IMAGE COMPRESSION USING SPATIAL PREDICTION

Ephraim Feig and Heidi Peterson

IBM Research
Yorktown Heights, NY

Viresh Ratnakar

Computer Sciences Department
University of Wisconsin-Madison, WI

ABSTRACT

This paper describes a new image compression technique, referred to as *spatial predicition*. Spatial prediction works in a manner similar to fractal-based image compression techniques, and is in fact a result of several experiments that we conducted to gain a better understanding of why fractal compression works. Spatial prediction compresses an image by storing, for each image block, either the quantized Discrete Cosine Transform (DCT) coefficients or the parameters of an affine transformation that constructs the block using another image block from the already encoded portion of the image. This technique does not require contractivity in the affine transformations and performs as well as or better than fractal compression. Spatial prediction does not out-perform pure DCT-based techniques (such as JPEG) in terms of PSNR/bit-rate tradeoff. However, at very low bit rates it results in far fewer blocky artifacts and markedly better visual quality.

1. INTRODUCTION

Probably the simplest fractal compression technique (from a conceptual point of view) is due to [1]. Its performance is almost as good as any of the more complicated variations. The underlying image model is that some parts of an image look very much like other parts of the image. In the actual fractal compression method, parts of an image are matched with other parts which have been shrunk and then perhaps acted upon by the dihedral group of rigid motions on the plane. The shrinking ensures contractivity of these matching transformations and is a mathematical technicality required *only* for the application of the fixed point theorem. That is, it is not necessitated by the image model itself.

We conducted experiments with the goal of developing an image compression scheme that exploits self-similarity but relaxes the contractivity requirement of the transformations used. The fixed-point theorem cannot be used for decompression, without contractivity. We overcame this obstacle by constraining the "predictor" for each block to be chosen from the region of the image to the left of or above that block. We encoded the first block of the image (in raster scan order) using quantized DCT coefficients. In addition, DCT was used to encode all those blocks for which no good predictor could be found. We describe the details in the next section.

2. SPATIAL PREDICTION (SP)

Let I be a grayscale image with height $H = h \cdot N$ and width $W = w \cdot N$. We segment the image into $h \cdot w$ blocks of size $N \times N$. Let $B_{i,j}$ denote the image block with its top-left corner located at (i,j). Let $b_{i,j}(k)$ denote the pixel values in $B_{i,j}$, for $1 \leq k \leq N^2$. Let $E(I)$ and $E(B_{i,j})$ denote the compressed image and the compressed blocks, respectively. Let $\tilde{I}$, $\tilde{B}_{i,j}$, and $\tilde{b}_{i,j}(k)$ denote the resulting (decompressed) approximations of I, $B_{i,j}$, and $b_{i,j}(k)$, respectively. The Spatial Prediction (SP) compression approach encodes image blocks either by using a block from the previously encoded area as a predictor, or by storing quantized DCT coefficients. The choice between the two options is made by comparing the squared error in the best prediction with a threshold $T_{i,j}$, which can be set in different ways which we discuss after presenting the basic outline of SP. Let $D(i,j)$ denote the set of locations of possible predictors for $B_{i,j}$. That is,

$$D(i,j) = \left\{ (i',j') \;\middle|\; \begin{array}{c} 1 \leq i' \leq i-N,\ 1 \leq j' \leq W-N \\ \text{or} \\ i-N+1 \leq i' \leq i,\ 1 \leq j' \leq j-N \end{array} \right\}.$$

Each possible predictor $\tilde{B}_{i',j'}$ is matched with $B_{i,j}$ in eight ways. Let the eight actions of the dihedral group acting on the plane be numbered $1, 2, \ldots, 8$, and let $\phi_1, \phi_2, \ldots, \phi_8$ denote the corresponding permutations on the indices $1 \ldots N^2$ of an $N \times N$ block. Then, *matching* $B_{i,j}$ with $\tilde{B}_{i',j'}$ according to the l^{th} group action

0-7803-2431-5/95 $4.00 © 1995 IEEE

means finding scalar constants s and o that minimize

$$\Delta = \sum_{k=1}^{N^2} \left(s\tilde{b}_{i',j'}(\phi_l(k)) + o - b_{i,j}(k)\right)^2.$$

This least-squares problem is easily solved, and we simply denote the optimal triple (Δ, s, o) by

$$\text{BestMatch}(B_{i,j}, \tilde{B}_{i',j'}, l).$$

Note that s and o must be quantized to a fixed number of bits. We defer this detail to a later discussion.

The following pseudo-code describes SP:

```
For i := 1, (h - 1)N + 1, N
  For j := 1, (w - 1)N + 1, N
    Δ* := ∞
    For each (i', j') ∈ D(i, j)
      For l := 1, 8, 1
        (Δ, s, o) := BestMatch(B_{i,j}, B̃_{i',j'}, l)
        If Δ < Δ* then
          Δ* := Δ
          (i*, j*, l*, s*, o*) := (i', j', l, s, o)
    If Δ* < T_{i,j} then
      {* Prediction *}
      E(B_{i,j}) := (i*, j*, l*, s*, o*)
    Else
      {* Intra-coding *}
      E(B_{i,j}) := DCT-Encode(B_{i,j})
    {* Reconstruct, for future predictions *}
    B̃_{i,j} := Decode(E(B_{i,j}))
```

The procedure DCT-Encode used above does runlength encoding of quantized DCT coefficients, in a JPEG-like fashion [3]. The quantization table used is the one that appears in the MPEG standard draft document [2]. The prediction parameters $(i^*, j^*, l^*, s^*, o^*)$ are not entropy-coded for simplicity. The encoding $E(I)$ of the image consists of the sequence of encoded blocks $E(B_{i,j})$ and one additional bit per block to identify whether prediction was used or intra-coding was done.

2.1. Variations in SP Parameters

The simplest choice of the threshold $T_{i,j}$ that we used was, $T_{i,j} = V_{i,j}$, where,

$$V_{i,j} = \sum_{k=1}^{N^2} \left(b_{i,j}(k) - \bar{b}_{i,j}\right)^2,$$

$\bar{b}_{i,j}$ being the average pixel value in $B_{i,j}$. Thus, $V_{i,j}$ is the squared error that would result if $B_{i,j}$ is intra-coded with every DCT coefficient except the DC being quantized to zero. We also experimented with setting $T_{i,j} = Q_{i,j}$, the quantization error resulting from the use of a DCT quantization table that gives a bit-rate (for the entire image) same as that of the predicted blocks.

For the prediction parameters $(i^*, j^*, l^*, s^*, o^*)$, we used several different bit-allocations. For 512×512 images, exhaustive search for predictors is very slow and necessitates 9 bits each for storing i^* and j^*. We also tried searching over a grid decimated by a factor of eight in each spatial dimension, which reduced the search time and required only 6 bits each for storing i^* and j^*. We always used 3 bits to encode l^*. With s^* and o^*, again we experimented with several settings, with a typical choice being 10 bits for o^* and 9 bits for s^*. When calculating BestMatch$(B_{i,j}, \tilde{B}_{i',j'}, l)$, the best value of s is first found and quantized, then o is calculated to minimize the error with the quantized value of s, and then o is quantized.

We will present performance results for three specific parameter settings, in Section 4.

3. SP WITH DCT CORRECTION (SP+DCT)

SP usually results in very low bit rates. The SP+DCT compression technique is an extension of the SP technique, which essentially encodes the residual error using DCT, in a JPEG-like fashion. We used SP+DCT to compare the performance of spatial prediction with the hybrid fractal compression scheme in [4]. The residual error $(I - \tilde{I})$ is compressed at different rates using different DCT quantization tables. These tables are chosen to maximize PSNR for the resulting bit-rates, using the optimization algorithm in [5]. The encoding of quantized DCT coefficients is done like in JPEG, except that the DC coefficients are not DPCM-coded.

4. PERFORMANCE RESULTS

We present performance results for three variants of SP and SP+DCT in this section. The test image used

Name	$T_{i,j}$	Bits for i^*, j^*	B
SP6V	$V_{i,j}$	6, 6	34
SP9V	$V_{i,j}$	9, 9	40
SP9Q	$Q_{i,j}$	9, 9	40

Figure 1: The three variants of SP used

was the well-known 512 x 512 grayscale image of Lena. The three variants are characterized by the different choices made for $T_{i,j}$ and the number of bits assigned for storing i^* and j^*. These choices, and the resulting total number B of bits used per predicted block, are listed in Figure 1. The parameters common to all three are: block size (N) = 16, 10 bits for o^*, 9 bits for s^*, and 3 bits for l^*.

We compared the PSNR/bit-rate tradeoffs for these three variants of SP, fractal compression, fractal compression with DCT correction, and JPEG. The results

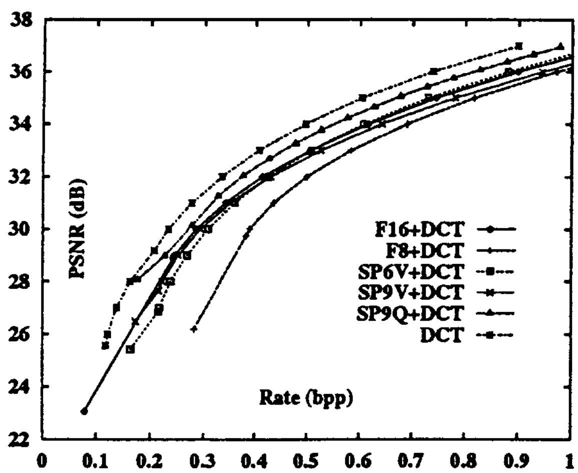

Figure 2: PSNR/bit-rate tradeoffs

are shown in Figure 2, as PSNR-Rate plots for each approach. The FN+DCT curves are for fractal compression with $2N \times 2N$ domain blocks mapped onto $N \times N$ range blocks, with the residual error coded in a JPEG-like fashion, again using quantization tables that maximize PSNR at different bit-rates. The leftmost points on each of the two FN+DCT curves correspond to no coding of the residual, that is, FN, the fractal method of [1]. Similarly, the leftmost points on each SP+DCT curve correspond to no residual coding. The curve marked DCT is for JPEG, i.e., just DCT-based compression.

5. DISCUSSION

Our results indicate that SP6V+DCT and SP9V+DCT techniques perform as well as F16+DCT compression, and better than F8+DCT compression, in terms of both PSNR and visual quality of the reconstructed images. SP9Q+DCT compression performs marginally better than these, and is closer to DCT, which outperforms all other techniques, in terms of PSNR. However,

Figure 3: DCT at 0.17 bpp, with PSNR = 28.3 dB

Figure 4: SP9Q at 0.17 bpp, with PSNR = 28.0 dB

at the very low bit rate of 0.17 bits per pixel, visual quality resulting from SP9Q is markedly better than DCT, as can be seen from the reconstructed images in Figure 3 and Figure 4. The DCT reconstruction is very blocky, while the SP9Q reconstruction is not, even though the former has a slightly higher PSNR. The SP9Q reconstruction also has better visual quality than the SP6V and SP9V reconstructions. Recall that the threshold $T_{i,j}$ used in SP9Q is the error in coding $B_{i,j}$ using DCT with a quantization table that gives a rate of $40/16^2 = 0.156$ bits per pixel for the whole image. We chose this strategy in order to come closer to the DCT curve in terms of PSNR. The fact that this strategy gives better *visual* quality suggests efficient exploitation of actual self-similarity in the image. In addition all SP compression techniques give visually more appealing reconstructions compared to fractal compression. These results indicate that the critical factor in fractal image compression is the self-similarity property, not the contraction-mapping construction (the so-called Collage Theorem).

The reduction in resolution of the search for predictors from SP9V to SP6V causes slightly lower PSNR and lower visual quality. However, when residual error coding is added (SP9V+DCT and SP6V+DCT), reduced resolution has little effect on PSNR or visual quality. This is important because reducing the resolution of the prediction parameters substantially reduces the prediction search time.

6. REFERENCES

[1] Jacquin, A. Image Coding Based on a Fractal Theory of Iterated Contractive Image Transformations. *IEEE Transactions on Image Processing*, 1:18–30, 1992.

[2] MPEG I draft: Coding of Moving Pictures and associated audio for digital storage, 1991. Document ISO/IEC-CD-11172.

[3] Pennebaker, W. B. and Mitchell, J. L. *JPEG Still Image Data Compression Standard.* Van Nostrand Reinhold, New York, 1993.

[4] Ratnakar, V., Feig, E., and Tiwari, P. Fractal Based Hybrid Compression Schemes. *Proceedings of SPIE's VCIP'94*.

[5] Ratnakar, V. and Livny, M. RD-OPT: An Efficient Algorithm For Optimizing DCT Quantization Tables. *Technical Report 1257, Computer Sciences Department, University of Wisconsin-Madison*, 1994.