On the limitations of fractal image texture coding

Geir E. Øien[1] *Raouf Hamzaoui*[2] *Dietmar Saupe*[2]

[1]Norges Teknisk-Naturvitenskapelige Universitet, Institutt for teleteknikk, N-7034 Trondheim, Norway. Phone: +47 73 59 26 74. Fax: +47 73 59 26 40. E-mail: oien@tele.unit.no
[2]Universität Freiburg, Institut für Informatik, Am Flughafen 17, 79110 Freiburg, Germany

ABSTRACT

Fractal-based image coders suffer from perceptually annoying distortion in textured areas. This paper discusses possible reasons for this limitation. A texture may be modelled as a stochastic process with place-independent autocorrelation properties. Two image areas should have similar correlation (spectral) properties to be perceived as belonging to the same texture. We show that stationary textures in the general case do not possess the self-similarity property on which current block-based fractal coding methods are based. A fractal collage approximation of such a texture may possess quite different spectral properties than is the case for the texture itself. We derive a formula for the transformed correlation function introduced by the decimation in the collage modelling process for the one-dimensional case, and provide coding examples and comparisons to JPEG for practical texture images. Areas with different spectral content, elsewhere in the same image, are needed in order to obtain an approximation with the desired spectral properties for a given texture area. The probability for this specific kind of image nonstationarity to be present may be rather small, which may explain why fractal texture coding often yields perceptually unsatisfying results, even with extensive domain searching.

1. INTRODUCTION TO FRACTAL COMPRESSION

Fractal compression [1] exploits *blockwise self-similarity over different scales* in an image, which is split into nonoverlapping *range* blocks to be coded. The similarity between each range and a set of nonlinearly transformed *domain blocks* is then considered. The domains are taken from a *decimated* version of *the same image* - hence the name *self-similarity*. Each range is approximated by the nonlinearly transformed domain which is the most similar to the range, out of all the domain candidates in the decimated image. This *collage* approximation can be almost perfectly decoded, in an iterative manner, from knowledge of only the domain *position*, and the *parameters* of the nonlinear domain transformation, for every range.

For simplicity of notation, we assume that all images have been ordered into column vectors, denoted by small boldface letters. Formally, an image $\mathbf{x}$ (containing M pixels) is then encoded as the *collage* $T\mathbf{x}$ of an *affine contractive map* $T : X \to X$ on $X = \mathbf{R}^M$ [2, 3]. T consists of a linear part $\mathbf{L} : X \to X$, and an *offset* $\mathbf{t}$:

$$T\mathbf{x} = \mathbf{L}\mathbf{x} + \mathbf{t} = \Big(\sum_{n=1}^{N_r} \alpha_2^{(n)} \mathbf{P}_n \mathbf{O}_n \mathbf{D}_n \mathbf{F}_{d(n)}\Big)\mathbf{x} + \sum_{n=1}^{N_r} \alpha_1^{(n)} \mathbf{P}_n \mathbf{b}_1^{(n)}. \quad (1)$$

N_r is the number of ranges. $\mathbf{F}_{d(n)} : X \to \mathbf{R}^{D_n}$ fetches the best domain-to-be (of position $d(n)$ and length D_n) for range n. $\mathbf{D}_n : \mathbf{R}^{D_n} \to \mathbf{R}^{B_n}$ decimates this block to a domain of range length B_n, by simple averaging ($D_n/B_n = 2$). $\mathbf{O}_n : \mathbf{R}^{B_n} \to \mathbf{R}^{B_n}$ *orthogonalizes* the domain with respect to the offset basis vector $\mathbf{b}_1^{(n)}$, a constant (DC) block. $\mathbf{P}_n : \mathbf{R}^{B_n} \to X$ places the sum of offset and transformed domain in range position n. $\alpha_1^{(n)}$, $\alpha_2^{(n)}$ are scalings for the offset and the transformed domain respectively. These parameters are usually *least squares optimized.* The code for each range consists of the optimized values of $\alpha_1^{(n)}$, $\alpha_2^{(n)}$, and $d(n)$. The *decoded* signal is the *attractor* of T [1],

$$\mathbf{x}_T = \sum_{k=0}^{K-1} \mathbf{L}^k \mathbf{t}, \quad (2)$$

where K is big enough to ensure convergence[1]. The decoded attractor can be made almost identical to the encoder collage approximation (by *The Collage Theorem* [1]).

2. FRACTAL MODELLING OF STATIONARY STOCHASTIC TEXTURES

Any range block $\mathbf{r}$ can be written as

$$\mathbf{r} = \mathbf{O}\mathbf{r} + \mathbf{m}, \quad (3)$$

[1]Typically $K = 4 - 8$.

where $\mathbf{m}$ is a constant block, with all samples equal to the block's mean value. This mean value is equal to the offset parameter α_1 for each range block [3]. Thus it is really only the *range residual*,

$$\mathbf{r}' = \mathbf{O}\mathbf{r}, \tag{4}$$

which is to be approximated by a zero-mean (orthogonalized) domain block. We term the domain-to-be (before decimation) $\mathbf{d}$. Note that whether decimation by averaging or orthogonalization (mean removal) is performed first is irrelevant for the final result. In the discussion in this paper it is practical to reverse the order used in practice, and assume that mean removal is performed before averaging.

Assume now that the range residual, for simplicity denoted $\mathbf{r} = [r_1, r_2, \ldots, r_N]^T$ from now on, can be viewed as consisting of samples from a stationary stochastic process $\{r\}$, with zero mean, characterized by its *autocorrelation function*,

$$\rho(l) = E[r_k r_{k+l}]. \tag{5}$$

Here l denotes *lag*. $\rho(l)$ is a common feature used to describe stochastic *textures* ([4], p. 397).

The above assumption implies that any *orthogonalized domain-to-be* $\mathbf{d}$ also belongs to this stationary stochastic texture, and thus has the same correlation/spectral properties as the range residual. For simplicity and without loss of generality we assume that the texture has been normalized to unit variance, i.e. $\rho(0) = 1$. What happens to the correlation/spectral properties of the domains-to-be when we perform the decimation which is part of the collage modelling process?

2.1. Acf transformation due to decimation by averaging

Assume that any orthogonalized domain-to-be, $\mathbf{d} = [d_1, d_2, \ldots d_{ND}]^T$, is decimated by a factor D by *sample averaging* over D samples at a time, i.e. any range sample r_k is to be approximated by a *collage* sample

$$c_k = \alpha_2 \frac{1}{D} \sum_{i=1}^{D} d_{(k-1)D+i} = \alpha_2 \mathbf{w}^T \mathbf{d}_k \tag{6}$$

where α_2 is the domain scaling factor, which will be least squares fitted to the signal, $\mathbf{d}_k = [d_{(k-1)D+1}, \ldots, d_{kD}]^T$, and $\mathbf{w} = \frac{1}{D}[1, \ldots, 1]^T$ (D 1's) is the decimation filter vector. Using the assumption of least squares optimal α_2, $\mathbf{c} = [c_1, \ldots, c_N]^T$ may be written as

$$\mathbf{c} = \frac{\mathbf{D}\mathbf{w}}{||\mathbf{D}\mathbf{w}||} \cdot \frac{(\mathbf{D}\mathbf{w})^T}{||\mathbf{D}\mathbf{w}||} \cdot \mathbf{r} \tag{7}$$

where $\mathbf{D}$ is the $N \times D$ matrix with $\mathbf{d}_1^T, \ldots, \mathbf{d}_N^T$ as row vectors.

Thus $\mathbf{c}$ is a projection of $\mathbf{r}$ onto the line spanned by the unit domain vector $\frac{\mathbf{D}\mathbf{w}}{||\mathbf{D}\mathbf{w}||}$, for a given domain-to-be vector $\mathbf{d}$. This means that the fit will be better the more parallel $\mathbf{r}$ is to the domain vector $\tilde{\mathbf{d}} = \mathbf{D}\mathbf{w}$, *no matter its norm.*

Assuming that the normalized autocorrelation (NACF) is a relevant measure of the "typical shape" of a process, we are therefore interested in computing the NACF of the domain process (now denoted $\{\tilde{d}\}$), defined as

$$\rho_d(l) = E[\tilde{d}_k \tilde{d}_{k+l}]/E[\tilde{d}_k^2], \tag{8}$$

and comparing it to $\rho(l)$. $\rho_d(l)$ should be equal as possible to $\rho(l)$, if we are to ensure that the domain process really can be used to find perceptually good approximations of range residual blocks.

After tedious but rather straightforward manipulations on the above equations and definitions, which we omit here, we obtain the relationship

$$\rho_d(l) = \frac{\rho(lD) + \frac{1}{D} \cdot \sum_{i=1}^{D-1} (D-i) \cdot [\rho(lD-i) + \rho(lD+i)]}{1 + \frac{2}{D} \sum_{i=1}^{D-1} (D-i) \cdot \rho(i)} \tag{9}$$

As an example, let us consider the much-used image model of an AR(1) process, with $\rho(l) = 0.95^{|l|}$. We assume that the decimation factor is $D = 4$. In Figure 1 $\rho(l)$ and $\rho_d(l)$ are plotted for $k = 0, 1, \ldots$.

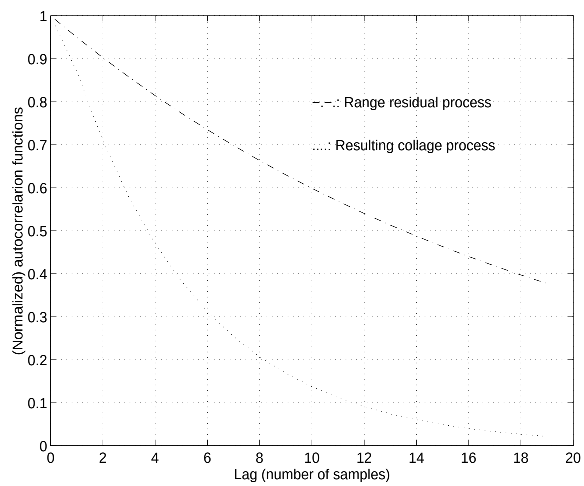

Figure 1. Autocorrelation function for range residual and domain processes.

It can be seen that the domain process in this case is much less correlated than the range residual process[2]. In fact, the domain process is not even an AR(1) process. Typically it will therefore not be visually similar to the process it is used to approximate. This seems to imply that there is a basic weakness in the assumption of self-similarity over different scales for stationary textures. Specifically, *searchless* fractal coding algorithms (such as the one proposed by

[2] Whether correlation increases or decreases will be dependent on the type of process we have.

Monro *et al.* [5]), which use domain-to-be blocks positioned directly above the ranges (implying that the statistical properties of the range and domain-to-be are basically the same) are probably not a good strategy for modelling textures[3].

The saving grace for high-quality fractal texture modelling, then, seems to be the existence of image nonstationarity, i.e. that a given texture can be approximated by a decimated version of *another* texture existing in the same image. The correlation/spectral properties of this second texture must then be such that they approximate those of the first *after the domain decimation.* If we were to investigate the probability of such matches existing, we would have to "invert" the formula for $\rho_d(l)$: I.e., given a desired range autocorrelation $\rho_d(l)$, which $\rho(l)$ must the domain-to-be process have in order to make a good collage approximation? This is not a one-to-one problem, but we may argue that there really is no particular reason why existence of such a "correct" domain-to-be texture should be guaranteed for a given range texture in an arbitrary image. The frequent failure of fractal texture coding in real-world images seems to confirm this suspicion.

3. EXPERIMENTS AND DISCUSSION

We shall perform the following experiments to illuminate the theories presented and to see how they apply to real world images:

1. We will show that fractal compression suffers more than JPEG when texture images are encoded in practice, by encoding *Brodatz texture* images [6] with both schemes, and comparing the visual quality.
2. We will demonstrate that stationary texture images suffer more from collage modelling than images containing several types of textured (and other) areas do.

3.1. Encoding of Brodatz textures

We have approximated a number of Brodatz texture images using

1. Collage approximation
2. JPEG encoding

In each case a uniform range partition of 8×8 blocks was used, and parameters were quantized to obtain a bit rate of 0.4 bits per pixel. In Figures 2 – 4, results are given for the "Metal" texture. These results are representative of every texture tried.

Unless the printing process has let us down, the perceptual difference between the collage and the JPEG image should be rather obvious: The collage looks "out-of-focus" (i.e. more lowpass),

[3]Two obvious, but rather trivial cases for which this is not true are those of a *white noise* (totally uncorrelated) signal and a *fully correlated* signal. In these cases the domain and the range processes will have the same correlation properties.

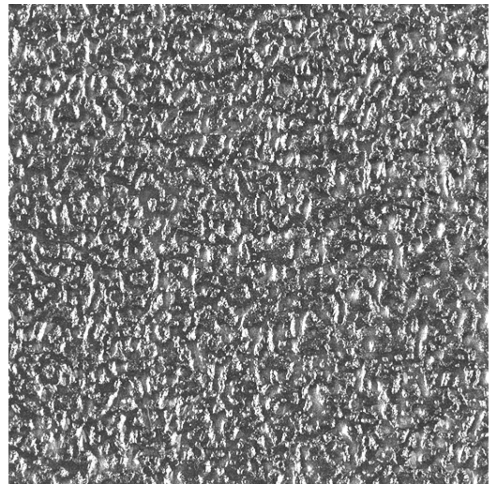

Figure 2. Original metal texture (512×512 pixel resolution).

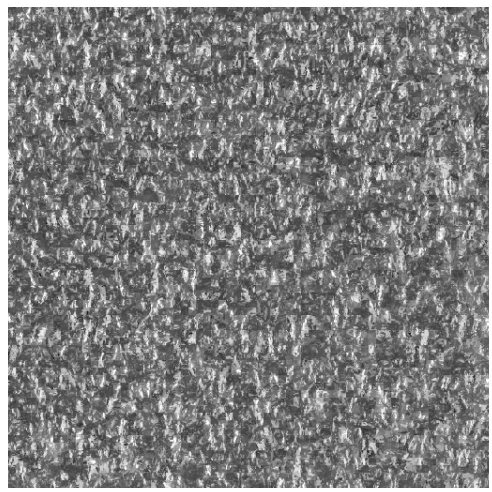

Figure 3. Collage coded metal texture (8×8 ranges, 16×16 domain-to-be's).

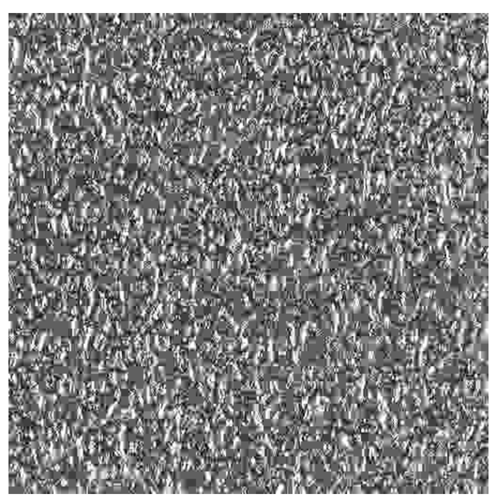

Figure 4. JPEG coded metal texture.

greyish, and of lower contrast than the original. The JPEG image, however, disregarding details at the pixel level, clearly comes from the same kind of texture as the original. This is rather obvious from the way JPEG chooses its parameters: For each block, the most perceptually important frequency components (and therefore, through the inverse Fourier relation, correlation properties) are kept. In the case of the collage, we have no such natural mechanism to retain perceptually important features. The pel-by-pel least squares block matching gives no guarantee of a good perceptual match.

3.2. Comparison between textures and real-world scene images

Collage modelling of the five Brodatz texture files *Stone.0005, woolencloth-1.1.5, roughwall-1.5.3, Fabric.0000*, and *Fabric.0004* yielded a *percentwise RMS NACF error* in the collage which on the average was about 35 % higher than the corresponding error for the collages of the following (nonstationary) real-world scene images: *peppers, Lenna, Airplane, Couple*, and *Boat*. This seems to confirm once more that stationary textures are not well suited for fractal modelling. However, we should comment that it is unlikely that the abovementioned percentwise RMS NACF error measure really provides the best way to measure the *perceptually important* differences in the NACF. An error measured in the frequency domain would probably be more appropriate, but this has not yet been implemented.

4. CONCLUSION

We have discussed the problem of fractal texture modelling in the light of a texture's correlation/spectral properties, and the way these are transformed by the domain-to-be decimation. It is seen that stationary textures do not possess self-similarity in the sense it is currently used in the fractal coding community, since their visual appearance and spectral properties may be significantly changed by this decimation. This is confirmed both by theoretical and experimental results. There are still some unresolved problems, most notably what error measure should be used to quantify the differences in correlation/spectral properties between the collage and the original texture, in order to best reflect the perceptual differences which occur.

REFERENCES

[1] Y. Fisher, ed., *Fractal Image Compression: Theory and Application.* New York: Springer Verlag, 1995.

[2] A. Jacquin, *A Fractal Theory of Iterated Markov Operators with Applications to Digital Image Coding.* PhD thesis, Georgia Institute of Technology, 1989.

[3] G. E. Øien, *L_2-Optimal Attractor Image Coding with Fast Decoder Convergence.* PhD thesis, The Norwegian Institute of Technology, Apr. 1993.

[4] A. K. Jain, *Fundamentals of digital image processing.* Englewood Cliffs: Prentice-Hall, 1989.

[5] D. M. Monro and F. Dudbridge, "Fractal approximation of image blocks," *Proc. Int. Conf. Acoust. Speech, Signal Proc.*, vol. 3, pp. 485 – 489, 1992.

[6] P. Brodatz, *Textures: A photographic album for artists and designers.* NY: Dover, 1966.

THE AUTHORS

Geir E. Øien received his Siv. Ing. degree in 1989, and his Dr. Ing. degree in 1993, both from the Department of Telecommunications at The Norwegian Institute of Technology in Trondheim. From 1994 to 1996 he was an associate professor at Stavanger College, Stavanger, Norway. He currently holds a position as an associate professor at The Norwegian University of Science and Technology in Trondheim, Norway. Dr. Øien's interests are in general signal and image processing, compression, modelling, and classification. He has been on the board of The Norwegian Association of Signal Processing (NORSIG) since 1994.

Raouf Hamzaoui received the Maîtrise de mathématiques from the Faculty of Science, Tunis, in 1986. From 1991 to 1993 he was at the University of Montreal where he obtained an M. Sc. in mathematics. He is currently a Ph. D. student at the University of Freiburg, Germany. His main research interests are in image coding and fractal geometry.

Dietmar Saupe is a Professor of Computer Science at the University of Freiburg, Germany. He received his Dr. rer. nat. in Mathematics in 1982 at the University of Bremen. He was Visiting Assistant Professor of Mathematics at the University of California, Santa Cruz, 1985–87, and from 1987 to 1993 Assistant Professor of Mathematics at the University of Bremen, where he was a researcher at the Center for Complex Systems and Visualization. Dr. Saupe's interests are in mathematical computer graphics, image processing, and scientific visualization. He is coauthor/-editor of "The Science of Fractal Images,", and "Chaos and Fractals", both on Springer Verlag.