free codes. In this paper, we examine the benefits of pre-processing when the quality requirements are not very high, and perceptible distortion results. Theoretical analysis, based on a weighted mean square distortion criterion, indicates that the benefit of noise removal no longer holds, as noise is removed in the discarding or quantizing of coefficients in the coding process. However, the use of data dependent anisotropic diffusion that maintains image structure, edges, and transitions in luminance or color is beneficial in controlling the spatial distribution of errors introduced by coding. Thus, the merit of pre-processing is no longer for mean square error reduction or SNR improvement, but for the control of coding errors. In this preliminary study, we only consider preprocessing prior to the use of the standard JPEG and MPEG coding techniques.

We study the pre-processing of images and video in two ways. First, by examining the effect of pre-processing on still image quality, as measured by a recently developed perceptually based picture quality scale (PQS). Second, we examine the effect of pre-processing on video coding performance, where noise has a more substantial effect, because the redundancy reduction sought by interframe coding is limited by the unpredictability of the noise from frame to frame. We evaluate the bit rates of encoded pre-processed and original video and compare for the same bit rate the quality of the resulting images either subjectively, using the PSNR, or objectively. We also examine briefly the benefits of post processing . We show an example of reduction of the blocking effect prevalent in DCT that is incorporated in the JPEG and MPEG standards.

2 Noise Reduction and Quantitative Quality Measure

2.1 Adaptive Noise Removal

The traditional rate-distortion results for a weighted mean-square distortion measure provide some insight in the methods for the encoding of information sources corrupted by additive noise. It can be shown that the optimum strategy in that case is to perform optimal (Wiener) filtering of the source and then to encode as if it were an noise free source.[1] If the noise is not removed, then the performance of the encoder will be degraded. The spectrum of the corrupted signal also indicates the importance of noise for the transform coding methods used in practice. At higher quality, some the additional transform coefficients retained in quantization are due to noise and increase substantially the bit rate required. Note that the common requirement of 46 dB of PSNR for an original image leads to a good visual appearance. But the additive noise for such a specification requires 2 bit/pixel to encode. At lower quality, quantization levels are set higher, and most of the noise, as well as part of the details in the image, are filtered out. Thus, filtering the image may be beneficial, but care has to be taken to maintain the important details within the image. However, it is well known that optimum Wiener filtering, or generalized Wiener filtering, such as is implemented in transform coders, results is significant image artifacts, not reflected in the improved SNR. This suggests an image dependent adaptive filtering approach, that selectively reduces the noise. We have already shown that such an approach has beneficial effects at high quality.[2] We wish to study here the benefits of noise removal and controlled image simplification on coding performance for lower quality image coding.

2.2 Mean Curvature Diffusion (MCD)

The basic algorithm for inhomogeneous diffusion, used in our approach to adaptive noise removal, is based on consideration of images as surfaces, and of noise removal as a regularization of the surface in a local neighborhood. Mean curvature diffusion is defined on the three-dimensional Euclidean space $\mathbf{E}^3$ and is interpreted geometrically. The image $I(x_1, x_2)$ is characterized as a surface $\mathcal{S}$ on $\mathbf{E}^2$.

$$\mathcal{S} \; : \; g(\mathbf{x}) = x_3 - A\, I(x_1, x_2) = B \tag{1}$$

where g is a differentiable real-valued function, A and B are real-valued constants, and $\mathbf{x} = (x_1, x_2, x_3)$ is the natural coordinate vector function of $\mathbf{E}^3$ defined such that for each point $\mathbf{p} = (p_1, p_2, p_3)$ in $\mathbf{E}^3$: $x_i(\mathbf{p}) = p_i$. The gradient vector field ∇g for the surface $\mathcal{S}$

$$\nabla g = \sum_{i=1}^{3} \frac{\partial g}{\partial x_i} \, \mathrm{U}_i \,, \tag{2}$$

where U_i is the unit vector field in the positive x_i direction, can be shown to be a non-vanishing *normal* vector field on the entire surface.[3] From (1) and (2) the magnitude of the surface normal can be expressed in terms of the image gradient

$$|\nabla g| = \sqrt{\mathrm{A}^2 \, |\nabla I|^2 + 1} \,. \tag{3}$$

The diffusion of g is modeled by

$$\frac{\partial g}{\partial t} = \nabla \bullet (C \, \nabla g) \,. \tag{4}$$

In our approach to inhomogeneous diffusion, the diffusion coefficient is the inverse of the surface gradient magnitude, i.e.,

$$C = \frac{1}{|\nabla g|}. \tag{5}$$

In terms of *unit normal vector field* on a neighborhood of $\mathbf{p}$

$$\mathcal{N} \triangleq \frac{\nabla g}{|\nabla g|}, \tag{6}$$

and we obtain

$$\frac{\partial g}{\partial t} = \nabla \bullet \mathcal{N} = \frac{\partial \mathcal{N}_1}{\partial x_1} + \frac{\partial \mathcal{N}_2}{\partial x_2} = 2\mathrm{H} \tag{7}$$

where the *mean curvature* H is the average value of the normal curvature in *any* two orthogonal directions. In this approach, diffusion occurs to decrease curvature. Noise produces random local curvatures that will be reduced. Image structures with zero curvature, such as edges, will be maintained. Note that the surface under this diffusion evolves at a rate twice the mean curvature of the image. Application of MCD to an isolated noisy edge[3] shows that the evolution of the surface results in surface area reduction (noise removal), arriving at a minimal surface at convergence (complete noise removal) with edge enhancement and an intact edge location.

2.3 Corner Preserving Filter (CPF)

The MCD algorithm preserves edges but rounds corners. By modifying the diffusion coefficient, however, we can force the inhomogenous diffusion algorithm to also preserve corners. The following diffusion coefficient meets these requirements.

$$C = \frac{1}{|\nabla g| \sqrt{1 + [(2\mathrm{H})\, |\nabla g| - (2\mathrm{H})]^2}} \tag{8}$$

At edges, H vanishes and the C reduces to its value in MCD. At corners, it reduces to $(|2\mathrm{H}|\, |\nabla g|^2)^{-1}$, which results in less diffusion than that performed along edges.[4]

3 PQS, A Picture Quality Scale

Research into the psychophysics of human visual perception has revealed that the human visual system (HVS) is not equally sensitive to various types of distortion in an image. This directly affects the perceived image quality. The Picture Quality Scale, PQS, that has been developed recently, is based on quantitative measures of several distortion factors.[5] Because these distortion factors are correlated, a principal component analysis is done to transform them into uncorrelated "sources of errors", and dominant sources are identified. These errors are then mapped to a PQS value by a model which was obtained from a linear regression analysis with the Mean Opinion Score (MOS).

3.1 Distortion Factors

The current version of the PQS includes five distortion factors of which the first two are derived from random errors and the last three from structural errors. Here we give only a description of these distortion factors. Formulas for computing the actual numerical measures are detailed in two references.[5,6]

Distortion Factor F_1 is a weighted difference between the original and the compressed images. The weighting function adopted is the CCIR television noise weighting standard. Here the viewing distance is assumed to be four times the picture height.

Distortion Factor F_2 is also a weighted difference between the original and the compressed images. The weighting function is from a model of the HVS. In addition, an indicator function is included to account for the perceptual threshold of visibility.

Distortion Factor F_3 reflects the end-of-block disturbances. The HVS is quite sensitive to linear features in images. In block coders, the error image contains discontinuities at the end of blocks, which explains blocking artifacts in the compressed image.

Distortion Factor F_4 accounts for general correlated errors. Textures with strong correlation are more perceptible than random patterns. Strong correlations in the error image suggest errors that are more apparent to human observers.

Distortion Factor F_5 is a measure of the large errors that occur for most coders in the vicinity of high contrast transitions (edges). Two psychophysical effects occur in the vicinity of high contrast edges. On the one hand, the visibility of noise decreases; this is referred to as "visual masking". On the other hand, the visibility of misalignments increases.

3.2 Principal Component Representation

Because the distortion factors $\{F_i\}_{1 \leq i \leq 5}$ are correlated, a principal component analysis is performed to decorrelate distortion measures and identify the dominant sources. This is done for a test set of distorted images obtained from representative coders which include transform, subband and DPCM coders for a range of quality scales. The transform matrix, $\{F_i\}$, that decorrelates C_F (shown in Table 1) is then computed using eigen-analysis. It was found that of the five eigenvalues of C_F the three largest ones accounts for 98% of the total error energy. Therefore, the three eigenvectors corresponding to the three largest eigenvalues can be chosen to transform $\{F_i\}$ into a principal component representation, $\{Z_i\}_{1 \leq i \leq 3}$.

3.3 Formation of the PQS

Since the various distortion factors collectively contribute to the overall perceived image quality, we seek a functional model mapping the distortion factors or measures to a single quality scale, the PQS. This model can be experimentally determined by studying the functional relationship between the distortion measures and the MOS, a five scale subjective ranking of image quality in terms of perceived distortions that is described in Table 2.[7] The simplest model is a linear one in which the PQS is expressed as a linear combination of uncorrelated principal distortion measures, $\{Z_i\}$, that is,

$$PQS = b_0 + \sum_{i=1}^{3} b_i Z_i \tag{9}$$

where $\{b_i\}_{0 \leq i \leq 3}$ are the partial regression coefficients obtained by multiple linear regression of $\{Z_i\}$ against the MOS.[5]

	F_1	F_2	F_3	F_4	F_5
F_1	1.00	0.97	0.95	0.03	0.97
F_2	0.97	1.00	0.99	0.15	0.91
F_3	0.95	0.99	1.00	0.17	0.88
F_4	0.03	0.15	0.17	1.00	0.11
F_5	0.97	0.91	0.88	0.11	1.00

Table 1: Covariance Matrix of F_i.

Grading Scale	Impairment
5	Imperceptible
4	Perceptible, not annoying
3	Slightly annoying
2	Annoying
1	Very Annoying

Table 2: The Scales of the MOS.

For a set of distorted images, the MOS values were obtained from an experiment involving nine observers under the conditions specified by the CCIR.[7] The observers were allowed to give half scale scores. A multiple linear regression analysis of $\{Z_i\}$ against the MOS gave $b_0 = 6.431$, $b_1 = -0.069$, $b_2 = -1.475$, $b_3 = -0.136$, with correlation coefficient $R = 0.88$.

4 Pre-processing of Still Images

We now discuss the methods and preliminary results if pre-processing is used prior to encoding. We first review briefly our previous results for high quality image coding, and then discuss strategies for higher compression and thus lower quality coding.

4.1 High Quality

At high image quality, also denoted perceptually transparent coding, the coding algorithm is designed such that the coding errors are not visible to a human observer. The transparent coding algorithms discussed in[2,8] are intended for application to high quality images having very high SNR. The compression ratio of such algorithms is improved by first filtering the image to remove non-visible noise and fine structure. This preprocessing filter must not induce any observable distortion.

Results have been obtained that compare error-free JPEG encoding and perceptually transparent coding techniques.[4,8] Results obtained from the latter indicates the benefits of CPF. For similar perceptual quality, CPF allows us to reduce the noise by more than an order of magnitude and, therefore, compress the image more effectively. The performance improvement for perceptually transparent coding , which uses pre-processing by inhomogeneous diffusion with an effective representation and coding strategy can be quite substantial. For the standard image Barbara, with a entropy of 7.35 bits/pixel, the best error free JPEG results in a bit rate of 5.58 bits/pixel, while the best perceptually transparent code yields a rate of 3.2 bit/pixel, a 43% improvement.

4.2 Pre-processing Strategies at Lower Quality

Image coding at lower quality levels means that the distortion due to the encoder has become quite perceptible. The purpose of pre-processing will then be to control the distribution of errors, so that they are less perceptible after coding, or to simplify the structure of the image so that fewer bits are necessary to encode it. We must now consider the following factors and their trade-offs.

1. The distortion introduced by pre-processing.
2. The distortion introduced by the encoder.
3. The bit rate of the encoder.

Since the pre-processing will introduce distortion, we have a situation in which distortion with respect to the original exists even before encoding. If that distortion is not perceptible, then we may be able to improve the performance of the encoding scheme by pre-processing for all quality levels. But if the distortion due to pre-processing is perceptible, then pre-processing will only become effective for some range of quality. In that range, the combined distortion due to pre-processing and encoding, is lower that the one due to encoding alone at the same bit rate.

4.2.1 Inhomogeneous Diffusion

Pre-processing by inhomogeneous diffusion is an effective way to smooth the image while maintaining important edge integrity. Since inhomogeneous diffusion makes use of the image gradient to determine the direction and amount of smoothing, we have also considered as a preliminary step, an additive decomposition that removes trends in flat portions of the image. Removing the trend may help in identifying residual noise in all smoothly varying portions of the image. This leads us to consider the following two strategies.

1. Preprocessing the original image.
2. Preprocessing the remainder of an additive decomposition of the original image.

4.2.2 Non Uniform Quantization

Although inhomogeneous diffusion is effective in removing noise that is not perceptible but which degrades performance at high quality, it is less effective a lower quality levels. This is because the encoding process, in the discarding of coefficients with successively larger amplitudes, performs a substantial removal of random noise by itself. What remains of the set of transform coefficients to be encoded, are larger coefficients that represent the active portions of the image, such as edges, or texture. We have considered, briefly, a new approach that quantizes the remainder of the image in a data dependent (non-uniform) manner. Although several adaptive quantization schemes simplify the structure of the remainder image, none of these schemes has helped, when combined with a DCT based coder that performs best for smoothly varying signals. Thus, the development of further image quantization schemes remains to be studied more extensively.

4.3 Examples

Using Lena as a test case in which to develop and evaluate our algorithms, we report encoding results (using JPEG) for the following cases:

1. No additional processing.
2. Noise reduction using anisotropic diffusion.
3. Additive decomposition using a bilinear approximation derived from an 8×8 grid, followed by inhomogeneous diffusion of the remainder. The remainder is encoded with JPEG, and the approximation is encoded using simple entropy encoding techniques.

The amount of noise removal is controlled by the number of times the iterative diffusion algorithms are applied to the images. We show, in Figure 1, the results obtained by preprocessing the original image, and in Figure 2, we show the same results when we pre-process the output of an additive decomposition of Lena. The additive decomposition uses as an approximation to the image a bilinear reconstruction from an 8×8 subsampled version of the original. The remainder, that contains all the details and structure of the original, is pre-processed with the CPF filter and JPEG encoded. The encoder reconstructs the approximation using error free DPCM , a JPEG option, at a cost of approximately 0.1 bit/pixel, and the JPEG encoded remainder image.

Discussion. The gains in performance achieved by such adaptive noise reduction are modest. The best results are achieved by using 20 iterations of the CPF inhomogeneous filter. This amount of processing distort substantially the image, so that at high quality or high bit rate, preprocessing is detrimental to the overall performance. Preprocessing begins to show some merit for a rate of about 1.4 bit/pixel, and an MOS or PQS quality of 3. If the original image is pre-processed, then the maximum gain due to preprocessing occurs at a bit rate of about 0.9 bit/pixel, and amounts to approximately 0.1 bit/pixel and 0.2 on the PQS scale. When an additive decomposition is used, then the maximum improvement (with respect to the unprocessed original) is about the same, except that it occurs near 1 bpp. These results indicate that preprocessing for JPEG still image coding may have merit, but that a more detailed study of preprocessing methods is necessary.

5 Preprocessing Video

In the encoding of video information, a major reduction in overall bit rate is possible because the successive video frames, representing motion in the scene, are highly redundant. Thus, in the MPEG coder, some frames are predicted from previous ones, while others are interpolated from previous and succeeding frames. After prediction, or interpolation, the residual frame difference images are the new information that must be represented and encoded. This interframe difference contains a contribution due to the motion and changes in the image, but also contains the independent noise contributions of original and predicted frames. Because of the subtraction process, the noise is increased, and represents an even larger fraction of the total residual signal.

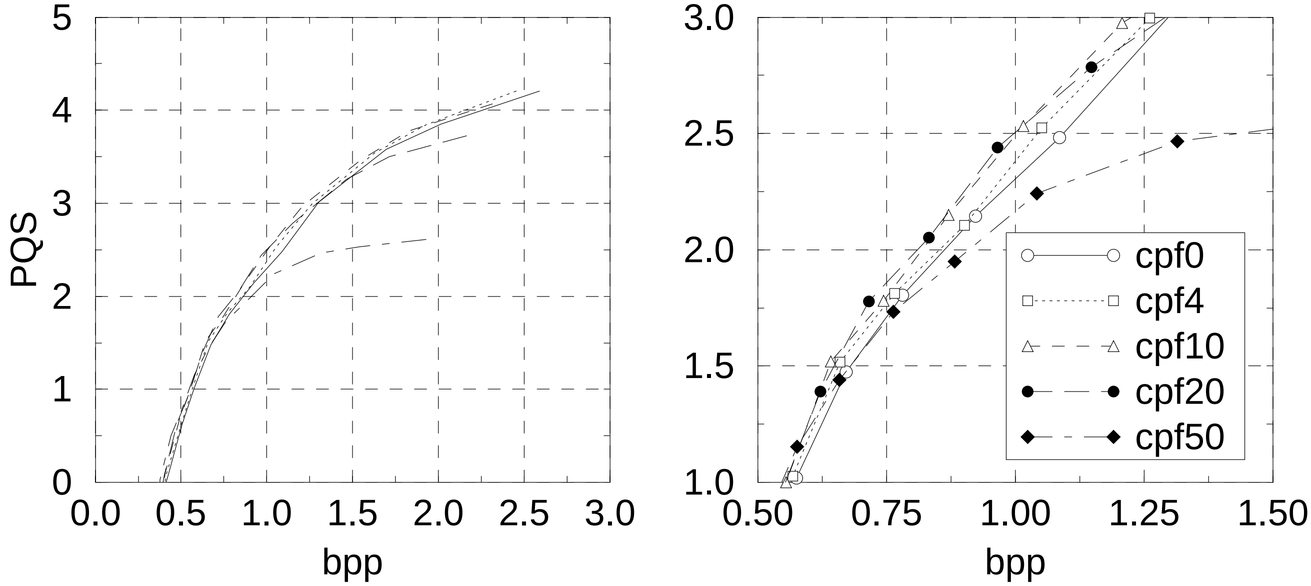

Figure 1: Benefit of preprocessing (Lena) before JPEG encoding.

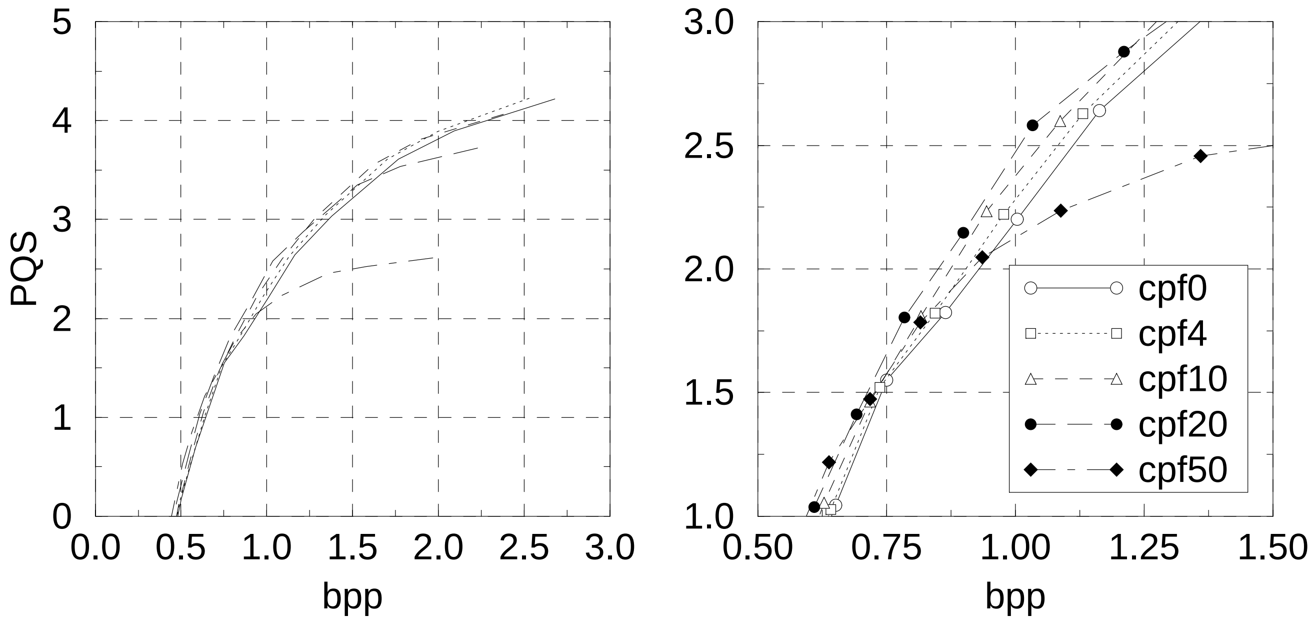

Figure 2: Benefit of preprocessing the remainder of an additive decomposition (of Lena) before JPEG encoding.

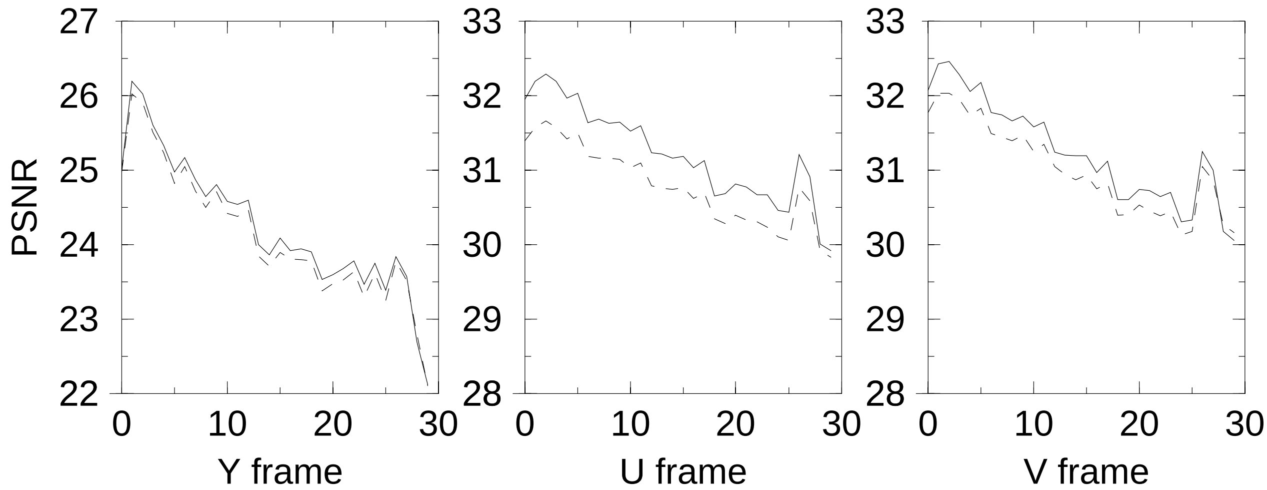

Figure 3: Benefits of preprocessing before MPEG encoding. These results are for the first 30 frames of the "mobile" sequence.

Two strategies, taken singly or in combination, can be used. The first one is to perform noise removal on the original signals, in the expectation that the interframe prediction difference will be easier to encode, because of the reduction in contributing noise. The second one is to perform a noise reduction on the interframe prediction error, and thus remove the noise at that stage to achieve encoding gain. Here again, the key issue is to develop a noise removal strategy that complements or reduces the effects of the coder on local features within each frame. We report here only on the results achieved by pre-processing the original frames in the sequence to be encoded by an MPEG2 coder. To compare performance, we compute the PSNR of the error between the original and encoded sequences, with and without preprocessing, for each of the YUV color components. Results are shown in the 3 graphs of Figure 3. The improvement in PSNR is approximately .2 dB for the luminance information and .5 dB for both the U and V chrominance components. The MPEG2 coder was designed so as to provide a nominal rate of 3 Mbps. Here again, the benefits of this frame based noise removal are limited. Further work on alternative methods for interframe processing is planned.

6 Improving Quality by Post Processing

Some of the more serious visual artifacts that are caused by block-based coders are not due to image noise or detailed image structures. They are due to the response at edges of the linear filters that are used in the signal representation. These errors produce visible end of block artifacts in flat portions of an image adjacent to edges. These artifacts are very visible, have a definite square grid pattern, but are of low contrast. The MCD adaptive filter strategy, that does not maintain corners, as contrasted to CPF, that does, is quite effective at reducing the sharp corners of these error patterns and making them much less visible. An example is shown in Figure 4. The MCD algorithm is very effective in reducing the the end of block effects. However, these effects occur at low quality levels where, after sufficient MCD iterations are used to remove blocking artifacts, the images become posterized and may not be acceptable for some specific applications either.

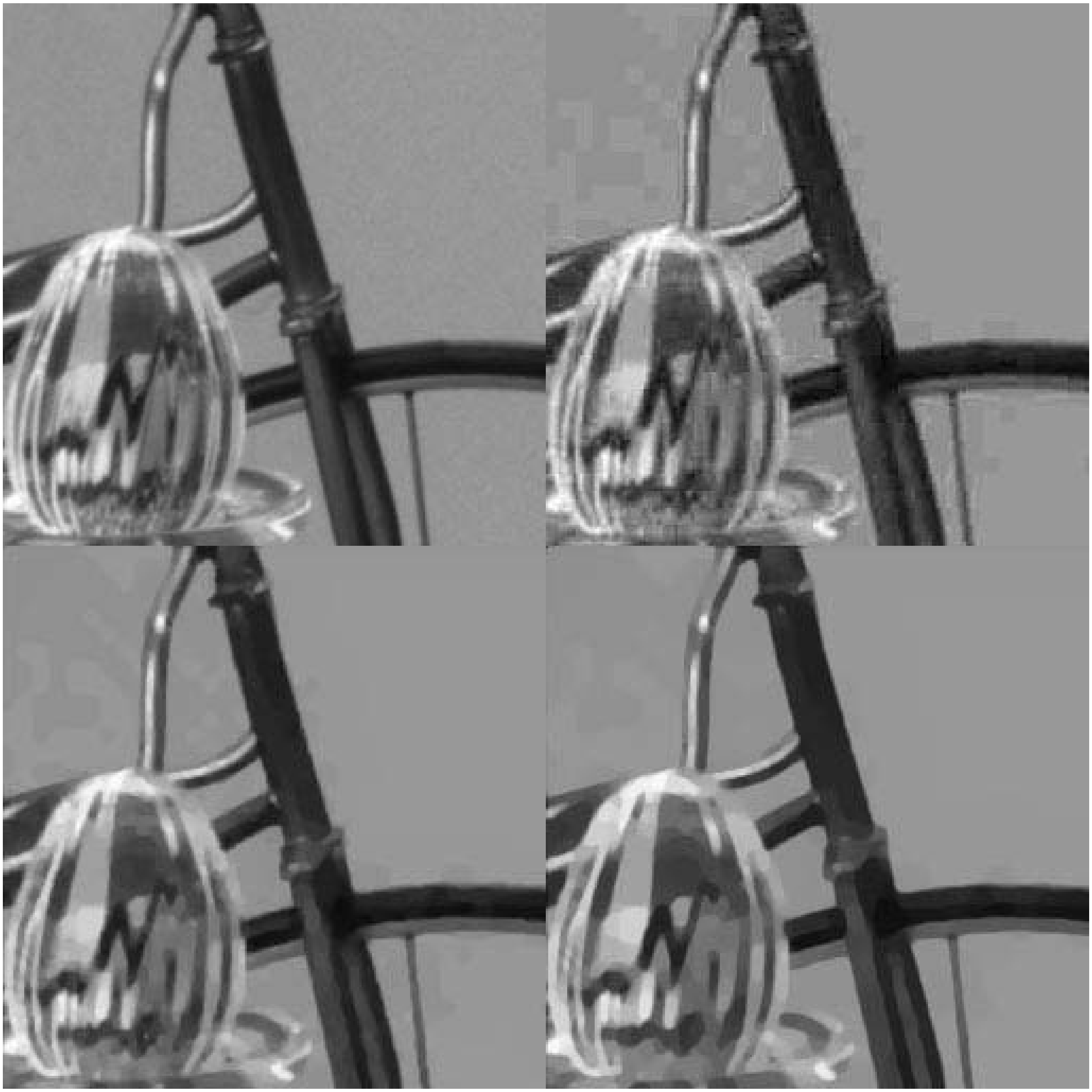

Figure 4: Effect of preprocessing on blocking artifacts. From top to bottom and left to right: the original, JPEG encoded, 40 and 100 iterations of MCD smoothing.

7 Discussion and Conclusions

- Basic rate distortion trade off a lower quality. As discussed previously, the preprocessing of image results in a lowering of the image quality, for high bit rates, but eventually proves beneficial . The methods that we have tried and reported are not very effective, but they form the basis for further systematic work.
- Improving the behavior at the knee of the curve. It interesting to note that when the PQS metric is used as a distortion or quality measure, the quality- rate curve has a distinct knee, which can also be stated as threshold phenomenon. Below a threshold rate, the quality degrades rapidly. Thus, the goal in improving coder performance by preprocessing is to extend the range of the quality-rate curve that is above the knee. That goal may also provide a formal framework for analysis and optimization of coder performance.
- Note we we have only examined the preprocessing of image or video to be encoded with the JPEG and MPEG standard codes. The philosophy of preprocessing to improve coding performance can be applied more broadly to any coding technique or to devise a new coding method. This is an important area for further research at high image quality. The postprocessing of coded images to improve quality is also very promising for the standard JPEG and MPEG codes because the errors introduced by the coders are highly structured, and can thus be identified and adaptively reduced.

8 ACKNOWLEDGMENTS

This research was supported in part by the UC MICRO program, Pacific Bell, Lockheed and Hewlett Packard.

9 REFERENCES

[1] V. R. Algazi, T. Reed, G. E. Ford, and R. R. Estes, "Image analysis for adaptive noise reduction in super high definition image coding," in *SPIE VCIP Proceedings*, vol. 1818, 1992.

[2] V. R. Algazi, T. R. Reed, G. E. Ford, E. Maurincomme, I. Hussain, and R. Potharlanka, "Perceptually transparent coding of still images," *IEICE Trans. Communications, Japan,*, vol. E75-B, no. 5, pp. 340–348, 1992.

[3] A. I. El-Fallah and G. E. Ford, "Nonlinear adaptive image filtering based on inhomogeneous diffusion and differential geometry," in *SPIE Proceedings*, vol. 2182, pp. 49–63, 1994.

[4] A. I. El-Fallah, G. E. Ford, V. Algazi, and R. R. E. Jr, "The invariance of edges and corners under mean curvature diffusions of images," in *Press, SPIE Proc. Electronic Imaging Symposium, San Jose Ca*, 1995.

[5] M. Miyahara, K. Kotani, and V. Algazi, "Objective picture quality scale (PQS) for image coding," in *Proc. SID Symposium for Image Display*, pp. 859–862, 1992.

[6] V. Algazi, Y. Kato, M. Miyahara, and K. Kotani, "Comparison of image coding techniques with a picture quality scale," in *Proceedings of SPIE, Applications of Digital Image Processing XV*, vol. 1771, 1992.

[7] "CCIR rec. 500-2, method for the subjective assessment of the quality of the television pictures," 1982.

[8] V. R. Algazi, G. E. Ford, R. R. Estes, A. El-Fallah, and A. Najmi, "Progressive perceptually transparent coder for very high quality images," in *SPIE Proceedings*, vol. 2298, pp. 13–24, 1994.