Progressive Perceptually Transparent Coding based on Image Analysis

V. R. Algazi G. E. Ford R. R. Estes Jr. A. El-Fallah

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

ABSTRACT: *For high quality applications, we have developed a coding technique that produces no visible degradation, that we denote as* perceptually transparent*. A compression significantly higher than error free codes is achieved. In a progressive scheme, a reduced resolution image is used for the transmission of high quality still images. To use the lower resolution images effectively, analysis based interpolation provides the estimate of the higher resolution image and reduces the incremental information transmitted. This approach may also lead to compressed images of higher quality that the original.*

1 INTRODUCTION

Most image coding techniques are directed at the efficient digital representation of original images of moderate quality. For very high quality and super high definition images, a more suitable goal is to allow coding errors, but only if they are imperceptible. We refer to such an approach as *perceptually transparent coding.*

In perceptually transparent coding, our primary goal is to control image quality. Secondarily, we wish to compress the image. Since quantization errors in transform domain techniques are distributed over the entire transform block, we restrict our attention to spatial domain processing, where we have sufficient control over the errors introduced.

In a differential quantization based approach, a good low frequency approximation of the image allows us to coarsely quantize the remainder and to exploit properties of the human visual system so that quantization errors are invisible in the reconstructed output image.

We further note that significant perceptual redundancy exists in images because of noise. As such, we preprocess the image using newly developed adaptive, anisotropic diffusion based techniques that removes this noise without introducing any perceptible artifacts.

Finally, we consider the error-free encoding of the non uniformly quantized remainder image. This image has characteristics similar to an edge map, with highly sparse structure. We propose to use analysis based techniques to encode it so that the encoding process is driven by (an analysis of) the data. In a pyramidal representation for the remainder image, higher resolution images are estimated from the lower resolution images using analysis based interpolation strategies. At each stage, only the error images must be transmitted.

2 DIFFERENTIAL QUANTIZATION

Visual artifacts encountered in coded images occur in the vicinity of high contrast edges, or near the transition between image regions, and are caused by inadequate control of the spatial distribution of errors. Artifacts with a spatial structure, such as contouring, or the end of block effect, are quite perceptible and highly objectionable. Differential quantization circumvents this problem by providing an excellent approximation in the flat portions of an image [4]. The approximation is not as good near edges or in active areas, but in these regions, visual masking, allows for substantial errors to occur below the visual threshold of perception. A diagram of the differential quantization approach is shown in Figure 1.

In this scheme, we compute a low frequency approximation to the original image, and then exploit the properties of this approximation by coarsely quantizing the difference between this approximation and the original image. That is, given an original $M \times N$ image, $I(m,n)$, we compute an approximation, $\hat{I}(m,n)$, which can be reconstructed from a smaller, subsampled image, $I_1(j,k)$.* We then coarsely quantize the difference $I_2(m,n) = I(m,n) - \hat{I}(m,n)$ to obtain the quantized remainder image, $\hat{I}_2(m,n)$, which we use as an alternate representation, $\{I_1, \hat{I}_2\}$, of the original image.

In this paper, we only consider a spline based low frequency approximation, computed from an 8×8 subsampled version of the original.

The subsampled array $\{I_1(j,k)\}$, from which we compute the spline approximation, $\hat{I}(m,n)$ is rep-

*In general, the approximation does not have to be in the form of an image, but typically, it is.

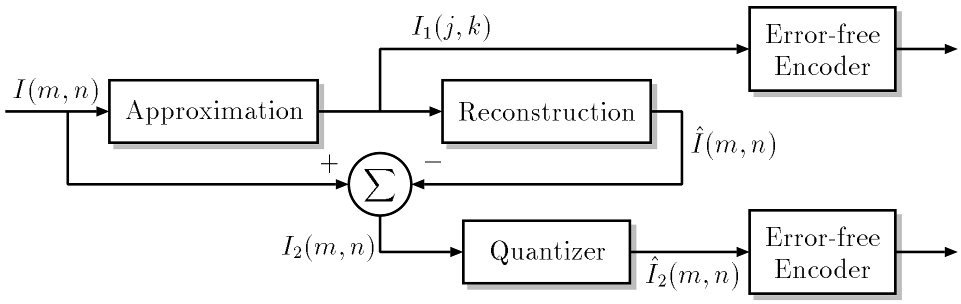

Figure 1: Differential Quantization

resented with 8-bit accuracy and is not further encoded. For an 8×8 subsampling grid, this adds 0.125 bits per pixel to the overall bit rate of the code.

2.1 Non uniform quantization and perceptual transparency

This additive decomposition results in a remainder image with a significantly lower variance than the original image. However, the number of quantization levels for the remainder remains high, but can be reduced.

Typically, for images, 8 bits of gray scale is used because at low luminance levels, the just noticeable difference (JND) in luminance is approximately 1/256. Further, any deviation in the mapping from the quantized signal to the brightness of the display will result in perceptible contouring in the low frequency subareas of the image if fewer quantization levels are used.

In our scheme , the spline approximation is best in the low frequency subareas of the images, making coarser quantization of the remainder now feasible.

Errors in images are substantially less visible in active portions of the image [11]. Since the remainder is the difference between a smooth approximation and the original, large values in the remainder correspond to the active portions of the image. At a viewing distance of 6 times picture height, visual masking will occur for a distance of up to six or seven pixels from a sharp transition. This suggests the use of a 8×8 subsampling grid, so that the maximum distance from the two dimensional grid of subsamples is less than six pixels.

Based on the above considerations, we devised a non uniform quantization scheme for the remainder. Such a scheme provides fine quantization for the low frequency subareas, where the remainder is small, and allows for coarse quantization in the active areas in the image.

We find that, for all images, quantization of the remainder to approximately 5 bits, or 32 levels, is sufficient to ensure perceptual transparency. Further work, based on a statistical analysis, results in a *universal* non uniform quantizer which provides perceptual transparency for *all* images. This non uniform quantizer has 31 levels in the range of -115 to $+115$ [3], and uses a uniform quantizer with step size 20 for larger errors. The theoretical maximum number of levels needed by this quantizer is then 45.

3 ADAPTIVE NOISE REDUCTION

Commonly used images are quite noisy. To quantify this statement, we have analyzed several images, including images from the Super High Definition (SHD) image test set provided by Nippon Telegraph and Telephone (NTT). Some results are shown in the first three columns of Table 1. No gamma correction was done.

The noise variance and the corresponding entropy are quite high, even for the SHD images. For additive Gaussian noise, at 45 dB PSNR, we predict an entropy of 2 bits/pixel [6], confirmed experimentally as shown. Such noise has a large effect on the performance of coders at high quality levels [6]. However, anisotropic diffusion based adaptive noise reduction techniques can substantially reduce this noise while maintaining the structured image details, important to image quality [2, 9].

3.1 Anisotropic diffusion

In adaptive noise reduction, an interative data dependent filtering algorithm is used. It can be shown that filtering with a family of Gaussian filter kernels $G(x,y,t)$ with variance parameter t, i.e.

$$I(x,y,t) = I(x,y) * G(x,y,t),$$

is equivalent to the partial differential diffusion equation

$$I_t = c\nabla^2 I = c(I_{xx} + I_{yy}),$$

where the subscripts denote partial derivatives, and ∇^2 is the Laplacian. In anisotropic diffusion, we allow the conduction coefficient, $C(x,y,t)$, to vary with respect to space and time, so that

$$I_t = C(x,y,t)\nabla^2 I + \nabla C \bullet \nabla I = \nabla \bullet \left[C(x,y,t)\nabla I\right],$$

where ∇ represents the gradient operation and $\nabla\bullet$, the divergence. Typically, $C = g(\nabla I)$, where g

image	before			after		
	MSE	PSNR	entropy	MSE	PSNR	entropy
bldg	0.705	49.65	1.869	0.078	59.23	0.241
daisys	8.702	38.73	3.559	0.436	51.74	0.959
lena	3.386	42.83	2.904	0.070	59.71	0.300
smile	0.163	56.00	1.037	0.020	65.18	0.018
wheel	1.231	47.23	2.255	0.058	60.47	0.262
wine	1.307	46.97	2.262	0.098	58.22	0.441

Table 1: Noise characteristics before and after 10 iterations of CPF based preprocessing.

is a nonlinear function to be specified. We have used adaptively scaled mean curvature diffusion (MCD) [7] by choosing

$$C = g(\nabla I) = \frac{1}{\sqrt{1 + A^2 |\nabla I|^2}},$$

where A is a scaling parameter (which is allowed to vary with time). As such, it can be shown [7] that the local rate of diffusion is equal to twice the mean curvature, H, of the image surface about each pixel.

This leads to a very effective, adaptive, iterative noise reduction technique. MCD preserves image structure, characterized by regions of consistently high gradients, and substantially reduces independent, random noise. It, however, also tends to round corners and other features characterized by higher order structure, such as edge intersections.

In more recent work [8], a diffusion coefficient,

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

is proposed which preserves corner structures much better. Using this filter, denoted the corner preserving filter (CPF), more iterations are allowed (yielding more noise reduction) while still maintaining perceptual transparency [8].

Ten CPF iterations results in more than 10 dB of noise reduction in the flat portions of the image (as can been seen in Table 1), while introducing no perceptible changes — as long as the PSNR of the original image is at least 46 dB, which implies that the noise is not perceptible. Note that, for noisy images, adaptive noise removal may actually *improve* the image quality.

4 ERROR FREE ENCODING OF THE QUANTIZED REMAINDER

In the proposed scheme, the non uniform quantization is as coarse as possible while still maintaining perceptual transparency. Additional image degradation is avoided when all subsequent coding of the quantized remainder is error free.

The quantized remainder image has just 45 levels, and we can encode it with any lossless encoding method available for gray scale images.

Since the prediction error is not guaranteed to fall on one of the allowed quantization levels, we enumerate the quantization levels and take differences between these indices to obtain an error-free encoding. Other techniques, such as error feedback [13], can also be used. Furthermore, the sums in the encoder and decoder can be taken modulo 45 (the number of levels in the non uniform quantizer) without affecting the error free nature of the code.

4.1 Interpolation based pyramids

In the progressive representation of the quantized remainder, the highest resolution image is estimated by interpolation from lower resolution subimages. The basic scheme is shown in Figure 2, where we show only 2 stages of the pyramid and haven't included the error free encoding that occurs between the encoder and decoder.

The lower resolution images supply non causal information about the signal and will lead to better predictions and lower bit rates than non analysis based techniques, such as DPCM.† We use bilinear interpolation and a directional interpolation strategy as the basis for the image pyramid [1].

The directional interpolation algorithm [1] (Figure 3) computes Robert's gradient angle estimates from the low resolution images in the pyramid and then interpolates these to the high resolution grid. From the high resolution angle estimates, a low frequency direction is determined that guides the interpolation algorithm. A measure of confidence in the gradient estimate is used to blend the results with those obtained with a more conservative, bilinear interpolation. As an alternative, we also considered an analysis based anisotropic diffusion interpolation algorithm.

Diffusion based interpolation starts with an initial approximation of the high resolution image, such as a bilinear interpolated version. The MCD algorithm is then applied while holding the subsampled pixel values fixed. By constraining the diffusion, we can obtain interpolated images which

†We will use DPCM to indicate our DPCM encoding technique for encoding the quantized remainder. We use JPEG-DPCM to refer to the standard error-free DPCM encoder used in the JPEG standard.

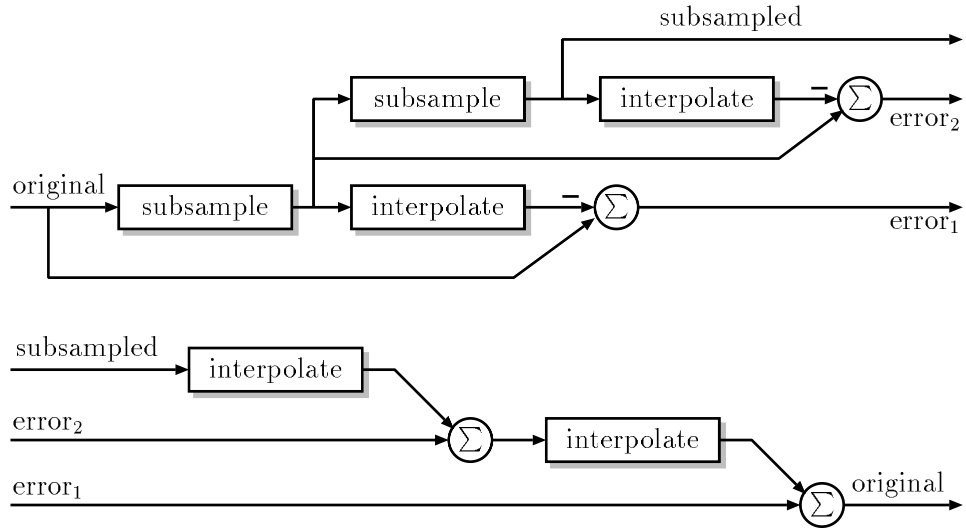

Figure 2: Two stage pyramidal code: encoder (top) and decoder (bottom).

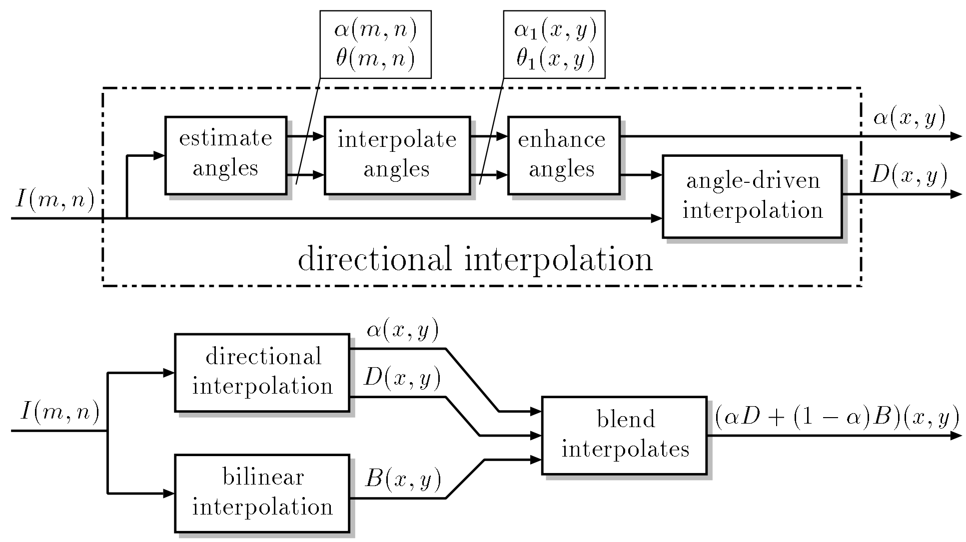

Figure 3: Overview of blended, directional interpolation strategy (top) and directional interpolation (bottom).

are good approximations of the original.

Typically, subsampling and interpolation are performed on a rectangular grid, in which case, subsampling yields a 4:1 reduction in the number of pixels in each stage of the pyramid. As an alternative, quincunx sampling only reduces the number of pixels by a factor of two at each stage, and we can cascade two stages of quincunx interpolation to get the equivalent of one stage of rectangular interpolation. The quincunx approach has advantages, since each pixel is encoded with respect to its four nearest neighbors at the *current* resolution, and the scheme is more regular than for rectangular interpolation. See our previous work [1] for more thorough discussion of the advantages of a quincunx pyramid.

4.2 Error free coding

For error free encoding, given the sometimes large percentage of zeros in these images, Huffman codes can be inefficient. Further, these zeros exhibit coherency, since the errors are localized in the active areas in the image. To exploit this coherency and make Huffman coding more efficient, we encode the position of the non zero pixels using a binary image encoder (the QM-code) with a seven pixel predictor, after which we encode the value of the non zero pixels with a Huffman code. This is a simple implementation of the color shrinking technique [5] for encoding sparse sources. This strategy results in an average 7% bit rate reduction for the analysis based pyramids (after noise preprocessing), but for the simpler images can be as high as 27%. The same numbers for the original (non preprocessed images) are 1.5% and 10%.

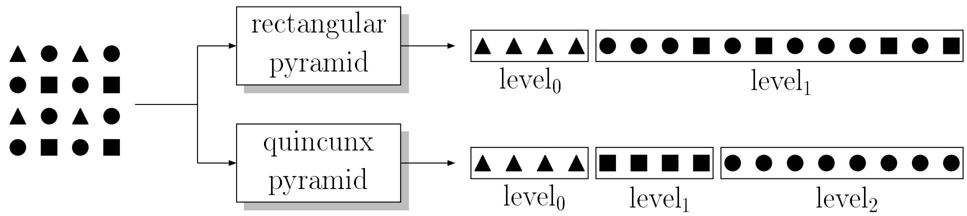

Figure 4: Quincunx sampling results in more symmetric neighborhoods, and better predictions than rectangular pyramids.

In addition, the 4096 subsamples from the original image transmitted by the differential quantization stage propagate through the the entire pyramid and, thus, the lowest resolution image in our pyramid consists entirely of zeros, and need not be encoded.

5 RESULTS

We summarize the results we have obtained with our perceptually transparent coding scheme, with and without adaptive noise removal. For each image, the remainder was encoded using DPCM and the pyramid based interpolation methods described in Section 4.1. The test set consists of the nine 512×512, 8 bit images.

The results are shown in Table 2. The first column, labeled JPEG, is the bit rate for the baseline JPEG-DPCM coder, where for each image we used the prediction kernel that gave the best results. The second column, labeled DPCM, is for perceptually transparent coding, using additive decomposition and coarse quantization of the remainder. The output of the DPCM predictor is one of the indices of the non linear quantization table. Since zeros are prevalent, color shrinking is used to encode efficiently the binary activity, zero or non zero output. A Huffman code used on the non zero source now gives results very close to the entropy.

The third column lists the best results obtained by a detailed study of all the interpolation schemes, which are rectangular, quincunx or diffusion based. We find that the directional quincunx interpolation yields consistently the best results. In particular, they are about 5% better than the results obtained with diffusion based interpolation.

In comparing the analysis and non analysis based pyramid encoding techniques, we find that we do obtain moderately better results using analysis, with the quincunx pyramid giving a 2% bit rate reduction with respect to the bilinear quincunx result. Quincunx sampling results are 5-6% better than the corresponding rectangular results. Finally, the analysis based rectangular pyramids are slightly better than bilinear interpolation.

The key to improving the performance, by analysis based interpolation, is an accurate gradient estimation and optimizing our interpolation rules, given this information. Alternatives have been presented [12, 14], which will be evaluated in future studies.

Note that with a good directional interpolation strategy, an adaptive quantization could be used while still maintaining perceptual transparency, since the errors are more localized near edges and other areas where masking is significant.

In Table 2, we show the effect of adaptive noise reduction. For the DPCM scheme, the noise reduction leads to a 14% decrease in bit rate, averaged over all images. For the quincunx pyramid it results in a 13% reduction. The gain of perceptually transparent coding, over JPEG-DPCM is slightly over 50%, i.e., we have compressed the images by an additional factor of two such that the error introduced are not visible.

6 DISCUSSION

This paper makes two major contributions. The first one is conceptual. It is possible to process and represent images so as to improve their compressibility without loss of image quality. In particular, adaptive noise reduction leads to a substantial increase in compressibility with no visible change in the image. The second contribution is that a carefully designed progressive code, when used in conjunction with image analysis, results in a hierarchical code which is as efficient as good full resolution, error free DPCM coder. Thus analysis leads to an effective progressive scheme, in contrast to the slight loss of performance generally associated with them.

Note that our approach to perceptually transparent coding, which controls image quality by first introducing imperceptible changes in the image, now requires increased attention to efficient error free coding schemes for quantized gray scale and color images.

7 ACKNOWLEDGMENTS

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

References

[1] V. R. Algazi, G. E. Ford, R. R. Estes, A. El-

Image	original			10 CPF iterations		
	JPEG	DPCM	directional	JPEG	DPCM	directional
baboon	6.572	4.262	4.251	6.470	4.122	4.114
bldg	4.233	2.603	2.575	3.782	2.310	2.310
daisys	5.453	3.336	3.304	5.143	3.038	3.026
flowers	3.524	1.935	1.939	3.246	1.827	1.831
lena	4.704	2.806	2.719	3.777	2.066	2.058
lynda	3.655	1.858	1.794	2.694	1.427	1.417
smile	2.909	1.439	1.442	2.029	1.008	1.036
wheel	3.640	2.037	2.049	2.623	1.453	1.496
wine	3.977	2.155	2.130	3.700	2.014	1.992
average	4.296	2.492	2.467	3.718	2.141	2.142

Table 2: Advantages of preprocessing and perceptually transparent coding. Note: the directional results are for a quincunx pyramid.

Fallah, and A. Najmi. Progressive perceptually transparent coder for very high quality images. In *Proceedings of the SPIE Symposium on Applications of Digital Image Processing XVII*, 1994.

[2] V. R. Algazi, G. E. Ford, M. Mow, and A. Najmi. Design of subband coders for high quality images, based on perceptual criteria. In *Proceeding of the SPIE Symposium on Applications of Digital Image Processing XVI*, volume 2028, pages 40–6, July 1993.

[3] V. R. Algazi, E. Maurincomme, and G. E. Ford. Error free and transparent coding of images using approximations by splines. In *Proceedings of the SPIE Symposium on Image Processing Algorithms and Techniques II*, volume 1452, pages 364–70, Boston, MA, November 1991.

[4] V. R. Algazi, T. R. Reed, G. E. Ford, E. Maurincomme, I. Hussain, and R. Potharlanka. Perceptually transparent coding of still images. *IEICE*, E75-B(5):340–8, March 1992.

[5] V. Ralph Algazi, Phillip L. Kelly, and Robert R. Estes. Compression of binary facsimile images by preprocessing and color shrinking. *IEEE Transactions on Communications*, 38(9):1592–8, September 1990.

[6] V. Ralph Algazi, T. R. Reed, Gary E. Ford, and R. R. Estes. Image analysis for adaptive noise reduction in super high definition image coding. In *Proceedings of the SPIE Symposium on Visual Communications and Image Processing*, volume 1818, Boston, MA, November 1992.

[7] A. I. El-Fallah and G. E. Ford. Mean curvature evolution and surface area scaling in image filtering. Submitted to the IEEE Transactions on Image Processing, June 1994.

[8] A. I. El-Fallah, G. E. Ford, V. R. Algazi, and R. R. Estes, Jr. The invariance of edges and corners under mean curvature diffusions of images. In *Proceedings of the SPIE, Image and Video Processing III*, San Jose, CA, February 1995.

[9] Adel El-Fallah and Gary E. Ford. Image filtering by gradient inverse inhomogeneous diffusion. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume V, pages V73–6, April 1993.

[10] T. A. Hentea and V. R. Algazi. Perceptual models and filtering of high–contrast achromatic images. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-14(2):230–45, 1984.

[11] A. Netravali and B. Haskell. *Digital Pictures.* Plenum Press, 1988. Chapter 4.

[12] John A. Robinson. Compression of super-high-definition multimedia images. In *Proceedings of the SPIE, Very High Resolution and Quality Imaging Conference*, volume 2663, San Jose, California, February 1996.

[13] K. M. Uz, M. Vetterli, and D. LeGall. Interpolative multiresolution coding of advanced television with compatible subchannels. *IEEE Transactions on CAS for Video Technology*, 1(1):88–99, March 1991. Special Issue on Signal Processing for Advanced Television.

[14] Xiaolin Wu, Nasir Memom, and Khalid Saywood. A context-based adaptive lossless/nearly-lossless coding scheme for continuous tone images. Submitted to ISO as a candidate algorithm for the next international standard for lossless compression of continuous-tone images, 1995.