

Fractal-Wavelet Image Denoising Revisited

Mohsen Ghazel, George H. Freeman, and Edward R. Vrscay

Abstract—The essence of fractal image denoising is to predict the fractal code of a noiseless image from its noisy observation. From the predicted fractal code, one can generate an estimate of the original image. We show how well fractal-wavelet denoising predicts parent wavelet subtree of the noiseless image. The performance of various fractal-wavelet denoising schemes (e.g., fixed partitioning, quadtree partitioning) is compared to that of some standard wavelet thresholding methods. We also examine the use of cycle spinning in fractal-based image denoising for the purpose enhancing the denoised estimates. Our experimental results show that these fractal-based image denoising methods are quite competitive with standard wavelet thresholding methods for image denoising. Finally, we compare the performance of the pixel- and wavelet-based fractal denoising schemes.

Index Terms—Fractal image coding, fractals, image denoising, image restoration.

I. INTRODUCTION

IN this paper, we present some results of ongoing work on fractal-based image denoising methods. The essence of these methods is to predict the fractal code of a noiseless image from its noisy observation. Such a prediction can be performed either in the pixel domain [9] or in the wavelet domain [8]. The predicted fractal code is then used to generate a denoised estimate of the original image by iteration of the associated fractal transform.

In [8] and [9], experimental results showed that our fractal-based methods significantly reduced the noise and errors in the noisy images (in all cases, the noiseless image was known). This suggests that the prediction of the fractal code of the noiseless image works quite well: As we report below, the prediction method produces assignments that are near optimal, i.e., they generally lie within the top one or two percent of the best available domain subtrees from the image, and, as expected, with increasing noise variance, the deviation of predicted subtrees from the most optimal ones increases. This portion of our work was inspired by some recent studies of the statistical properties of domain-range matching in images [1]. In general, there are many domain subtrees that match a given range subtree very

Manuscript received November 10, 2004; revised December 12, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ron Kimmel.

M. Ghazel was with the Department of Electrical and Computer Engineering, Faculty of Engineering, University of Waterloo, Waterloo, ON N2L 3G1 Canada. He is now with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4 Canada (e-mail: mghazel@ece.ubc.ca).

G. H. Freeman is with the Department of Electrical and Computer Engineering, Faculty of Engineering, University of Waterloo, Waterloo, ON N2L 3G1 Canada (e-mail: freeman@uwaterloo.ca).

E. R. Vrscay is with the Department of Applied Mathematics, Faculty of Mathematics, University of Waterloo, Waterloo, ON N2L 3G1 Canada (e-mail: ervrscay@uwaterloo.ca).

Digital Object Identifier 10.1109/TIP.2006.877377

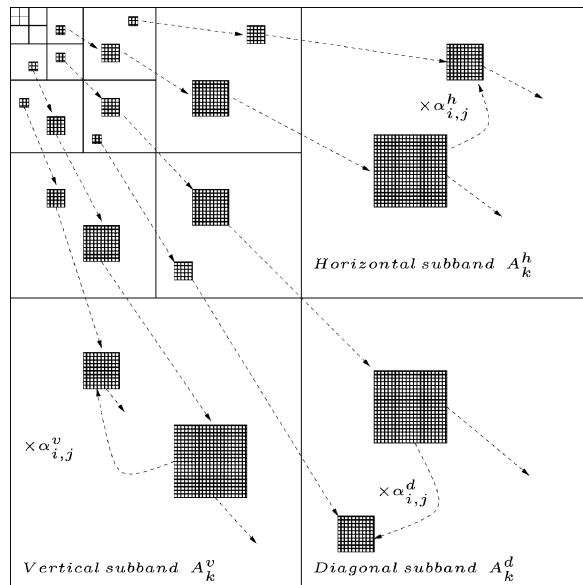


Fig. 1. Two-dimensional fractal-wavelet transform.

well. In other words, images generally do possess a high degree of local self-similarity.

II. FRACTAL-WAVELET IMAGE DENOISING

A. Basics of Fractal-Wavelet Image Coding

Fractal-wavelet (FW) transforms, discovered independently by a number of researchers ([5], [6], and [12] to name only a few), were introduced in an effort to reduce the blockiness and computational complexity that are inherent in fractal image compression. Their action involves a scaling and copying of wavelet coefficient subtrees to lower subtrees, quite analogous to the action of fractal image coders in the spatial domain.

Recall that the discrete wavelet transform (DWT) coefficients of a two-dimensional (2-D) signal (image) are conveniently arranged in a standard matrix fashion, the first few blocks of which are shown in Fig. 1 (here, we are assuming that the 2-D wavelet basis functions are constructed in the usual way by using suitable tensor products of one-dimensional scaling and wavelet functions—see [14]). Each of the blocks $\mathbf{A}_k^h, \mathbf{A}_k^v, \mathbf{A}_k^d, 0 \leq k \leq K$, for some $K > 0$, contains 2^{2k} coefficients $a_{kij}^h, a_{kij}^v, a_{kij}^d$, respectively. The three collections of blocks comprise the fundamental *horizontal*, *vertical*, and *diagonal* quadtrees of the coefficient tree. Now, consider any wavelet coefficient $a_{kij}^\lambda, \lambda \in \{h, v, d\}$ in this matrix and the unique subtree, denoted by A_{kij}^λ , with this element as its root.

The “collage coding” procedure to produce the FW code for an image proceeds as follows: First consider a fixed set of parent-child level values (k_1^*, k_2^*) , where $k_1^* < k_2^*$. Then,

for each *uncoded* child subtree, $A_{k_2^*, i, j}^\lambda$, $i, j = 1, 2, \dots, 2^{k_2^*}$, find the parent subtree $A_{k_1^*, i', j'}^\lambda$ and the corresponding scaling coefficient $\alpha_{i, j, i', j'}$ so that the so-called ‘‘collage distance’’

$$\Delta_{i, j, i', j'}^\lambda = \left\| A_{k_2^*, i, j}^\lambda - \alpha_{i, j, i', j'} A_{k_1^*, i', j'}^\lambda \right\|_2 \quad (1)$$

is minimized. The resulting FW code consists of the following.

- 1) The parent-child index pair (k_1^*, k_2^*) .
- 2) The wavelet coefficients $\hat{c}_{k, i, j}$, for $1 \leq k \leq k_2^* - 1$: A total of $4^{k_2^*}$ coefficients. These wavelet coefficients are unchanged by the FW transform.
- 3) The scaling factors $\alpha_{i, j}^\lambda$ and parent subtree indices, (i^λ, j^λ) , for all elements $a_{i, j}^\lambda$ in each of the three blocks $A_{k_2^*}^\lambda$, for $\lambda \in \{h, v, d\}$: A total of $3 \times 4^{k_2^*}$ scaling factors and $2 \times 3 \times 4^{k_2^*}$ indices.

In the *standard* FW scheme (as proposed in [5] and [12], for example) it is assumed that common parents and scaling factors are used for the three fundamental subbands. Of course, such a restriction is performed at the expense of image fidelity. Historically, it was employed for purposes of data compression.

For FW decoding, the usual procedure is to start with a wavelet coefficient tree that contains the stored wavelet coefficients and zeros elsewhere. Iteration of the fractal-wavelet scaling and copying procedure produces a ‘‘fixed point’’ wavelet coefficient matrix that is an approximation to that of the original image. The smaller the ‘‘collage distance’’ in (1), the better the approximation [15].

Given a wavelet coefficient tree, let the range subtree $R_{i, j} = A_{k_2^*, i, j}^\lambda$ be represented by the vector $\mathbf{y} = [y_1, y_2, \dots, y_n]$ and the domain subtree $D_{i', j'} = A_{k_1^*, i', j'}^\lambda$ be represented by the vector $\mathbf{x} = [x_1, x_2, \dots, x_n]$. In practice, the \mathcal{L}^2 norm is used so that collage coding of the domain-range pair minimizes the the \mathcal{L}^2 error $\Delta = \|\mathbf{y} - \alpha \mathbf{x}\|$. In what follows, we shall regard the wavelet transform of an image as a random signal so that the wavelet coefficients \mathbf{x} and \mathbf{y} can be considered as random samples drawn from the random variables X and Y representing the wavelet coefficient distributions of a parent subtree D and its corresponding child subtree R , respectively. The optimal least-squares scaling coefficient can then be written as

$$\alpha^* = \frac{E[XY]}{E[X^2]} \quad (2)$$

where

$$E[XY] = \frac{1}{n} \sum_{m=1}^n x_m y_m \text{ and } E[X^2] = \frac{1}{n} \sum_{m=1}^n x_m^2. \quad (3)$$

Strictly speaking, the above expressions are approximations to the statistical quantities of the random variables X and Y since they represent limited sample statistics. The fact that n will not be very large in our applications will contribute to errors in estimating the local image statistics and, subsequently, suboptimal fractal codes for the noiseless images.

B. Remarks

Recall that fractal-wavelet transforms exploit the local self-similarities exhibited by wavelet coefficients across resolution

scales [15]. However, noisy structures have little or no resemblance across resolution levels. One would, therefore, expect that the FW coding of a noisy image would result in some denoising. It was observed in [7] that this was, indeed, the case. However, even though noise reduction has been performed, the denoised image is clearly not an ideal estimate of the original noiseless image. This is to be expected since the collage-based FW coding was performed using *noisy* image data. As we have shown for fractal coding in the pixel domain [9], one can go a step further and predict the fractal-wavelet code of the *original* image from these noisy observations.

C. Predicting the FW Code of the Noiseless Image

If we regard the wavelet transform of an image as a random signal, then the fractal-wavelet coding process can be reduced to the following mean-squared error (MSE) estimation problem: For each *uncoded* child subtree Y_k , $k = 1, 2, \dots, 2^{k_2^*}$, find the optimal parent subtree, X_i , and its corresponding scaling coefficient, α_{ik}^* , for which the MSE, given by

$$\begin{aligned} \Delta_{ik}^2 &= E \left[(Y_k - \alpha_{ik}^* X_i)^2 \right] \\ &= E \left[Y_k^2 \right] + \alpha_{ik}^{*2} E \left[X_i^2 \right] - 2\alpha_{ik}^* E \left[X_i Y_k \right] \end{aligned} \quad (4)$$

is minimized. This yields the MSE scaling coefficient estimate of (2)

In practice, however, we must work with noisy images. In what follows, random variables representing coefficients that correspond to a noisy image will have hats, e.g., \hat{X} and \hat{Y} . In the case of an AWGN noise and when using an *orthogonal* wavelet basis, the relationship between the subtrees of wavelet coefficients corresponding to the noisy and the noiseless images is as follows:

$$\hat{X}_i = X_i + W_{X_i} \text{ and } \hat{Y}_k = Y_k + W_{Y_k}. \quad (5)$$

Here, W_{X_i} and W_{Y_k} are identically distributed AWGN processes, $N(0, \sigma_w^2)$, which are also assumed to be statistically independent. The independence can be achieved by ensuring that the child and parent subtrees do not overlap. We also assume that the image and the noise signals are independent, which is reasonable and practical. In view of the above assumptions, it can easily be shown that

$$\begin{aligned} E \left[X_i^2 \right] &= E \left[\hat{X}_i^2 \right] - \sigma_w^2 \\ E \left[Y_k^2 \right] &= E \left[\hat{Y}_k^2 \right] - \sigma_w^2 \\ E \left[X_i Y_k \right] &= E \left[\hat{X}_i \hat{Y}_k \right]. \end{aligned} \quad (6)$$

It is important to note here that in [8, (14)], the noise variance term σ_w^2 was incorrectly added to instead of subtracted from the noisy second moment estimates. This may account for the difficulties experienced by Barthel *et al.* [2] in obtaining any gain in the denoising behaviour of the fractal coder.

In view of the above derivations, we make the following important observations.

- The above derivations allow us to estimate the statistics of the original noiseless image (unknown in practice) from

those of its noisy observation. When these estimates are robust, they can be used to estimate the FW code of the noiseless image. As in the case of pixel-based fractal image denoising [9], we also estimate the squared collage errors Δ_{ik}^2 in the noiseless image in terms of the statistics of the noisy observation. For a given child subtree, we select the parent subtree for which the estimated error Δ_{ik}^2 is minimized.

- The robustness of these statistical estimates is achieved when the energy of the noisy parent and child subtrees are sufficiently larger than the noise variance, that is

$$E \left[\hat{X}_i^2 \right] \geq \kappa \sigma_w^2 \text{ and } E \left[\hat{Y}_i^2 \right] \geq \kappa \sigma_w^2 \quad (7)$$

for some parameter $\kappa \gg 1$. It was observed that for various test images, one obtains better results for κ values in the range between 1.5 and 2.5. For our experimental results, it was assumed that $\kappa = 2$.

- When the robustness criteria in (7) is not satisfied, (6) may yield negative second moment estimates and the predicted scaling coefficient in (2) may be unbounded. Thus, in the case of sparse signal information, our predictive methodology cannot be applied. In our previous work [8], we simply encoded this type of noisy subtrees using the FW scheme. No attempt was made to predict the FW code of the corresponding noiseless subtrees. In [2], it was suggested that reducing the scaling coefficient, in this case, may be beneficial. For such predominantly noisy subtrees, a reduction in the magnitude of the noisy scaling coefficient would suppress some of the noise. Here, we acknowledge the above observation and employ a slightly modified method to reduce the magnitude of the noisy scaling coefficient $\hat{\alpha}_{ik}^*$ as follows:

$$\alpha_{ik}^* \approx \min \left(\frac{E \left[\hat{Y}_k^2 \right]}{\kappa \sigma_w^2}, \frac{E \left[\hat{X}_i^2 \right]}{\kappa \sigma_w^2} \right) \times \hat{\alpha}_{ik}^*. \quad (8)$$

- Although the above algorithm was outlined for the purpose of using the standard FW scheme, it is, indeed, a straightforward matter to generalize it to other FW schemes, such as the quadtree-based FW scheme which uses a collage error decomposition criterion [10]. It is recommended that one use a FW that combines the three subbands in order for the child and parent subtrees to have sufficiently large size. Otherwise, poor estimates of the local statistics may lead to poor results.

D. Enhancing the Fractal Denoised Estimates Using Cycle Spinning

Similar to most wavelet-based methods, the FW denoised estimates generally exhibit some pseudo-Gibbs artifacts. Also, as is the case for most block-based spatial methods, the denoised estimates obtained by pixel-based fractal denoising schemes tend to suffer from blockiness artifacts. Motivated by [3], we examine the use of cycle spinning to reduce such disturbing artifacts.

For the case of the predictive FW scheme, cycle spinning can be incorporated as follows:

$$\hat{\mathbf{x}}_N = \frac{1}{N} \sum_{h=0}^{N-1} D_{-h} (IDWT(FW(DWT(D_h(\mathbf{y}))))). \quad (9)$$

Here, the noisy image (\mathbf{y}) is first shifted using a diagonal shifting operator D_h . The DWT of the resulting image is then computed and the fractal-wavelet (FW) denoising scheme is then applied. The inverse DWT is then computed and the image is unshifted. This process is repeated for each shift, $h = 0, 1, \dots, N-1$. The respective results are then averaged to obtain one enhanced and denoised estimate with hopefully weaker artifacts. Similarly, one may also incorporate cycle spinning within the (pixel-based) predictive fractal denoising scheme of [9]. The main difference is that there is no need to compute the wavelet transform of the image.

E. Experimental Results

Table I illustrates the results obtained by two versions (fixed and quadtree partitioning) of the predictive wavelet and pixel-based fractal schemes as well as *BayesShrink* and *OracleShrink* wavelet thresholding methods [4], before and after cycle spinning, as applied to the *Lena* image. *BayesShrink* is a spatially adaptive wavelet thresholding method based on context modeling to adapt the threshold selection to changing image characteristics, and *OracleShrink* is the optimal level-wise uniform thresholding technique (in the MSE sense) when the original noiseless image is assumed to be known [4]. Similar qualitative results were obtained for a number of other test images.

Before incorporating cycle spinning, we note that the proposed FW predictive scheme yields results that are comparable to those obtained by the pure (in the pixel domain) fractal-based denoising schemes as well as the standard wavelet thresholding methods. In terms of PSNR, the quadtree-based FW predictive scheme yields higher values than *BayesShrink*. For the quadtree-based FW scheme, we have also found it beneficial to threshold the stored wavelet coefficients that are located at the first few decomposition levels. A neighborhood-based, level-dependent thresholding strategy was applied, as proposed in [11]. A collage error decomposition criterion for the quadtree-based FW predictive scheme was also employed.

For a more equitable comparison between the fractal-wavelet and the wavelet thresholding methods for image denoising, we need to also assess and compare their respective computational requirements. As detailed at the beginning of this section, for the FW denoising scheme, each child subtree, Y_k , $k = 1, 2, \dots, 2^{k_2^*}$, is matched and compared with each potential parent subtree, X_i , $i = 1, 2, \dots, 2^{k_1^*}$. The noiseless statistics in (6) are then estimated. If these estimates are robust, as defined in (7), the noiseless scaling coefficient and the collage error, as given in (2) and (4), respectively, are then computed. When the robustness condition in (7) fails, the scaling coefficient and the collage error are computed from the noisy statistics, except for an adjustment of the scaling coefficient, as given by (8). This completes prediction of the FW code of the original image. During the fast FW decoding process, the truncated wavelet coefficients are extrapolated from the stored wavelet coefficients using the predicted FW code by scaling and copying of

TABLE I
COMPARISON BETWEEN THE RESULTS OBTAINED BY THE VARIOUS IMAGE DENOISING METHODS BEFORE AND AFTER INCORPORATING THE CYCLE SPINNING ALGORITHM WITH $N = 16$ SHIFTS, FOR THE NOISY *LENA* IMAGE WITH $\sigma_w = 25$

Denoising Method	Before Cycle Spinning		After Cycle Spinning	
	RMSE	PSNR	RMSE	PSNR
Predictive FW scheme with $(k_1^*, k_2^*) = (5, 6)$	10.31	27.86	8.53	29.51
Predictive quadtree-based FW scheme with collage error decomposition criterion.	9.24	28.82	8.37	29.68
Predictive pixel-based fractal scheme with uniform partitioning: $(M, N) = (32, 64)$	10.03	28.10	8.39	29.66
Predictive pixel-domain quadtree-based fractal scheme with collage error decomposition criterion.	9.10	28.95	8.09	29.97
Soft thresholding BayesShrink method	9.93	28.19	8.66	29.38
Soft thresholding OracleShrink method	8.88	29.16	8.41	29.63

the parent subtrees to lower child subtrees, as detailed in [5], [6], and [12]. This summarizes the computational requirement involved in the proposed FW denoising scheme.

In summary, the *BayesShrink* wavelet thresholding technique [4], performs soft thresholding using an adaptive, subband (*sub*) and decomposition level (*j*)-dependent near optimal threshold given by

$$\hat{\lambda}_j^{\text{sub}*} = \begin{cases} \frac{\hat{\sigma}_w^2}{\hat{\sigma}_{\mathbf{X}_j^{\text{sub}}}^2}, & \text{if } \hat{\sigma}_w^2 < \hat{\sigma}_{\mathbf{Y}_j^{\text{sub}}}^2 \\ \max_{m=1,2,\dots,M_j} \{|Y_{j,m}^{\text{sub}}|\}, & \text{otherwise} \end{cases} \quad (10)$$

where $\mathbf{X}_j^{\text{sub}}$ and $\mathbf{Y}_j^{\text{sub}}$ denote the original and noisy wavelet coefficients located in subband, *sub*, and decomposition level, *j*, and $\hat{\sigma}_{\mathbf{X}_j^{\text{sub}}}$ is estimated by

$$\hat{\sigma}_{\mathbf{X}_j^{\text{sub}}} = \sqrt{\max(\hat{\sigma}_{\mathbf{Y}_j^{\text{sub}}}^2 - \hat{\sigma}_w^2, 0)}. \quad (11)$$

Clearly, the computational complexity of *BayesShrink* is lower than that of the predictive FW denoising scheme. Quantitatively, it was observed that *BayesShrink* executed in 14 s while the FW denoising scheme required 63 s to execute, using a Pentium 4, 1.9-GHz computer system platform. It should be noted that the implemented computer programs were not necessarily optimized to yield the fastest execution times. It should also be noted that the fractal-wavelet transforms were originally developed for the purpose of image coding and compression. Thus, the FW denoising method may also be optimized to yield simultaneous image compression and denoising, while wavelet thresholding methods only aim for noise reduction.

After incorporating cycle spinning, we note that the performance of the proposed fractal denoising methods slightly surpasses that of two wavelet thresholding methods (especially *BayesShrink*), at least as far as PSNR is concerned. It should

be mentioned that, in spite of the obvious advantages of cycle spinning, clearly, this algorithm may be rather computationally expensive. Indeed, when incorporating this algorithm with N shifts within any fractal or wavelet-based denoising method, the computational complexity is multiplied by a factor of N .

III. EXAMINING THE PERFORMANCE OF THE FW CODE PREDICTION

We now examine how well the predicted FW code obtained from the noisy test image of *Lena*, as corrupted by an AWGN noise with noise intensity $\sigma_w = 25$, compares with the exact FW code of the original (noiseless) image of *Lena*, which is assumed to be known for comparison purposes only.

Fig. 2(a) and (b) illustrates the distribution of the scaling coefficients, α , corresponding to the original and noisy images as well as the predicted scaling coefficients obtained using the standard FW predictive scheme with $(k_1^*, k_2^*) = (5, 6)$. Note that the bimodal distribution of the noisy image scaling coefficients (with no zero scaling coefficients) differs significantly distinct from that of the noiseless case. This difference can be explained in the following way: In the FW coding of a noise-free image, flat regions yield low energy subtrees with mostly zero or near-zero coefficients. This results in zero or near-zero scaling coefficients α . In the presence of noise, however, these regions can become dominated by the (higher frequency) noise. The energies of the corresponding wavelet subtrees are increased, resulting in an increase of the magnitudes of the α coefficients.

As illustrated in Fig. 2(c), the distribution of the predicted scaling coefficients remains bimodal, although to a lesser degree than for the noisy image case. This is because many flat subregions of low activity yield child subtrees dominated by noise. As explained above, these low-energy subtrees are simply fractally coded and then their corresponding scaling coefficients are reduced, as given in (8). Consequently, the distribution of the

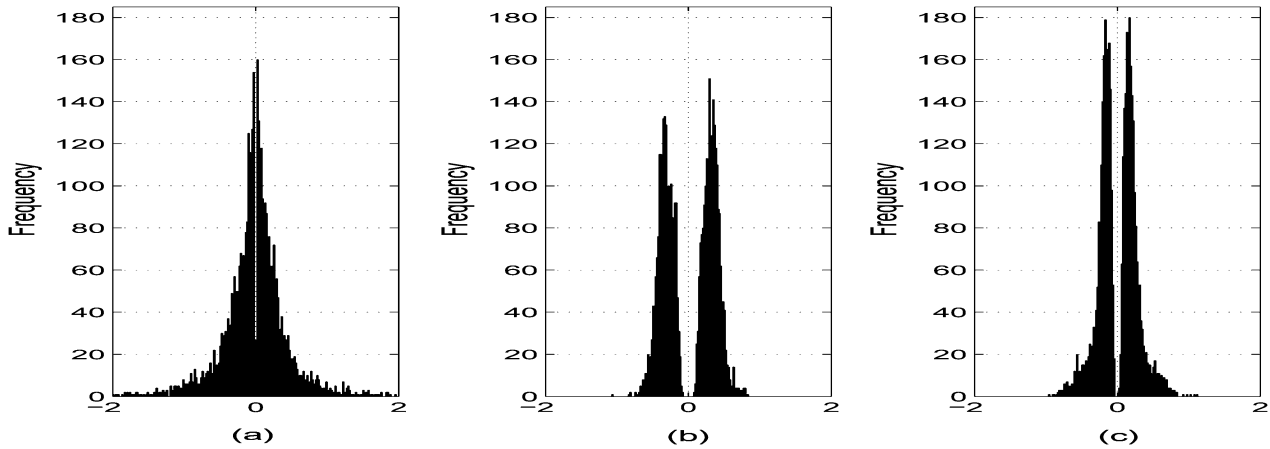


Fig. 2. Comparison between the scaling coefficients corresponding to (a) the original image, (b) the noisy image, and (c) the predicted coefficients.

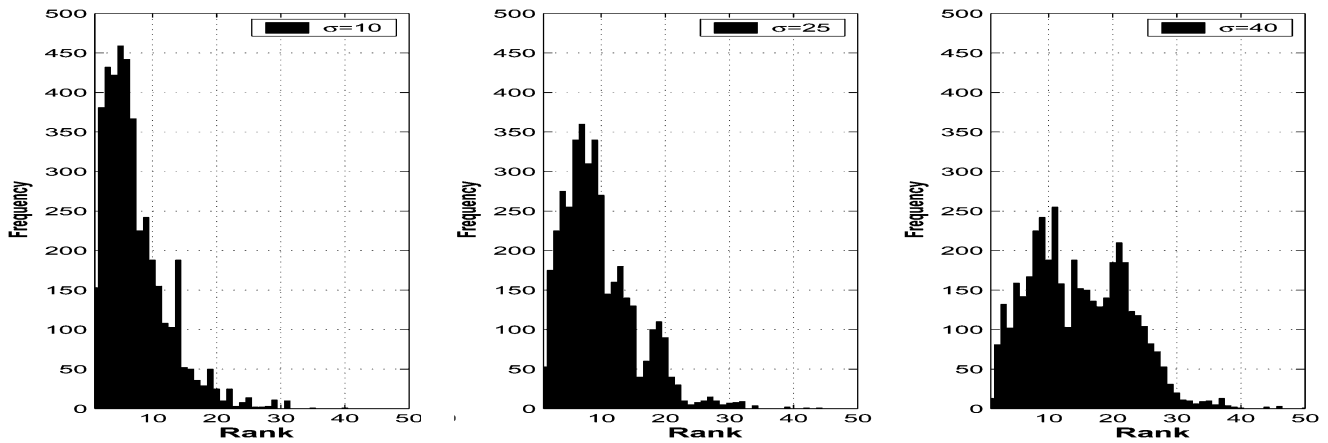


Fig. 3. Histogram of the rank of the predicted parent subtrees corresponding to the noisy test image corrupted with AWGN noise with various intensities $\sigma = 10, 25, 40$. A rank of 1 means that the optimal parent subtree was predicted, the rank is 2 when the second best parent subtree was predicted, etc. For each child subtree, a domain pool of 1024 parent subtrees was examined.

predicted scaling coefficients is now closer to that of the exact scaling coefficients corresponding to the noiseless image.

For a large number of images, we have observed that for most child subtrees, a good number of parent subtrees yield collage errors that are quite close to the minimum collage error. This is in accord with observations made in [1]. In general, therefore, one does not need to predict the optimal parent subtree corresponding to each child subtree in order to obtain reasonably good suboptimal FW codes. Fig. 3 shows the histogram distribution of the rank of the predicted parent subtrees for different values of the noise intensity σ_w . For the chosen FW resolution, $(k_1^*, k_2^*) = (5, 6)$, all 4096 child subtrees were examined. For each child subtree, all 1024 potential parent subtrees were tested, and the predicted parent subtree was ranked with respect to the collage distances of parent subtrees of the noiseless image. For example, a rank of one means that the best parent tree was selected by the prediction; a rank of ten means that there were nine parent subtrees for which the collage error was lower. We note that most of the predicted parent subtrees are among the first few closest parent subtrees, especially for smaller noise intensity. As expected, the histogram distribution spreads out as the noise variance increases since it becomes more difficult to estimate the noiseless collage errors from the noisy data since

the robustness criteria in (7) are less likely to hold in the presence of higher intensity noise. The fact that the predicted parent subtrees are usually among the first few closest (i.e., near-optimal) parent subtrees, at least for low noise variance, accounts for the good estimates of the noiseless image provided by the predictive FW schemes.

Finally, we should mention that performing such a fractal coding predictive scheme in the *pixel* domain yields similar results [7]. For low noise variance, the predicted parent subblocks rank among the best subblocks. The histogram distributions of the ranks are virtually identical in structure.

IV. PIXEL VERSUS WAVELET-BASED FRACTAL PREDICTIVE SCHEMES: A COMPARISON

Table II illustrates the results of applying the pixel- and wavelet-based fractal predictive schemes to denoise three different test images corrupted by an AWGN noise with different noise intensities, σ_w . Note that the pixel-based fractal predictive image denoising scheme performs consistently better than the fractal-wavelet method, before and after the use of cycle spinning. Possible explanations include the following.

- The pixel-based fractal predictive scheme employs a set of contractive geometric maps that geometrically transform

TABLE II
COMPARISON BETWEEN THE RESULTS OBTAINED BY THE PREDICTIVE FRACTAL (PIXEL-BASED) AND FRACTAL-WAVELET SCHEMES, BEFORE AND AFTER APPLYING CYCLE SPINNING WITH $K = 16$ DIAGONAL SHIFTS, FOR A FEW TEST IMAGES

		Predictive Fractal Scheme				Predictive Fractal-Wavelet Scheme			
Cycle Spinning		Before		After		Before		After	
Image	σ_w	RMSE	PSNR	RMSE	PSNR	RMSE	PSNR	RMSE	PSNR
<i>Lena</i>	10	8.05	30.01	6.13	32.38	8.10	29.96	7.54	30.59
	20	9.53	28.55	7.79	30.29	9.87	28.24	8.37	29.68
	30	10.89	27.39	8.57	29.47	11.09	27.23	9.16	28.89
	40	12.11	26.47	9.79	28.32	12.95	25.89	10.31	27.86
<i>Boat</i>	10	9.24	28.81	8.65	29.39	9.68	28.42	9.06	28.99
	20	10.67	27.56	8.89	29.15	10.86	27.41	9.15	28.90
	30	12.10	26.47	10.41	27.78	12.78	26.00	11.00	27.31
	40	13.67	25.41	11.87	26.64	14.11	25.14	12.26	26.35
<i>Peppers</i>	10	7.83	29.21	6.47	31.09	8.08	29.99	7.51	30.61
	20	8.61	29.43	7.44	30.70	9.02	29.02	7.78	30.31
	30	10.65	27.58	8.94	29.10	11.07	27.25	9.32	28.74
	40	11.72	26.75	10.22	27.94	12.64	26.09	10.66	27.57

and decimate a parent pixel subblock so that the subimage it supports can be deformed to match the image supported on a child subblock. In fact, as discussed in [9], the decimation associated with the contractive spatial maps is probably responsible for most of the denoising. However, the fractal-wavelet scheme uses no such smoothing or preprocessing step before fitting a parent subtree to a child subtree. This may indeed be the main reason behind the difference in the performance of the two schemes.

- Another significant feature of the pixel-based fractal denoising scheme is that it adopts a quantization strategy of the gray-level coefficients that yields a fractally denoised estimate with pixel values that lie in the range $[0, 255]$. As the noise variance increases, it becomes more difficult for the FW predictive scheme, as well as other conventional denoising methods, to yield denoised estimates that satisfy this requirement.
- In the fractal-wavelet denoising scheme, a set of noisy wavelet coefficients is stored at the top of the wavelet coefficient tree. These coefficients, along with the FW code, are then used by the FW decoder to estimate the remaining wavelet coefficients. As a result, the noise in the stored coefficients is redistributed among the predicted coefficients. Although the stored wavelet coefficients are located in the higher decomposition level (lower scale/resolution, top left corner of the matrix in Fig. 1) which are smoothed by the wavelet filters, the role of these coefficients in redistributing the noise may be significant. In an effort to reduce this problem, we examined the use of a context-based thresholding method [11] that was applied to some of the stored wavelet coefficients. An improvement in results was noted when the quadtree-based fractal-wavelet scheme was used. However, the method was counterproductive in the

case of the standard FW scheme [fixed parent-child levels (k_1^*, k_2^*)].

V. SUMMARY AND CONCLUDING REMARKS

The aim of this work was to shed some further insight into the fractal-based image denoising methods previously proposed in [8], [9]. The essence of fractal-based denoising, both in the wavelet as well as pixel domains, is to predict the fractal code of a noiseless image from its noisy observation. We have experimentally shown that the fractal-wavelet denoising scheme is able, at least for moderate noise variances, to locate near-optimal parent subtrees that lie among the best domain subtrees in terms of collage distance. The procedure is assisted by the high degree of local self-similarity of an image. In general, in the pixel domain, a good number of domain subblocks approximate a given range subblock very well. And the fractal-based denoising method works well in finding one of these subblocks. This is also reflected in the wavelet domain.

We have incorporated cycle spinning into these fractal-based denoising methods to produce enhanced estimates of the denoised image. We have also found that pixel-based fractal denoising schemes consistently perform slightly better, in terms of PSNR, than their wavelet-based counterparts (this does *not*, however, imply that the visual quality of the images is higher). Some possible explanations have been forwarded. However, the significant computational savings of the wavelet-based fractal denoising scheme may make it advantageous to use.

In [9], the pixel-based predictive fractal denoising scheme was shown to be competitive with some of the standard pixel-based image denoising schemes such as the Lee filter [13]. In terms of PSNR, the wavelet-based predictive fractal denoising schemes yield results that are competitive with the efficient wavelet thresholding methods, such as *BayesShrink*

and *OracleShrink*. This is particularly the case when using the quadtree partitioning of the wavelet decomposition tree and even without cycle spinning. Our FW denoising scheme is computationally more expensive, roughly by a factor of four when no cycle spinning is employed. However, we believe that this factor could be reduced significantly with a more efficient coding of the method. In conclusion, we hope that our work, along with [8], [9], and [12], has shown that the scope of applicability of fractal-based methods has been broadened.

REFERENCES

- [1] S. Alexander, "Multiscale Methods in Image Modelling and Image Processing," Ph.D. dissertation, Dept. Appl. Math., Univ. Waterloo, Waterloo, ON, Canada, 2005.
- [2] K. U. Barthel, H. L. Cycon, and D. Marpe, "Image denoising using fractal and wavelet-based methods," *Proc. SPIE*, vol. 5266, pp. 10–18, 2003.
- [3] R. R. Coifman and D. L. Donoho, A. Antoniadis, G. Oppenheim, and editors, Eds., "Translation-invariant denoising," in *Wavelets and Statistics*. New York: Springer-Verlag, 1995, vol. 103, pp. 125–150.
- [4] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive image thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.
- [5] G. Davis, "A wavelet-based analysis fractal image compression," *IEEE Trans. Image Process.*, vol. 7, no. 2, pp. 141–154, Feb. 1998.
- [6] B. Forte and E. R. Vrscay, "Inverse problem methods for generalized fractal transforms," in *Fractal Image Encoding and Analysis*, ser. NATO ASI F 159, Y. Fisher, Ed. New York: Springer-Verlag, 1998.
- [7] M. Ghazel, "Adaptive Fractal and Wavelet Image Denoising," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Waterloo, Waterloo, ON, Canada, 2004.
- [8] M. Ghazel, G. H. Freeman, and E. R. Vrscay, "Fractal-wavelet image denoising," in *Proc. IEEE Int. Conf. Image Processing*, 2002, pp. 836–839.
- [9] —, "Fractal image denoising," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1560–1578, Dec. 2003.
- [10] —, "An effective hybrid fractal-wavelet image coder using quadtree partitioning and pruning," presented at the IEEE CCECE Halifax, NS, Canada, May 7–11, 2000.
- [11] —, "Wavelet thresholding for image denoising using localized thresholding operators," presented at the ICIAR Toronto, ON, Canada, 2005.
- [12] H. Krupnik, D. Malah, and E. Karnin, "Fractal representation of images via the discrete wavelet transform," presented at the IEEE Conv. Elect. Eng. Tel-Aviv, Israel, Mar. 7–9, 1995.
- [13] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Pattern Anal. Mach. Intell.*, vol. 2, no. 2, pp. 165–168, Feb. 1980.
- [14] S. G. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.
- [15] E. R. Vrscay, "A generalized class of fractal-wavelet transforms for image representation and compression," *Canad. J. Elect. Comput. Eng.*, vol. 23, no. 1–2, pp. 69–84, 1998.

Mohsen Ghazel received the B.Sc. degree in engineering mathematics from the University of Arizona, Tucson, in 1992, the M.Math. degree in applied mathematics, and the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 1994 and 1999, respectively.

He is currently a Postdoctoral Researcher at the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC.

George H. Freeman received the B.A.Sc. and Ph.D. degrees in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1979 and 1984, respectively.

He then joined the Continuous Speech Recognition Group, IBM T. J. Watson Research Center, Yorktown Heights, NY, where he worked on statistical language modelling and array signal processing algorithms for speech recognition. In 1985, he returned to the University of Waterloo, where he is currently an Associate Professor and Associate Chair for Undergraduate Studies in the Department of Electrical and Computer Engineering. A one-year sabbatical leave from 1993 to 1994 was spent working with NCR Canada, Ltd., Waterloo, exploring fractal and wavelet methods for document image compression. Past research interests have also included lossless data compression, trellis quantization, and resource management for multimedial wideband CDMA systems. He is now interested in wavelet analysis for biomedical image processing, particularly transectal ultrasound images of the prostate gland, and articulatory models for speech recognition.

Edward R. Vrscay received the B.Sc. degree in chemistry and the M.Math. and Ph.D. degrees in applied mathematics from the University of Waterloo, Waterloo, ON, Canada, in 1975, 1977, and 1984, respectively.

From 1984 to 1986, he was an NSERC Postdoctoral Fellow and Visiting Assistant Professor at the School of Mathematics, Georgia Institute of Technology (Georgia Tech), Atlanta. In 1986, he returned to the Department of Applied Mathematics, University of Waterloo, as an NSERC University Fellow. His original research interests were in mathematical physics, in particular, quantum mechanics. At Georgia Tech, he became involved with dynamical systems, fractal geometry, and fractal image coding. He has worked on the mathematical formulation and solution of inverse problems for fractal and fractal-wavelet approximation and is now interested in other problems of mathematical imaging, including image superresolution, and coregistration. He is the Co-Editor (with M. Barnsley and D. Saupe) of the book *Fractals in Multimedia* (Springer Verlag, 2002), a collection of invited papers at a workshop sponsored by the Institute of Mathematics and its Applications, University of Minnesota, January 2001, as part of the IMA's Mathematics in Multimedia Program, 2000 to 2001.