FRACTAL REPRESENTATION OF IMAGES VIA THE DISCRETE WAVELET TRANSFORM

H. Krupnik, D. Malah and E. Karnin*

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel; E-mail: hagai@noga.technion.ac.il

ABSTRACT

Fractal representation of images is based on mappings between similar regions within an image (also known as IFS). Such a representation can be applied to image coding and to increase image resolution. One of the main drawbacks of conventional fractal representation is the fact that the mappings are between *blocks*. As a result, the reconstructed image may suffer from disturbing blockiness.

In this work we present a method for mapping similar regions within an image in the *wavelet* domain. we first show how to use the *Haar wavelet* transform coefficients to find mappings which are *identical* to conventional blockwise mappings. The union of these mappings, between sets of wavelet coefficients, can be interpreted as a prediction of higher bands of a signal from its lower band. Changing the mother-wavelet to other than Haar, creates mappings which are between regions which smoothly decay towards their borders, thus reducing the blockiness, as well as improving the PSNR of the reconstructed image.

I. INTRODUCTION

1.1 Mathematical Background.

Fractal representation of a signal is based on the theory of contractive transformations. A transformation T from a metric space (E,d) to itself is contractive iff it has the property :

$$d(x,y) \le a \cdot d(T(x),T(y)) \quad \forall x,y \in E \qquad (1)$$

Where a is called the contractivity factor and must be strictly smaller then 1, and d is the metric.

For a contractive transformation T there *exists* a *unique* point $x^f \in E$, called the *fixed point* of the transformation such that $x^f = T(x^f)$. This unique fixed point is the limit of applying T repeatedly on an arbitrary point $x_0 \in E$.

Fractal representation of a signal x is based on finding a contractive transformation T such that the distance $d(x, x^f)$ between the transformation's fixed point and the signal x is minimal. In practice x^f is not known before T is found. Instead, it is common to find T that minimizes the distance $d(x,T(x))$. This sub-optimal choice is justified by the *collage Theorem* [1]:

1.2 Conventional Fractal Representation

In order to obtain a compact representation of the signal, and to be able to find the best transformation, the family of the considered transformations should be restricted.

Let us describe first the commonly used transformations [2][3] for a discrete, finite support, one-dimensional signal. The transformations are based on a piecewise partition of a vector signal, into non-overlapping sub-vectors R_i of length B each, called *Range Blocks.* A set of sub-vectors of length 2B is also extracted from the signal. Each of these sub-vectors is scaled down using a scaling function $\varphi(X)$ that averages pairs of adjacent samples, to length B. Each of the scaled-down sub-vectors is called a *Domain Block* and is labeled as D_j. The whole set of domain blocks is called the *Domain Pool.* A technical alternative to sub-vector scaling is to scale, first, the whole signal X to get $X_{1/2}=\varphi\{X\}$ and then extract domain blocks D_j of size B from the scaled-down signal. This alternative will serve us later on.

The transformation assigns to each range block R_i a mapping T_i whose parameters consist of a domain block index j, a *scaling factor a,* and an *offset b,* resulting in an affine mapping of the domain block,

$$\hat{R}_i = a \cdot D_j + b, \qquad (2)$$

The parameters are selected so that $\hat{R}_i$ is an *approximation* of the range block R_i which minimizes the squared-error $\|R_i - \hat{R}_i\|_2^2$.

---

* E karnin is with IBM Israel Science and Technology, Haifa.

The transformation T is the union of the mappings T_i, and $T(X)$ is the union of the reconstructed range blocks $\hat{R}_i$. The fixed point of the transformation is approached by applying T over and over again on an arbitrary initial signal, as implied by the name 'Iterated Function System' (IFS), for the transformation T. For a detailed discussion on the contractivity of T see [4]

Fractal representation of images as 2D signals is similar to the 1D representation just described [2][3]. The main difference is that the range blocks are squares of BxB pixels. The domain pool is obtained by 2^2:1 scaling down of 2Bx2B extracted blocks or by extracting BxB blocks from a 2^2:1 scaled-down image. In the 2D case The domain pool is usually enriched by applying isometries (such as rotations and flips) on its domain blocks [2][3]. These isometries can be considered as distinct members of the domain pool.

1.3 Fixed Point Pyramid

Let us use the terminology $X_{1/2^k} = \varphi(X_{1/2^{k-1}})$ where $k = 1,2,\ldots,\log_2(B)$ to represent the lower resolutions (scaled-down versions) of a 1D or 2D signal $X_1 \equiv X$. It has been shown [5] [6] that each of the lower resolutions of X_1^f, the fixed point of T, is a fixed point of a related transformation, denoted as $T_{1/2^k}$. This transformation has in fact the same parameters as $T \equiv T_1$, but is applied with a smaller range block length of $B/2^k$. The relations between lower resolutions of the fixed point is demonstrated in Fig. 1.

Since the conventional IFS involves extracting domain blocks from $X_{1/2}$ and copying them to an appropriate range blocks in X_1, and since this is also the situation for higher levels of the pyramid, one can get X_1^f as a *zoom-in* interpolation procedure, that starts from $X_{1/B}^f$ and uses all mappings in $T_{1/2^{k-1}}$ to apply the affine transformation on domain blocks of size $B/2^{k-1}$ from $X_{1/2^k}^f$, and tile the next higher resolution $X_{1/2^{k-1}}^f$.

1.4 DC Orthogonalization

A Modification of the conventional mapping in (2) is a mapping with orthogonalization to the block averages (DC) [7]. The reconstructed block and the transformation can be described as:

$$\hat{R}_i = a_i \cdot (D_{j(i)} - \overline{D}_{j(i)}) + \overline{R}_i \quad T_i \equiv \{a_i, \overline{R}_i, j(i)\} \qquad (3)$$

Where $\overline{R}_i$ and $\overline{D}_j$ are the averages of the corresponding

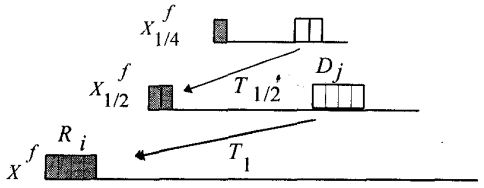

Fig. 1 - Mapping relations for different scales of the fixed point of an IFS (B=4)

blocks. One of the main advantages of defining T as in (3) over the one in (2) is that its fixed point can be found in exactly $\text{Log}_2(B)+1$ iterations [7], i.e. $T^{l+m}=T^l$ for all $m \geq 0$ and $l = \text{Log}_2(B)+1$. In this case there exist a tighter bound on the collage error [8].

The DC block orthogonalization does not affect the pyramidal relations between the different scales of the fixed point. Therefore, it is possible to combine DC orthogonalization with the fixed point pyramid representation, and perform a non-iterative hierarchical decoding of the signal from its IFS as follows [7] :

1. Construct the pyramid level $X_{1/B}^f$ in which the *point* $X_{1/B}^f(i)$ is the range block average $\overline{R}_i$.
2. "Zoom-in", $\log_2(B)$ times, starting from $X_{1/B}^f$, to get $X_1^f = X^f$ as described in section 1.3. This time use (3) to copy domain blocks from a given pyramid level to tile the next lower level (higher resolution).

In the next section we refer to this procedure as the "*reference algorithm*" and "transfer" this algorithm to the wavelet domain.

II. IFS MAPPINGS IN THE HAAR WAVELET DOMAIN

2.1 Subband Decomposition and Wavelet Transforms

Let us describe the *Discrete Wavelet Transform* (DWT) of a 1D *discrete* signal X as an octave-band subband decomposition (Fig. 2, bottom), using a pair of *Quadrature Mirror Filters* (QMF) [9]. The octave-band filtering is implemented by splitting over and over the lowest subband with the two-band pair of filters. Such a decomposition with three splits is shown in Fig. 2. At each split the higher branch represents highpass filtering and 2:1 decimation, and the lower branch represents the corresponding lowpass filtering and decimation. The outputs of the lower branches are, in fact, lower scales of the signal, Therefore they are labeled $X_{1/2^k}$, as in section 1.3, The the higher branches outputs are labeled $X_{\text{H}1/2^k}$.

Mallat [9] has shown the connection between Orthogonal

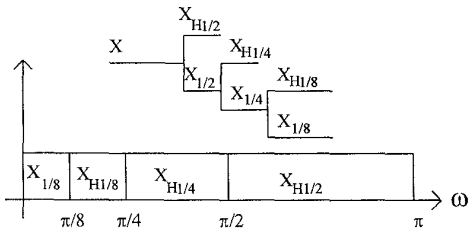

Fig. 2 - The DWT of a signal. Top - the splitting tree; Bottom - frequency division into subbands.

Wavelet Transforms, and QMF subband decomposition. He has shown that given a perfect reconstruction QMF pair, such that the reconstruction filters are the same as the analysis filters, the output of the octave-subbands are, in fact, the coefficients of an orthonormal DWT. Hence the sub-band decomposition can be represented by a unitary matrix U..

2.2 IFS and the Haar-DWT

The Haar-DWT is an orthonormal transform that its equivalent QMF pair is the lowpass filter H_L=[1,1] and the highpass H_H=[1,-1] (up to a gain factor). Note that 2:1 decimation of the output of the convolution of X with H_L results in $X_{1/2}=\varphi\{X\}$, the scaled-down signal described in section 1.2. The aim of this section is to show the following

1. Given an IFS, the Haar-DWT coefficients of the higher subbands of its fixed point can be calculated from the fixed point's lowest band ($X^f_{1/B}$).
2. The IFS mapping parameters, obtained by the reference algorithm in section 1.4 can be found directly from the Haar-DWT coefficients of the input signal X.

Consider a signal X^f that is a fixed point of an IFS. X^f is tiled with range blocks such that for any block, R, there is a corresponding domain block D in $X^f_{1/2}$ such that $R = a(D-\overline{D})+\overline{R}$. Let us decompose separately R and D by the Haar-DWT with L=$\log_2$(B) splits. The transform could be described by a unitary BxB matrix U:

$$\begin{aligned} [r^1\ r^2\ \ldots r^B]^t &= UR; \quad R = U^t[r^1\ r^2\ \ldots r^B]^t \\ [d^1\ d^2\ \ldots d^B]^t &= UD; \quad D = U^t[d^1\ d^2\ \ldots d^B]^t \end{aligned} \tag{4}$$

where $[r^1 \ldots r^B]$ and $[d^1 \ldots d^B]$ are the coefficients of R and D, respectively. r^1 and d^1 are the only lowpass coefficients and provide (up to a constant gain factor) the averages of the blocks ($\overline{R}$ and $\overline{D}$). Since for $X=X^f$, $R=\hat{R}$, applying the unitary matrix U on both sides of (3)

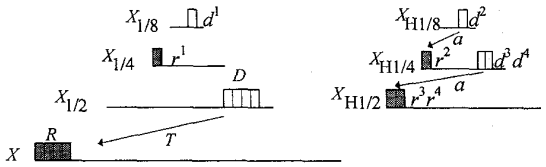

Fig. 3 - Relations between the Haar-DWT coefficients Left - Lowpass pyramid; Right - Highpass pyramid

gives :

$$U(R-\overline{R}) = a\cdot U(D-\overline{D}) \Rightarrow r^p = a\cdot d^p,\ 2\le p\le \mathrm{B} \quad (5)$$

Suppose, now that we decompose the *whole* fixed point X^f with a L+1=$\log_2$(2B) split, Haar-DWT, as in Fig. 2, and build two pyramids. The first is a lowpass pyramid that consists of $X^f_{1/2^k}$ (k=1,...,L+1), where only $X^f_{1/2B}$ is part of the DWT. The second is a highpass pyramid that consist of $X^f_{H1/2^k}$, and is fully contained in the DWT. The two pyramids are illustrated in Fig. 3 (in this figure B=4).

If we choose to construct the domain pool from blocks that tile $X_{1/2}$ without overlap, it is easy to see that the Haar-DWT coefficients $[r^2 \ldots r^B]$ and $[d^2 \ldots d^B]$ of R and D are part of the Haar-DWT of the *whole* fixed-point signal X^f. This is due to the fact that a $\log_2$B split Haar-DWT of a vector, consists of a set of independent Haar-DWT on non-overlapping sub-vectors of size B. In Fig. 3, the DWT coefficients of a four sample range block in X (shaded) and domain block (white) in $X_{1/2}$, are marked in the highpass and lowpass pyramids.

Since all r^p and d^p ($1\le p\le$B) are parts of the L+1 split DWT of X^f, and since for each p, d^p is from a subband of lower frequency than r^p, (5) shows that the coefficients of the higher bands can be calculated from those of the lower bands, i.e. one can start with $X^f_{H1/2B}$, calculate all r_i^2 which are in $X^f_{H1/B}$, and so on. This calculation can be interpreted as a recursive *extrapolation*, or *prediction* of the higher bands from the lowest one. At each stage another subband is extrapolated and so a higher resolution of the fixed point can be obtained, quite similar to the hierarchical decoder that is described in section 1.3.

Can we find the IFS code in the wavelet domain as well ? The answer depends on the metric. If one uses the standard l_2 inner product metric (a very common choice), then, due to the norm preserving characteristics of the DWT transform, the approximation error minimization can be performed in the DWT domain:

$$d^2(\hat{R}_i, R_i) = d^2(U\hat{R}_i, UR_i) = \sum_{p=2}^{B} (r_i^p - a_i \cdot d_j^p)^2 \qquad (6)$$

Thus, the minimization of the distance can be performed by comparing the blocks' wavelet coefficients that have been extracted from the DWT of the whole signal. In each block there are 2^L-1 coefficients as in the right pyramid of Fig. 3. Each range and domain block consists of coefficients from *different* subbands, representing the *same location* of the signal. A summary of the encoding/decoding process is as follows

Finding an IFS (Encoding):

1. Compute the DWT with L+1 splits of the signal
2. Extract range and domain blocks of AC coefficients.
3. For each range block R_i, find the best domain block D_j and scaling factor a_i that minimizes (6).
4. The IFS code consists of the upper lowpass and highpass levels of the pyramid ($X_{1/2B}$ and $X_{H1/2B}$), the scale factors a_i, and the indices $j(i)$ for each block.

Finding the IFS Fixed point (Decoding)

1. Copy the lowpass and highpass levels of the pyramid from the code.
2. Extrapolate downwards in the highpass pyramid using the scale factors and the indices.
3. Compute the Inverse DWT.

2.3 2D IFS in the Haar-Wavelet Domain

In order to apply the above algorithm to images there is a need to extend it to two dimensions. This is most easily done with separable 2D QMF's. The subband decomposition of a separable octave-band QMF is a quadtree partition of the 2D frequency domain as illustrated in the left side of Fig. 4. In such a decomposition every split consists of four bands, labeled "LL", "HL", "LH" and "HH". Exactly as it was done for 1D signals, all the highpass coefficients that represent the 2D separable Haar transform (2D Haar-DWT) of BxB domain and range blocks, can be found in the 2D Haar-DWT decomposition ,with L+1 splits of the *whole* signal. In the left side of Fig. 4, there are 15 shaded 2D Haar-DWT coefficients. These coefficients are the DWT of the 4x4 (shaded) range block of an image shown in the right. The 15 white coefficients represent a single domain that represent a scaled-down version of the 8x8 square that is also shown on the right of Fig. 4 Therefore :

1. One can extract range blocks and domain blocks of B^2-1 HP coefficients from the image Haar 2D-DWT coefficients in order to find the best IFS.
2. Given an IFS, one can extrapolate the higher bands in all three directions (HL LH and HH) *simultaneously* (i.e. using the same mappings).

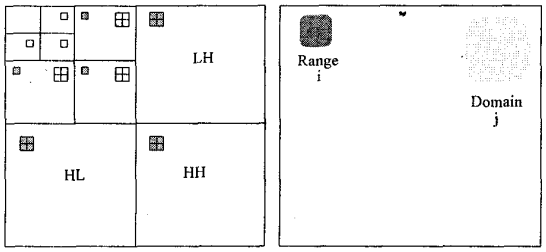

Fig. 4 - The coefficients of the 2D-DWT of a Domain (white) and Range (shaded) blocks B=4
Left - The DWT domain Right - the corresponding regions

Extending the coding algorithm described in section 2.2 to 2D needs an explanation on how to implement the isometries in the 2D-DWT transform domain. It can be shown that all the conventional isometries of a block means nothing else but :

1. applying the same isometries in each of the subbands separately.
2. switching between HL and LH subbands (for some of the isometries)
3. inverting the sign of the coefficients (for some of the isometries and in some of the bands).

III A BLOCKLESS IFS

The investigation of the IFS in the Haar-DWT domain gives a better understanding of the frequency characteristics of the code. Although many practical benefits can be achieved from the subband interpretation, the main benefit comes, probably, from changing the QMF pair. Suppose that we decompose an image into its subbands, as in Fig. 4, with a separable QMF pair other than Haar. Then, collect a range block of B^2-1 coefficients from different subbands that belong to some area of the image, such as the shaded coefficients in Fig. 4 (left). Suppose that we pick the white coefficients from a one level shifted down of the decomposition to be a domain. The new "Domain" coefficients still represent a decomposition of a scaled-down version of the image. However the scaling function φ in not a four pixel averager. Instead it is the LP QMF 2D separable filter followed by a decimator. Such a filter has a much smoother characteristics as a scaling operator then the averager. Finding the best domain block of coefficients and the best scaling factor a, that describe each Range block of coefficients, is equivalent to finding similar regions in an image and in its scaled-down version. However those regions overlap since the support of basis vectors overlap. Morover they have smooth decaying borders. Thus a code that is not in blocks have been obtained. To summarize, if one uses the algorithm for

finding an IFS and its fixed point, as in section 2.3, using a QMFs pair other than Haar, he effectively obtains a blockless reconstructed image.

The chosen QMF should have the following properties. First and most important, since the IFS finds correlation *between* different frequency bands (in different locations) it should have zero (linear) phase. In addition it should be relatively short due to the finite size of the image. It is sufficient however that the QMF will provide near-perfect reconstruction.

IV RESULTS AND DISCUSSION

In order to evaluate the quality of the IFS that can be obtained in the DWT domain (DWT-IFS), Its ability to compress images have been examined. The wavelet coefficients of the lower bands were quantized with 7 bits/coeff for the LP band and 6 bit/coeff for the three HP bands. The scaling factors were also quantized with 6 bits each. All the quantizers used, were uniform. The quantization was followed by adaptive arithmetic encoding of each of the quantized variables.

Several separable QMF's were examined on a few images, and all of the filters performed better than the Haar-QMF. The best results, in terms of PSNR, as well as subjectively, were obtained with Adelson's et. al. 9 taps QMF's [10]. As an example, we compare the results of coding the picture "lena" of size 512x512 with three different block sizes (actually levels of prediction): L=2 (equivalent to B $=2^{L}=4$), L=3 and L=4. The bit rate, and the PSNR values (in dB) obtained are given in Table 1.

	PSNR	PSNR	Bit/pel	Bit/pel
	n=2(Haar)	n=9	n=2	n=9
L=2	36.95	38.67	1.56	1.57
L=3	30.48	31.63	0.35	0.35
L=4	25.85	26.81	0.08	0.08

Table 1: Results (bit rate and PSNR in dB) of coding the 512x512 image "lena"

The reduction of the artifacts due to the use of the 9 taps QMF is shown in Fig. 5, where a part of the image "lena" that was coded with L=4 is magnified. It is easy to see that the blockiness of the reconstructed Haar-IFS image (Fig. 5a) has vanished in the 9 taps QMF image (Fig. 5b). However, the results in Table 1 do not compete with the state of the art wavelet coders[11]. The reason for that might be that in order to achieve better results one has to adapt the size of the blocks to their activity. The conventional IFS coder does it by Quadtree partitioning of the image[3] A similar approach for the DWT-IFS is currently under investigation.

The pyramidal interpretation of an IFS code may also be used to increase image resolution. The proposed

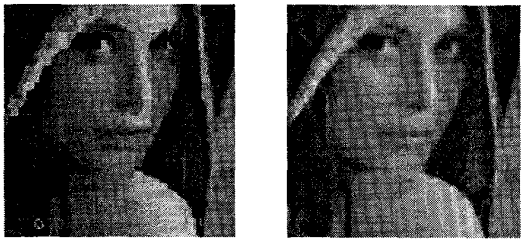

Fig. 5 - parts of the reconstructed image "lena"
(a) Haar DWT, L=4 (b) Adelson 9 Taps, L=4

algorithm offers an opportunity to locally estimate the (missing) high frequency band, i.e extrapolate in the wavelet domain in order to interpolate in the spatial domain. The use of IFS-DWT to increase image resolution is another issue that needs further investigation.

REFERNCES

[1] M. F. Barnsley, "*Fractals Everywhere*", London, Academic Press, 1988.

[2] E. Jacquin, "Fractal Image Coding: A Review", *Proc. of the IEEE*, Vol. 81, No 10, Oct. 1993, pp. 1451-1464.

[3] E. W. Jacobs, Y. Fisher and R. D. boss, "Image Compression: A Study of the Iterated Transform Method", *Signal Processing*, Vol. 19, 1992, pp. 251-263.

[4] B. Hurtgen, T. hain, "On the Convergence of Fractal Transforms", *ICASSP*, 1994, Vol. V, pp. 561-564.

[5] Z. Baharav, D. Malah, E. Karnin, "Hierarchical interpretation of fractal image coding and its application to fast decoding", *Proc. Digital Signal Processing Conference*, Cyprus, July 1993, pp. 190-195.

[6] Z. Baharav et.al., CH. 5 in Y Fisher ed. "*Fractal Compression: Theory and Applications to Digital Images*", Springer, Berlin, Sept. 1994.

[7] G. E. Oien, S. Lepsoy, "Fractal Based Image Coding with Fast Decoder Convergence", *Signal Processing*, No 40, 1994, pp. 105-117.

[8] G. E. Oien, Z. Baharav, S. Lepsoy, E. Karnin, D. Malah, "A New Improved Collage Theorem With Applications to Multiresolution Fractal Image Coding", *ICASSP*, 1994, Vol. V, pp. 565-568.

[9] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Trans. of PAMI*, Vol. 11 No 7, July 1989, pp. 674-693..

[10] H. Adelson and E. Simoncelli, "Orthogonal Pyramid Transforms for Image Coding", *SPIE*, Vol. 845, 1987, pp. 50-58.

[11] J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", *IEEE Trans. on Sig. Proc.*, Vol. 41, No 12, Dec. 1993, pp. 3445-3462.