A NON UNIFORM SEGMENTATION OPTIMAL HYBRID FRACTAL/DCT IMAGE COMPRESSION ALGORITHM

Gerry Melnikov and Aggelos K. Katsaggelos

Northwestern University
Department of Electrical and Computer Engineering
Evanston, IL 60208
Email : {gerrym,aggk}@ece.nwu.edu

ABSTRACT

In this paper a hybrid fractal and Discrete Cosine Transform (DCT) coder is developed. Drawing on the ability of DCT to remove inter-pixel redundancies and on the ability of fractal transforms to capitalize on long-range correlations in the image, the hybrid coder performs an optimal, in the Rate-Distortion sense, bit allocation among coding parameters. An orthogonal basis framework is used within which an image segmentation and a hybrid block-based transform are selected jointly. A Lagrangian multiplier approach is used to optimize the hybrid parameters and the segmentation. Differential encoding of the DC coefficient is employed, with the scanning path based on a 3^{rd}-order Hilbert curve. Simulation results show a significant improvement in quality with respect to the JPEG standard.

1. INTRODUCTION

Fractal image coding takes advantage of image self similarities on different scales. Most fractal algorithms, beginning with Jacquin's implementation [4], operate on an image segmentation consisting of non-overlapping square regions, called ranges. Each range block r_i is encoded by a non-expansive transformation T_i operating on the whole image and mapping a domain block, d_i, twice the size of the range block, onto r_i. For each range block, the encoder seeks the transformation minimizing the collage error $\|r_i - T_i x\|$, where x is the original image. The collage theorem guarantees an upper bound on the reconstruction error as a function of the collage error and the contractivity of T_i.

The structure of each transformation is fixed and consists of a decimation operator, an isometry, a multiplication by a scalar, and addition of an intensity translation block. A specification of these parameters together with the domain block index constitutes a block code. Most of the fractal coders classify blocks into classes to reduce computation and to achieve some local adaptiveness by spending fewer bits in relatively uniform areas. At the decoder, an approximation to the original image is generated by iterative application of these blockwise transformations, starting with an arbitrary image.

Blocks for which a good approximation, under a contractive transformation, can be found elsewhere in the image, can be efficiently coded using a fractal transform. The self-similarity assumption which is central to fractals, however, may not be justified for all blocks. In these cases, spending more bits on the fractal transform, by employing more isometries or finer quantizers is not very efficient [2], [8].

The Discrete Cosine transform (DCT) has been the transform of choice for most codecs due to its decorrelation and energy compaction properties. Complicated image features, however, require a significant number of DCT coefficients for good fidelity. The coarse quantization of the DCT coefficients in this case results in blocking artifacts and unsharp edges.

The proposed approach combines fractal based coding with DCT coding and within the chosen algorithmic structure, arrives at an optimal code. DCT is used to encode low frequency coefficients at predetermined positions in a block. The rest of the coefficients are encoded through the fractal transform. The encoder chooses from several sets of predetermined coefficient positions, with each set producing a rate and distortion pair.

The image segmentation and the optimal hybrid code are found jointly. The Lagrangian multiplier method is used in Sec. 2.1 to convert the constrained problem of finding the code and segmentation which minimizes the distortion measure for a given bit budget, into an unconstrained minimization problem. Sec. 2.3 describes how fractal and DCT transforms are combined in the coder. The use of interblock predictive coding in the algorithm couples code selection and segmentation decisions among blocks. A dynamic programming first-order backward dependency scheme, described in Sec. 2.4, is used to arrive at the optimal scanning path of leaves in a quad-tree (QT) decomposition. The quantization of the parameters is addressed in Sec. 2.5. Finally, experimental results and conclusions are covered in Sec. 3.

2. HYBRID FRACTAL/DCT TRANSFORMATION

2.1. Problem Formulation

The problem at hand is to simultaneously segment an image into blocks of variable sizes and, for each, to find a code such that any other choice of segmentation and coding parameters would result in a greater distortion for the same rate. Mathematically, for a given image x, we want to solve the following optimization problem:

$$\min_{s\in S, c\in C_s} D(x_{s,c}, x) \quad subject\ to\ \ R(x_{s,c}) \leq R^*, \qquad (1)$$

where $x_{s,c}$ is the encoded image, D the distortion metric, s a member of the set of all possible image segmentations S, c a member of C_s, the set of all possible codes given segmentation s, R the bit rate associated with segmentation s and code c, and R^* the target bit budget. The distortion metric chosen here is the l_2-norm. The constrained optimization problem stated in Eq. (1) is converted into an unconstrained problem using the Lagrangian multiplier method, that is the following problem is solved

$$\min_{s\in S, c\in C_s} G(x_{s,c}) = \{D(x_{s,c}, x) + \lambda \cdot R(x_{s,c})\}. \qquad (2)$$

The multiplier $\lambda = \lambda^*$, with corresponding optimal segmentation s^* and code c^*, for which $R(x_{s^*,c^*}) \approx R^*$, can be efficiently found using the monotonicity of the operational rate-distortion (ORD) curve, as in [9]. Using this formulation, we overcome a disadvantage of conventional fractal coders which is their inability to provide good rate control when high fidelity is required, that is, the allocation of more bits per transformation, by allowing a greater domain pool or more isometries, after a certain point, does not lead to corresponding increases in reconstruction quality.

2.2. Segmentation

The set of all possible segmentations S is restricted to be on the quad-tree lattice as a compromise between local adaptivity and simplicity of description. For a 256×256 input image it is a 3-level quad-tree with maximum block size of 16×16 pixels and minimum size 4×4 pixels. At each level of the quad-tree only one bit is required to signal a splitting decision, with no such bit required at the lowest level.

2.3. Code Structure

Making components of the fractal transform orthogonal to each other carries many benefits [6]. Among these are fast convergence at the decoder, non-iterative determination of scaling and intensity translation parameters, and removal of the magnitude restriction on the scaling coefficient for convergence. Here we explore another such benefit, the local continuity of the DC value of the intensity translation term. Frequency domain interpretation lends itself naturally to the concepts of orthogonality and energy compactness, which is why we perform the collage error minimization in the DCT domain. In agreement with the terminology used in [7], let vector r_i, of size $M^2 \times 1$, represent the DCT coefficients of range block i, of size $M \times M$, scanned in the zigzag order. Similarly, vector d_i comes from the chosen domain block, after decimation, DCT, and the application of isometry operators. Let the intensity translation term be represented by a linear combination of $N-1$ fixed vectors f_{ik}, of size $M^2 \times 1$. Then range vector i is approximated by N orthogonal vectors as follows:

$$r_i \approx \beta_i \cdot d_{io} + \sum_{k=0}^{N-2} c_{ik} \cdot f_{ik}, \qquad (3)$$

where d_{io} is the projection of d_i onto the orthogonal complement of the subspace spanned by vectors f_{ik} for $k = 0, \ldots, N-2$. A non-iterative procedure for finding β_i and c_{ik} yielding a least-squares approximation of r_i was given in [7]. A similar approach was tried in [3] in which multiple orthogonalized domain vectors were allowed. DCT domain modeling of range vectors was used in [1], but that approach suffered from the lack of orthogonality of d_i to all fixed vectors f_{ik}, a uniform segmentation, a restriction placed on the scaling coefficient, the need to specify significant DCT coefficient positions, and an independent encoding of transform coefficients.

In our implementation, we allow a bank of fixed subspaces to be used to model a range vector of a given size. To illustrate how a fixed subspace is formed, let us, for simplicity, consider a 2×2 block of DCT coefficients. Let us also assume that the lower 3 of these coefficients are used to form a subspace of dimension 3, when the full space is of dimension 4. Each f_{ik} corresponds to one coefficient in the two-dimensional DCT of size $M \times M$, as shown in Fig. 1 for $M = 2$. Each vector f_{ik} has zeros in all positions, except the one corresponding to the order in which the DCT coefficient it represents was scanned, where it has 1. Generally, in larger blocks more DCT coefficients tend to be significant. Hence, the fixed space used for the coding of a 16×16 range block will be allowed to be of a higher dimension than that of a 4×4 block. Limiting the number of available subspaces has the advantage that only its index needs to be sent to the decoder, whereas in [1] individual coefficient positions had to be sent as well. Shown in Fig. 2 are the subspaces used for block sizes 4×4. The subspaces for block sizes 8×8 are generated similarly from the first 1, 3, and 4 diagonals, and for block sizes 16×16 from the first 1, 4, and 5 diagonals, respectively. Thus we allow

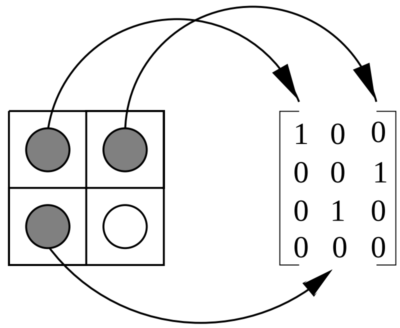

Figure 1: Mapping of selected DCT coefficients into basis vectors

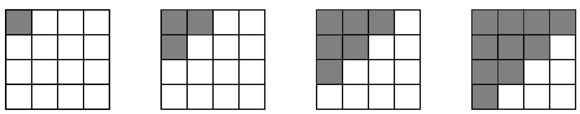

Figure 2: The 4 subspaces for block size of 4

energy carrying low-frequency DCT coefficients to provide the fixed subspace vectors f_{ik}. For each domain vector, only the component orthogonal to the subspace spanned by f_{ik} is used to approximate a given range block. Clearly, the subspace dimension uniquely identifies the basis vectors used.

2.4. Dynamic Programming Solution

The c_{i0} coefficients in Eq. (3) are encoded differentially, due to their correlation. Hilbert curve is known to satisfy certain adjacency requirements [9] and is efficient for predictive coding. We use the 3^{rd}-order Hilbert curve for the 256×256 input image.

The overall problem of finding the optimal segmentation and the hybrid fractal/DCT code is posed as that of finding the shortest path through the leaves of the quad-tree decomposition, with each leave having 1 to 3 possible codes, corresponding to 1 to 3 predecessors of a block in our Hilbert curve. If $g_{i,j} = d_i + \lambda \cdot r_{i,j}$ denotes the transition cost associated with encoding block i with block j as its predecessor, the optimal scanning path has the optimal substructure property, i.e., it consists of optimal segments. This motivates the use of dynamic programming (DP) to find it. Details on how (DP) can be used to find this optimal path can be found in [9].

2.5. Quantization of Parameters

The relative position of a domain block with respect to the range block was encoded using a variable length code (progressively more bits are used for farther distances). A 32×32 search window centered around the range block, with step size of 2 pixels was used to generate codebooks. Isometries were limited to identity, horizontal and vertical symmetries, and rotation by 180^o. The scaling parameter β_i was non-uniformly quantized using a Max-Lloyd quantizer with 5 bits. The DCT coefficients c_{ik} were encoded using JPEG's variable length codes, except that no zero-runs were used. The overhead information consists of: 1 or 2 bits to indicate what fixed subspace is used (depending on the block size), and 1 bit to indicate whether a fractally transformed domain vector d_{io} is used as the N^{th} basis vector or not.

3. EXPERIMENTAL RESULTS AND CONCLUSIONS

Figure 3 shows the performance of the hybrid fractal/DCT algorithm and JPEG under different bit budgets on a 256×256 Lena image. The hybrid algorithm is shown to outperform JPEG by 1.5-3.0 dB across the range of bit rates.

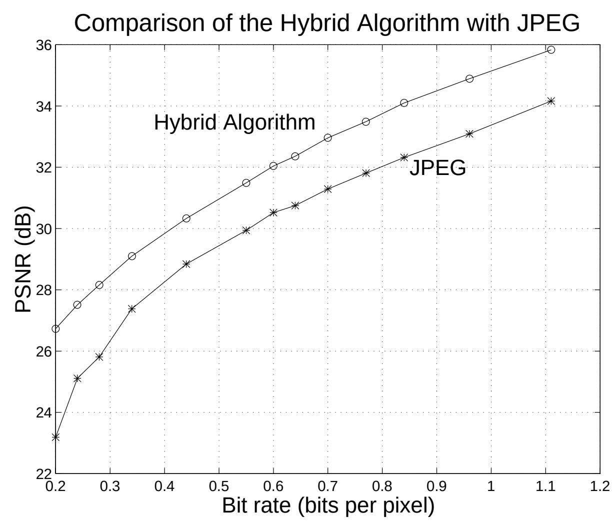

Figure 3: Hybrid Fractal/DCT algorithm vs JPEG

The images compressed by JPEG and the proposed algorithm at roughly the same bit rate (0.20bpp) are shown in Figs. 4 and 5, respectively. The hybrid algorithm reduces the blocking artifacts by modeling high frequency information through the fractal transform. Fig. 6 shows how the optimal segmentation adapts to the image by using larger size range blocks in relatively uniform areas. Overall, about 30% of the bits were spent on the fractal component of the transformation.

These results show how fractal and DCT transforms can be combined to yield an optimal segmentation and code. They also show that the fractal transform is efficient for representing high frequency information, and the DCT for representing low frequency information.

Figure 4: Hybrid algorithm (R=0.20bpp, PSNR=26.71 dB)

Figure 5: JPEG (R=0.20bpp, PSNR=23.19 dB)

4. REFERENCES

[1] K. Barthel, J. Schuttemeyer, T. Voye, P. Noll, A new image coding technique unifying fractal and transform coding. Proceedings of IEEE International Conf. on Image Processing, vol.3, pp.112-116, 1994.

[2] Y. Fisher, Fractal Image Compression - Theory and

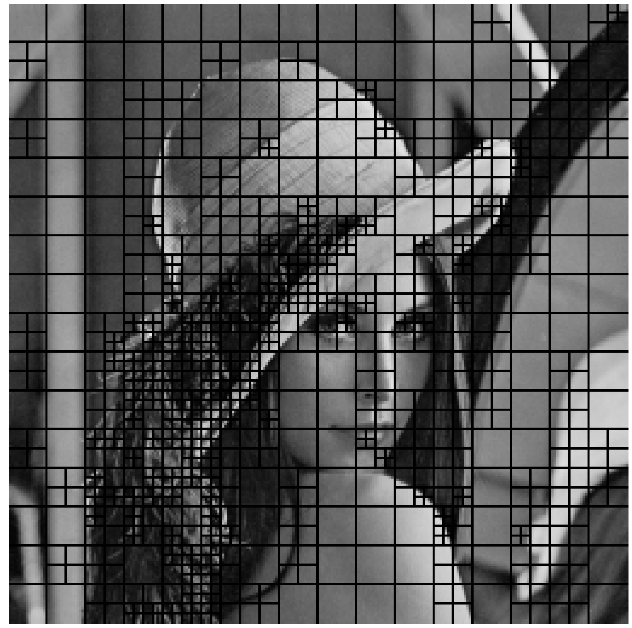

Figure 6: Optimal Segmentation (R=0.44bpp, PSNR=30.33 dB)

Applications. New York: Springer-Verlag, 1994.

[3] M. Gharavi-Alkhansari, T. Huang, Fractal-based techniques for a generalized image coding method. Proceedings of IEEE International Conf. on Image Processing, vol.3, pp.122-126, 1994.

[4] A. E. Jacquin, A fractal theory of iterated Markov operators with applications to digital image coding. Ph.D. thesis, Georgia Institute of Technology, 1989.

[5] G. Melnikov, Hybrid Fractal/DCT image compression algorithms using an orthogonal basis and non-square partitions. M.S. thesis, Northwestern University, 1997.

[6] G. E. Oien, S. Lepsoy, Fractal-based image coding with fast decoder convergence. Signal Processing, vol. 40, pp. 105-117, 1994.

[7] B.-B. Paul, M. H. Hayes III, Video coding based on iterated function systems. In ICASSP-95, IEEE, vol. 4, pp. 2269-2272, 1995.

[8] D. C. Popescu, A. Dimca, H. Yan, A nonlinear model for fractal image coding. IEEE Trans. on Image Proc., vol. 6, no. 3, March 1997

[9] G. M. Schuster, A. K. Katsaggelos, Rate-Distortion Based Video Compression Kluwer Academic Publishers, 1997.