FAST PYRAMIDAL SEARCH FOR PERCEPTUALLY BASED FRACTAL IMAGE COMPRESSION

H. Lin and A. N. Venetsanopoulos

Department of Electrical and Computer Engineering,
University of Toronto, Toronto, Ontario, Canada, M5S 3G4
E-mail: lin@dsp.toronto.edu, anv@dsp.toronto.edu

ABSTRACT

In this paper, we present a fast algorithm for perceptually based fractal image compression. The algorithm is based on a refinement of the fractal code from an initial coarse level of a pyramid. Assuming the block matching error is modeled as a first order Laplacian autoregressive process, we derive the threshold sequence for the objective function in each pyramidal level. Computational efficiency depends on the depth of the pyramid and the search step size, and could be improved by up to two orders of magnitude over the computational effort required for a full search of the original image. The algorithm is quasi-optimal, in terms of minimizing the weighted least absolute error. Its main advantage is the greatly decreased computational complexity, when compared to full search algorithms.

1. INTRODUCTION

During the last decade, fractal geometry has captured increasing attention and interest. The application of fractal models to image compression has been promoted by Barnsley et al [1]. Fractal image compression is based on the observation that all real-world images are rich in affine redundancy. That is, under suitable affine transformations, larger blocks (domain blocks) of the image look like smaller blocks (range blocks) in the same image. The encoding process consists of finding for every range block an affine transformation of a domain block, which fits best in the sense of the used image metric. These affine maps give a compact representation of the original image and are used to regenerate that image, usually with some amount of loss. Since Jacquin presented the first automatic fractal image compression method [2], many additional contributions [3] have been made. In [4], we introduced a perceptually appropriate criterion for perceptually lossless fractal image compression. The method significantly improved the encoding fidelity by using the human visual (HVS) model. The initial definition of the new metric is given in terms of the average local contrast of the block. After the conversion between the physical and digital representation under the given display condition, for perceptually lossless image compression, we have the following inequality:

$$E = \frac{1}{N}\sum_{n=1}^{N} |I_n - \hat{I}_n| \le \bar{I}\, T_B \qquad (1)$$

where N is the block size, I_n is the original image, $\hat{I}_n$ is the fractal encoded image, $\bar{I}$ is the block mean and T_B is the visual contrast threshold. $\hat{I}_n$ can be represented as an affine transform of the contracted domain block I_{M-1}, i.e., $\hat{I}_n = s I_{M-1,n} + t$, where s is the contrast scaling factor and t is the brightness offset. Obviously, the encoding error is measured in terms of weighted l_1 norm. Thus, the encoding process needs to make the block matching under the least absolute deviation (LAD) criterion. Like other fractal-based methods, its major disadvantage is the high computational cost. This is mainly due to the fact that a full search of the domain blocks is needed in order to find the fractal code. To speed up the encoding process, in this paper, we extend the pyramidal algorithm for l_2 norm [5] to the one for l_1 norm. However, the encoding error threshold sequence for l_2 norm is invalid for l_1 norm. Based on the Markov random process theory, we redesign the encoding error threshold for each pyramidal level. For the perceptually lossless compression, the encoding error threshold at the finest pyramidal level will be the same as the visual threshold $\bar{I} T_B$. Furthermore, the least square line fit is different from the LAD line fitting, whose parameters' determination needs an iterative procedure. Therefore, for the parameters of the LAD line, we need to reconsider the propagation rule from the coarse to fine pyramidal levels. The pyramidal search is first carried out on an initial coarse level of the pyramid. This initial search increases the encoding speed significantly, because not only the number of the domain blocks to be searched is reduced, but also the data within each domain block are only a fraction of those in the finest level. Then, only a few numbers of the fractal codes from the promising domain blocks in the coarse level are refined through the pyramid to the finest level with little effort.

2. FAST PYRAMIDAL DOMAIN BLOCK SEARCH

Pyramidal image models employ several copies of the same image at different resolutions. Let $f(x, y)$ be the original image of size $2^M \times 2^M$. An image pyramid is a set of image arrays $f_k(x, y)$, k=0, 1, ..., M, each having size $2^k \times 2^k$. The pyramid is formed by low pass filtering and subsampling of the original image. The pixel $f_k(x, y)$ at level k of a mean pyramid is obtained from the average of its four neighbours $f_{k+1}(x', y')$ at level (k+1):

$$f_k(x,\ y) = \frac{1}{4}\sum_{r=0}^{1}\ \sum_{s=0}^{1} f_{k+1}(2x+r,\ 2y+s) \qquad (2)$$

At the coarsest level (k=0), the image has size 1 and represents the average grey level of the original image. The finest level image f_M is the original image of size $2^M \times 2^M$. As the number of the levels decreases, the image details are gradually suppressed and spurious low spatial frequency components are introduced due to the effect of aliasing. Because the pyramidal structures offer an abstraction from image details, they have been proven to be very efficient in certain image analysis, motion estimation and image compression applications.

Notice that the contracted domain block image $I_{M-1,n}$ in the last section is the corresponding block in (M-1)-th level of the pyramid $f_{M-1}(x,\ y)$. When range blocks are of size $2^m \times 2^m$, the previous optimization objective function for the best matched domain block search can be rewritten as:

$$E = \frac{1}{4^m} \sum_{n=0}^{4^m-1} |D_n(s,\ t)\ -\ R_n| \qquad (3)$$

where $D_n(s, t)=s\, I_{M-1,\, n}+t$ is an affine of the scaled domain block and $R_n= I_n =f(x, y)$ is the range block to encode. Note that for consistency with (1), we use a single subscript n as the index of the pixel at location (x, y). Clearly, $n\ =\ 2^m y + x$.

From the original image a pyramid is created, the depth of which is determined by the range block size. Because the range block is defined in the image, the range block pyramid will be contained in the image pyramid with the k-th level of the range block pyramid corresponding to the (M-m+k)-th level of the image pyramid. Instead of a direct search of the minimum of the objective function at the finest level m, we propose a fast algorithm by introducing a smaller, approximate version of the problem at a coarser level k of the range block pyramid:

$$E^k = \frac{1}{4^k}\sum_{n=0}^{4^k-1} |D_n^{\,k}(s^k,\, t^k) - R_n^{\,k}| \qquad (4)$$

for $k_0 \le k \le m$.

Therefore, at range block pyramid level k, the encoding amounts to finding the best matching domain block of size $2^k \times 2^k$ in the image of the size $2^{M-m+k} \times 2^{M-m+k}$. For example, for an original image of size 512×512 (M=9) and range block size 32×32 (m=5), the search complexity at k_0=2 is that of image size 64×64 and range block of size 4×4. The k=k_0 level of the range block pyramid is said to be initial and every location of the image from the (M-m+k_0)-th level of the image pyramid needs a test. A new feature of the algorithm is the need of the parameter (fractal code) optimization during the block matching. Now, generate a $2^{k+1} \times 2^{k+1}$ promising location matrix G^{k+1}:

$$(G^{k+1})_{2u,2v} = \begin{cases} 1, & \text{if } E^k(u,v) < T^k \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

where (u, v) is the upper left corner coordinates of the domain block and T^k is the threshold at level k. Matrix G^{k+1} is used as a guide in the search of the domain location at the next level k+1. Tests are to be performed only at the locations (i, j) for $(G^{k+1})_{i,j}$ =1 and its neighbour locations. Other parameters P^k of the promising locations are also propagated to P^{k+1} for further refining at level k+1. For the rotation index and the domain block location, we have $\theta^{k+1}=\theta^k$, $D_x^{k+1}=2D_x^{\,k}$, $D_y^{k+1}=2D_y^{\,k}$. Parameters s^{k+1} and t^{k+1} can be obtained from the refinement of s^k and t^k. For the LAD line fitting, it is known that the desired absolute minimum line must pass through at least two points of the given data [6]. Since the refinement of the line parameters only slightly change the position of the absolute minimum line, the first reference point in the iterative process of the fine level k+1 is best chosen as the last reference point through which the absolute minimum line of coarse level k passes. Such a choice of the initial reference point will lead to a smaller number of iterations to locate the line of the minimum deviations of level k+1. The algorithm provides a gradual refinement of the fractal code. The process is repeated recursively until the finest level m is reached. At the finest level, if there exist more than one locations (u, v), such that $(G^M)_{u,v}$=1, select the parameters with the smallest match error as the fractal code. Different thresholds are used to determine promising locations in the corresponding pyramidal levels. The next section shows how to estimate these thresholds using a Markov random process model.

3. DETERMINATION OF THRESHOLDS

The two dimensional encoding error image can be converted into a one dimensional time series X_n after row-by-row scanning. According to equation (1), $X_n = I_n - s\, I_{M-1,n} - t$. Then, the time series is modeled as a stationary first-order Markov process. Since the series has marginal Laplacian distribution, it can be represented as a first-order Laplacian autoregressive (LAR(1)) process:

$$X_n = \rho\, X_{n-1}\ +\ \varepsilon_n \qquad (6)$$

where $|\rho| < 1$,a nd $\{\varepsilon_n\}$ is a sequence of independent, identically distributed (iid), zero mean random variables (RVs). It can be shown that $E^{(k)}$ is an approximate gamma variable. Given the finest block size 4^m, parameter of Laplacian distribution α, correlation coefficient ρ, probability of finding the best match P_0, then, the thresholds are derived as follows.

(1) At the finest level k=m:

$$T^{(m)} = \frac{\sqrt{\beta}}{\gamma}\mu \tag{7}$$

where

$$\beta \approx 4^m \frac{1-\rho_E}{1+\rho_E}, \quad \gamma = \beta\,\alpha, \quad \rho_E = \frac{2\rho^2}{1+|\rho|} \tag{8}$$

(2) At level k ($k_0 \le k < m$):

$$T^{(k)} = \frac{\sqrt{\beta_k}}{\gamma_k}\mu_k \tag{9}$$

where

$$\beta_k \approx 4^k \frac{1-\rho_{E_k}}{1+\rho_{E_k}}, \quad \gamma_k = \beta_k \alpha_k \tag{10}$$

Parameters ρ_{E_k} and α_k follow the iterative equations:

$$\begin{aligned}
\rho_{L_k} &= \frac{\rho_{L_{k-1}}(1+\rho_{L_{k-1}})(1+\rho^2_{L_{k-1}})^2}{2(2+\rho_{L_{k-1}}+\rho^2_{L_{k-1}})} \\
\rho_{E_k} &= \frac{2\rho^2_{L_k}}{1+|\rho_{L_k}|} \\
\alpha_k &= \frac{2\sqrt{2}\,\alpha_{k+1}}{\sqrt{(1+\rho_{L_{k-1}})(2+\rho_{L_{k-1}}+\rho^2_{L_{k-1}})}}
\end{aligned} \tag{11}$$

with the initial $\alpha_m = \alpha$, $\rho_{L_m} = \rho$. In the above, μ_k is a number listed in the incomplete gamma function table [7].

Example. Consider a block of size 8x8 (m=3), given correlation coefficient ρ=0.2 and the probability P_0=0.9. By equation (7) and (9), we get the thresholds for pyramid level k=3, 2 and 1 as follows:

$$\begin{aligned}
T^{(3)} &= \frac{1.1739}{\alpha} \\
T^{(2)} &= \frac{0.7718}{\alpha} \\
T^{(1)} &= \frac{0.5048}{\alpha}
\end{aligned} \tag{12}$$

where $1/\alpha$ is the mean value of the encoding error at the finest level. From the example, we notice that $T^{(k)}$ is a monotonic increasing function of the pyramidal level, that is, fine levels have larger thresholds than the coarser levels.

4. COMPUTATIONAL EFFICIENCY

In this paper, we have used Karst's iterative procedure [6] to determine the parameters s and t. Computer simulation results for general least absolute deviations curve-fitting showed that the actual computation complexity grows linearly with the number of data points [8]. Thus, the computation efficiency analysis in [5] is still valid and is listed here:

$$\begin{aligned}
Q &= \frac{C_1}{C_2} \\
C_1 &= 8\left(\frac{2^{M-1}-2^m}{h}+1\right)^2 2^{2m} \\
C_2 &= 8\left(\frac{2^{(M-m+k_0-1)}-2^{k_0}}{h(k_0)}+1\right)^2 2^{2k_0} + n_p n_s \sum_{i=k_0+1}^{m} 2^{2i} \\
h(k_0) &= \max\left(1, \frac{h}{2^{(m-k_0)}}\right)
\end{aligned} \tag{13}$$

where C_1 and C_2 are the computational cost of full search and pyramidal search, respectively, $h(k_0)$ is the search step size at the initial pyramidal level k_0, and Q is a computational saving factor. As in [5], the pyramid computational saving factor Q (relative to the LAD full search) depends on the depth of the pyramid and the search step size. For the extreme case, the computation efficiency could be improved up to two orders of magnitude when compared with the LAD full search of the original image.

5. EXPERIMENTAL RESULTS

The algorithm in this paper is implemented serially on KSR computer. Figure 1 shows the 512×512 original image Lenna. Quadtree partition is used for range blocks. The initial range block size is 16×16. The encoding error was determined for each range block. Blocks which had an error exceeding the visual suprathresholds (when Gain Factor = 5)[4], were split into four 8×8 blocks. The initial level k_0 is set to 1. The contractive factor s and grey level shift t are coded using 5 and 7 bit uniform quantizers followed by Huffman encoding, respectively. Figure 2 is the plot of the speed up factor Q as a function of the search step size h. When h=16, full search took 601 minutes while pyramidal search took 8 minutes. Figure 3 shows the reconstructed image using full search method at compression ratio 26.3:1 and PSNR=30.5 dB. Figure 4 is the result of the pyramidal search at the same compression ratio and PSNR=30.2 dB. Considering that other fast search techniques such as the conjugate direction search, the 3-step search and the 2D logarithmic search will lead to relative larger matching errors as in motion estimation, thus, we conclude that the pyramid search algorithm is quasi-optimal in terms of minimizing the least absolute error. The main advantage of the pyramidal algorithm is the greatly reduced computational complexity, when compared to full search.

REFERENCES

[1] M. F. Barnsley, and L. P. Hurt, *Fractal Image Compression*, Wellesley, MA: A. K. Peters, 1993

[2] A. E. Jacquin, "Image coding based on a fractal theory of iterated contractive image transformation," *IEEE Trans. Image Process*., vol. 1, no. 1, p18-30, Jan., 1992

[3] Y. Fisher (Ed.), ***F**ractal Image Compression: Theory and **A**pplication*, Springer-Verlag, p55-90, 1995

[4] H. Lin, and A. N. Venetsanopoulos, "Perceptually lossless fractal image compression," Proc. of the Symposium on Visual Communications and Image Processing'96, Spie vol. 2727, p1394-1399, Orlando, USA, 1996

[5] H. Lin, and A. N. Venetsanopoulos, "A pyramid algorithm for fast fractal image compression," *Proceedings of 1995 IEEE International Conference on Image Processing*, vol.3, p596-599, Washington, USA, 1995

[6] J. O. Karst, "Linear curve fitting using least deviations, " *J. **A**mer. Statist. Assoc.***,** 53, p118-132, 1958

[7] K. Pearson (Ed.), *Tables of the Incomplete* Γ*-Function***,** Cambridge University Press for the Biometrika Trustees, Cambridge, 1965

[8] P. Bloomfield, and W. L. Steiger, "Least absolute deviations curve-fitting," *SIAM J. Sci. Stat. Comput.* vol. 1, no. 2, p290-301, 1980


Figure 1. Original image Lenna


Figure 3. Full search, CR=26.3:1, PSNR=30.5 dB

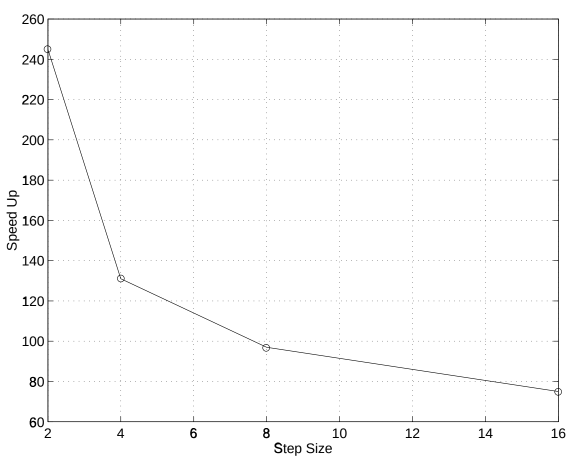

Figure 2. Speed up as a function of search step size


Figure 4. Pyramidal search, CR=26.3:1, PSNR=30.2 dB