SIAM J. IMAGING SCIENCES
Vol. 2, No. 2, pp. 470–507

Measure-Valued Images, Associated Fractal Transforms, and the Affine Self-Similarity of Images*

D. La Torre†, E. R. Vrscay‡, M. Ebrahimi§, and M. F. Barnsley¶

Abstract. We construct a complete metric space (Y, d_Y) of measure-valued images, $\mu : X \to \mathcal{M}(\mathbb{R}_g)$, where X is the base or pixel space and $\mathcal{M}(\mathbb{R}_g)$ is the set of probability measures supported on the greyscale range $\mathbb{R}_g$. Such a formalism is well suited to nonlocal (NL) image processing, i.e., the manipulation of the value of an image function $u(x)$ based upon values $u(y_k)$ elsewhere in the image. We then show how the space (Y, d_Y) can be employed with a general model of affine self-similarity of images that includes both same-scale as well as cross-scale similarity. We focus on two particular applications: NL-means denoising (same-scale) and multiparent block fractal image coding (cross-scale). In order to accommodate the latter, a method of fractal transforms is formulated over the metric space (Y, d_Y). Under suitable conditions, a transform $M : Y \to Y$ is contractive, implying the existence of a unique fixed point measure-valued function $\bar{\mu} = M\bar{\mu}$. We also show that the pointwise moments of this measure satisfy a set of recursion relations that are generalizations of those satisfied by moments of invariant measures of iterated function systems with probabilities.

Key words. measure-valued images, multifunctions, nonlocal image processing, self-similarity, nonlocal-means denoising, fractal transforms, iterated function systems

AMS subject classifications. 28B20, 94A08, 68U10

DOI. 10.1137/070696763

1. Introduction.

Most practical as well as theoretical works in image processing and mathematical imaging consider images as real-valued functions, $u : X \to \mathbb{R}_g$, where X denotes the *base space* or *pixel space* over which the images are defined and $\mathbb{R}_g \subset \mathbb{R}$ is a suitable *greyscale space*. A variety of function spaces $u \in \mathcal{F}(X)$ may be considered depending on the application. In standard image processing schemes, $\mathcal{F}(X)$ is usually assumed to be $\mathcal{L}^2(X)$, the space of square-integrable functions on X.

There are, however, situations in which it is useful to consider the greyscale value of an image u at a point x as a random variable that can assume a range of values $\mathbb{R}_g \subset \mathbb{R}$.

*Received by the editors July 11, 2007; accepted for publication (in revised form) December 17, 2008; published electronically April 22, 2009. This work was supported in part by a Discovery grant (ERV) from the Natural Sciences and Engineering Research Council of Canada (NSERC).
http://www.siam.org/journals/siims/2-2/69676.html
†Department of Economics, Business and Statistics, University of Milan, 20122 Milan, Italy (davide.latorre@unimi.it).
‡Department of Applied Mathematics, University of Waterloo, Waterloo N2L 3G1, ON, Canada (ervrscay@uwaterloo.ca).
§Department of Applied Mathematics, University of Waterloo, Waterloo N2L 3G1, ON, Canada. Current address: Imaging Research Group, Sunnybrook Health Sciences Centre, 2075 Bayview Ave., Toronto M4N 3M5, ON, Canada (mehran@sri.utoronto.ca). This author's research was supported by the Province of Ontario (in the form of an Ontario Graduate Scholarship) as well as the Faculty of Mathematics, University of Waterloo.
¶Department of Mathematics, Australian National University, Canberra, ACT 0200, Australia (mbarnsley@aol.com).

One example is the characterization of the statistical properties of a class of images, e.g., MRI brain scans, for a particular application, say, image compression. Another example is statistical image processing as applied to the problem of image restoration (denoising or deblurring). Of course, it is not enough to know the greyscale values that may be assumed by an image u at a point x: One must also have an idea of the probabilities (or frequencies) of these values. As such, it may be more useful to represent images by *measure-valued* functions, for example, $\mu : X \to \mathcal{M}(\mathbb{R}_g)$, where $\mathcal{M}(\mathbb{R}_g)$ is the set of probability measures supported on $\mathbb{R}_g$. This is a particular example of an image being represented by a *multifunction*. In this paper, we employ such a multifunction/measure-valued representation of images for the purpose of *nonlocal (NL) image processing*, that is, the manipulation of the value of an image function $u(x)$ based upon values of $u(y_k)$ elsewhere in the image.

NL image processing has recently received a great deal of attention, fueled in part by the exceptional success of the *NL-means* image denoising method [7]. *Fractal image coding* [20, 3] is another example of an NL image processing method. Both of these methods, which will be described briefly below, may be viewed under the umbrella of a more general model of affine image self-similarity [2], in which subblocks of an image are approximated by other subblocks of the image. Indeed, a number of other image processing methods that exploit self-similarity, for example, [26] and the various *example-based methods* [9, 10, 16], also fit naturally under this NL, self-similar framework.

Up to now, the functional representation of images has been used very successfully in the above schemes. The general procedure is as follows: Given an image function $u : X \to \mathbb{R}_g$, the value $u(x)$ is replaced by a transformed value $Tu(x)$ which is determined by one or more values $u(y_k)$, where the points $y_k = y_k(x)$ lie elsewhere in the image and not necessarily close to x. (This is in contrast to standard image processing methods which are *local* in nature, i.e., the points y_k lie in a neighborhood of x.) The purpose of this paper is to show how measure-valued images may be useful in these schemes, for at least a couple of reasons that are not unrelated:

1. as a kind of intermediate step before the final "projection" of values $u(y_k)$ to produce the transformed value $Tu(x)$;
2. using the measure $\mu(x)$ to characterize the *local self-similarity* of the image I at a point $x \in X$.

One of the main results of this paper is that we formulate a method of *fractal transforms* over an appropriate space of measure-valued functions.

The structure of this paper is as follows. In section 2, we construct an appropriate complete metric space (Y, d_Y) of measure-valued images $\mu : [0,1]^n \to \mathcal{M}(\mathbb{R}_g)$. In section 3, we outline the general model of affine image self-similarity and how it includes NL-means denoising and fractal image coding as special examples. We then illustrate how the space (Y, d_Y) of measure-valued images can easily be employed in "nonfractal," i.e., *same-scale*, affine self-similarity models. In particular, we consider the NL-means denoising method. The full mathematical structure of the space (Y, d_Y) is not utilized in these methods, however, since they are not iterative in nature.

In section 4, we formulate a method of fractal transforms over the metric space (Y, d_Y). The iterative nature of this *cross-scale* self-similarity method utilizes more of the mathematical structure of the space. Under suitable conditions, a fractal transform operator $M : Y \to Y$

is contractive, implying the existence of a unique fixed point measure $\bar{\mu} = M\bar{\mu}$. We also show that the pointwise moments of this measure satisfy a set of recursion relations that can be viewed as generalizations of the relations satisfied by moments of invariant measures for iterated function systems with probabilities (IFSP).

The results of section 4 may be viewed as a natural extension of previous work [23] in which a method of fractal transforms was formulated over *image multifunctions*, that is, set-valued mappings.

For the benefit of readers who are not very familiar with fractal-based methods of coding, we present some basics of the method of "iterated function systems" (IFS) in the appendix. It is necessary to present the formulation of IFS-based methods over both image and measure spaces, since the method of fractal transforms over the space (Y, d_Y) presented in section 4 represents a kind of "fusion" of these two methods.

2. A complete metric space (Y, d_Y) of measure-valued images. We first set up our space of measure-valued images. In what follows, $X = [0,1]^n$ will denote the "base space," i.e., the support of the images. $\mathbb{R}_g \subset \mathbb{R}$ will denote a compact "greyscale space" of values that our images can assume at any $x \in X$. (The following discussion is easily extended to $\mathbb{R}_g \subset \mathbb{R}^m$ to accommodate color images, etc.) Now let $\mathcal{M}(\mathbb{R}_g)$ denote the set of all Borel probability measures on $\mathbb{R}_g$ and let d_H denote the Monge–Kantorovich metric on this set:

$$d_H(\mu,\nu) = \sup_{f \in Lip_1(X,\mathbb{R})} \left[\int_X f d\mu - \int_X f d\nu \right], \tag{1}$$

where

$$Lip_1(X,\mathbb{R}) = \{f : X \to \mathbb{R} \mid |f(x_1) - f(x_2)| \le d(x_1,x_2)\ \forall x_1, x_2 \in X\}. \tag{2}$$

For a given $M > 0$, let $\mathcal{M}_1 \subset \mathcal{M}(\mathbb{R}_g)$ be a complete subspace of $\mathcal{M}(\mathbb{R}_g)$ such that $d_H(\mu,\nu) \le M$ for all $\mu, \nu \in \mathcal{M}_1$. We now define

$$Y = \{\mu(x) : X \to \mathcal{M}_1, \mu(x) \text{ is measurable}\} \tag{3}$$

and consider on this space the metric

$$d_Y(\mu,\nu) = \int_X d_H(\mu(x),\nu(x)) d\mu_{\mathcal{L}}(x). \tag{4}$$

Here $\mu_{\mathcal{L}}$ denotes Lebesgue measure on X. We observe that d_Y is well defined, since μ and ν are measurable functions and d_H is bounded, and so the function $\xi(x) = d_H(\mu(x),\nu(x))$ is integrable on X.

Theorem 2.1. *The space (Y, d_Y) is complete.*

Proof. It is trivial to prove that this is a metric when we consider that $\mu = \nu$ if $\mu(x) = \nu(x)$ a.e. $x \in X$. To prove the completeness we follow the trail of the proof of Theorem 1.2 in [21]. Let μ_n be a Cauchy sequence in Y. Thus for all $\epsilon > 0$ there exists n_0 such that for all $n, m \ge n_0$ we have $d_Y(\mu_n, \mu_m) < \epsilon$. Let $\epsilon = 3^{-k}$ so that you can choose an increasing sequence n_k such that $d_Y(\mu_n, \mu_{n_k}) < 3^{-k}$ for all $n \ge n_k$. Thus choosing $n = n_{k+1}$ we have $d_Y(\mu_{n_{k+1}}, \mu_{n_k}) < 3^{-k}$. Let

$$A_k = \{x \in [0,1] : d_H(\mu_{n_{k+1}}(x), \mu_{n_k}(x)) > 2^{-k}\}. \tag{5}$$

Then

$$\mu_{\mathcal{L}}(A_k)2^{-k} \le \int_{A_k} d_H(\mu_{n_{k+1}}(x), \mu_{n_k}(x))d\mu_{\mathcal{L}} \le 3^{-k} \tag{6}$$

so that $\mu_{\mathcal{L}}(A_k) \le \left(\frac{2}{3}\right)^k$. Let $A = \bigcap_{m=1}^{\infty} \bigcup_{k\ge m} A_k$. We observe that

$$\begin{aligned} \mu_{\mathcal{L}}\left(\bigcup_{k\ge m} A_k\right) &\le \sum_{k\ge m} \mu_{\mathcal{L}}(A_k) \\ &\le \sum_{k\ge m} \left(\frac{2}{3}\right)^k = \frac{\left(\frac{2}{3}\right)^m}{1-\left(\frac{2}{3}\right)}. \end{aligned} \tag{7}$$

Therefore

$$\mu_{\mathcal{L}}(A) \le 3\left(\frac{2}{3}\right)^m \tag{8}$$

for all m, which implies that $\mu_{\mathcal{L}}(A) = 0$. Now for all $x \notin X\backslash A$ there exists $m_0(x)$ such that for all $m \ge m_0$ we have $x \notin A_m$ and so $d_H(\mu_{n_{m+1}}(x), \mu_{n_m}(x)) < 2^{-m}$. This implies that $\mu_{n_m}(x)$ is Cauchy for all $x \notin X\backslash A$ and so $\mu_{n_m}(x) \to \mu(x)$ using the completeness of $\mathcal{M}_1$. This also implies that $\mu : X \to Y$ is measurable; that is, $\mu \in Y$. To prove $\mu_n \to \mu$ in Y we have that

$$\begin{aligned} d_Y(\mu_{n_k}, \mu) &= \int_X d_H(\mu_{n_k}(x), \mu(x))d\mu_{\mathcal{L}} \\ &= \int_X \lim_{i\to+\infty} d_H(\mu_{n_k}(x), \mu_{n_i}(x))d\mu_{\mathcal{L}} \\ &\le \liminf_{i\to+\infty} \int_X d_H(\mu_{n_k}(x), \mu_{n_i}(x))d\mu_{\mathcal{L}} \\ &= \liminf_{i\to+\infty} d_Y(\mu_{n_k}, \mu_{n_i}) \le 3^{-k} \end{aligned} \tag{9}$$

for all k. So $\lim_{k\to+\infty} d_Y(\mu_{n_k}, \mu) = 0$. Now we have

$$d_Y(\mu_n, \mu) \le d_Y(\mu_n, \mu_{n_k}) + d_Y(\mu_{n_k}, \mu) \to 0 \tag{10}$$

when $k \to +\infty$. ■

3. A simple class of models for image self-similarity. In this section, for simplicity, we consider images to be $n_1 \times n_2$-pixel arrays. (The extension to continuous support, i.e., $X = [0,1]^2$, is rather straightforward.) In all computations, we work with normalized images, i.e., $\mathbb{R}_g = [0,1]$. The components of our model, introduced in [2], are as follows:

1. A set $\mathcal{R}$ of $n\times n$-pixel *range* subblocks R_i, $1 \le i \le N_R$, such that (i) $R_i \cap R_j = 0$ if $i \ne j$ and (ii) $X = \cup_i R_i$. In other words, $\mathcal{R}$ forms a partition of X. We let $u(R_i) = u|_{R_i}$ denote the portion of u that is supported on R_i.
2. A set $\mathcal{D}$ of $m \times m$-pixel *domain* subblocks D_j, where $m \ge n$. The set of blocks $\mathcal{D}$ should cover X, i.e., $\cup_j D_j = X$, but they need not be nonoverlapping.

3. The one-to-one geometric transformations $w_{ij}^{(k)}$ that map a domain block D_j to range block R_i. For simplicity, we consider only *affine* transformations. Since both blocks are square, there are eight possible mappings (four rotations and four inversions about the center) which are accommodated in the index $1 \le k \le 8$. (For notational convenience, the k superscripts will be omitted unless absolutely necessary to the discussion.) In the case that $m > n$, i.e., D_j is larger than R_i, it is also assumed that the *contractive* map w_{ij} includes an appropriate pixel decimation operation.
4. Affine greyscale maps $\phi : \mathbb{R}_g \to \mathbb{R}_g$ having the form $\phi(t) = \alpha t + \beta$.

Given an image function u, we examine the degree to which subimages $u(R_i)$ are approximated by subimages $u(D_j)$, i.e.,

$$u(R_i) \approx \phi_{ij}(u(w_{ij}^{-1}(R_i))) = \alpha_{ij} u(w_{ij}^{-1}(R_i)) + \beta_{ij}, \tag{11}$$

which we shall write symbolically as

$$u(R_i) \approx \alpha_{ij} u(D_j) + \beta_{ij}. \tag{12}$$

The error Δ_{ij} associated with this approximation is given by

$$\Delta_{ij} = \min_{\alpha,\beta \in \Pi} \|u(R_i) - \alpha u(D_j) - \beta\|. \tag{13}$$

Here, $\|\cdot\|$ denotes the $\mathcal{L}^2(X)$ norm. In all calculations reported in this paper, the $\mathcal{L}^2$ distance between two $n \times n$ image subblocks $u(R_i)$ and $v(R_i)$ will be the root-mean-square (RMS) distance. $\Pi \subset \mathbb{R}^2$ denotes the feasible (α, β) parameter space which is appropriate for the self-similarity scheme being examined (one of the four cases listed below) and which guarantees that $\phi : \mathbb{R}_g \to \mathbb{R}_g$. We assume that Π is compact.

There are four important cases of this self-similarity model:

1. *Purely translational.* Domain and range blocks have the same size, i.e., $m = n$. As such, the w_{ij} are translations, and $\alpha_{ij} = 1$, $\beta_{ij} = 0$. The approximation error is simply

$$\Delta_{ij}^{(Case\ 1)} = \|u(R_i) - u(D_j)\|. \tag{14}$$

2. *Translational + greyscale shift.* The w_{ij} are again translations. We set $\alpha_{ij} = 1$ and optimize over β:

$$\beta_{ij} = \bar{u}(R_i) - \bar{u}(D_j), \qquad \Delta_{ij}^{(Case\ 2)} = |\beta_{ij}|, \tag{15}$$

where the bars denote mean values of the subblocks.

3. *Affine, same-scale.* The w_{ij} are translations, and we optimize over α, β. The standard regression formulas,

$$\alpha_{ij} = \frac{Cov(u(R_i), u(R_j))}{Var(u(R_i))}, \qquad \beta_{ij} = \bar{u}(R_i) - \alpha_{ij}\bar{u}(R_j), \tag{16}$$

apply in the case that $(\alpha_{ij}, \beta_{ij}) \in \Pi$, the feasible parameter space for the problem. In this case,

$$\Delta_{ij}^{(Case\ 3)} = \left[Var(u(R_i)) - \alpha_{ij}^2 Var(u(R_j))\right]^{1/2}. \tag{17}$$

Otherwise, the problem is a quadratic programming problem over Π.

4. *Affine, two-scale.* The w_{ij} are affine spatial contractions (which involve decimations in pixel space). We optimize over α, β. If applicable, the formulas for α_{ij}, β_{ij}, and $\Delta_{ij}^{(Case\ 4)}$ are obtained from (16) and (17) by replacing $u(R_j)$ with $\tilde{u}(D_j)$, the appropriate decimation of $u(D_j)$, so that D_j is mapped onto R_i.

As emphasized in [2], this is a very simple model. Very briefly, the use of square, nonoverlapping blocks of the same size is an effort to standardize the method at a given scale. One may wish, or even need, to examine self-similarity statistics at various scales. And regarding the use of affine greyscale maps: Such a family is simple in form yet sufficiently flexible.

From the viewpoint of computer vision, our model is a *low-level* one since it is based on block similarities and not on patterns or context. We also mention that particular applications may dictate the use of specific cases. For example, it may be permissible (and, as we show, quite effective) to employ affine greyscale maps for the purpose of denoising natural images, e.g., faces or woodland. But in medical imaging applications, only translational symmetry may be permitted. For example, regions of diseased tissue (represented by whiter pixels in a PET image and darker pixels in an MRI image) will be compared only to other regions of diseased tissue.

In [2] we examined the distributions of approximation errors Δ_{ij} for a variety of test images for the purpose of characterizing their overall degrees of self-similarity. In general, as we show below, the Δ-distributions associated with affine greyscale transformations—Cases 3 and 4—demonstrate some degree of peaking. A more concentrated peaking near zero error suggests a greater degree of self-similarity. A very simple, yet important "benchmark" self-similar image is the constant image $u = C$. Here, the Δ-distributions for Cases 1–4 are identical: a single peak at $\Delta = 0$.

The effects of (additive, Gaussian, zero-mean, independent) noise were also investigated in [2]. The Δ-distributions of noisy images are shifted outward from their noiseless counterparts—the greater the variance of the noise, the greater the shift. We shall discuss this feature below, with particular reference to NL-means denoising.

The fact that the Case 3 and Case 4 Δ-error distributions of images generally demonstrate significant peaking near zero error indicates that subblocks of an image are generally well approximated by many other subblocks. This affine self-similarity property can be exploited for image processing purposes, for example, denoising, as we show below.

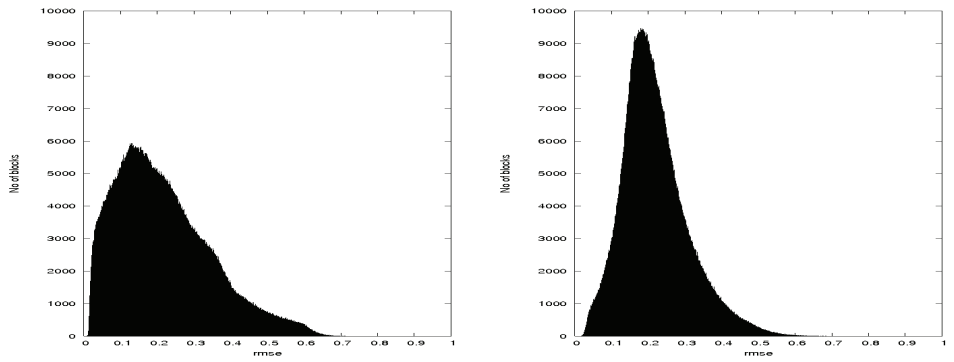
3.1. Cases 1–3: Same-scale self-similarity. In these three cases, the domain and range blocks have the same size. For a given domain/range pairing (D_j, R_i), the approximation errors of (13) are related as

$$0 \le \Delta_{ij}^{(Case\ 3)} \le \Delta_{ij}^{(Case\ 2)} \le \Delta_{ij}^{(Case\ 1)}, \tag{18}$$

since we optimize over more parameters in moving from Case 1 (no parameters) to Case 2 (one parameter) to Case 3 (two parameters).

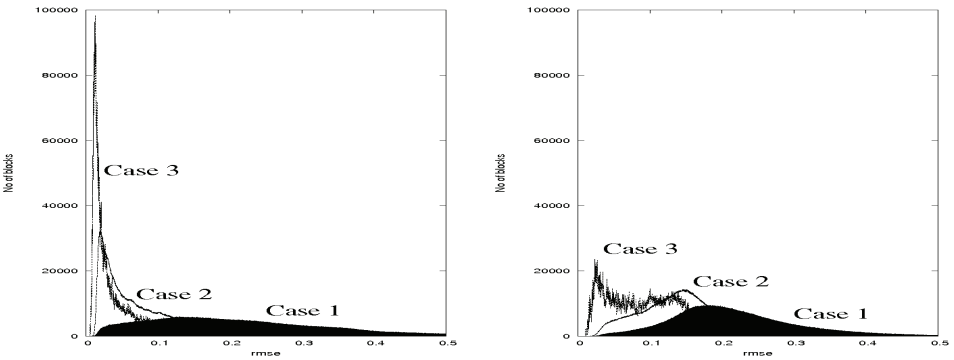
In all numerical results reported below, the domain and range blocks were taken from the same set of nonoverlapping 8×8-pixel blocks of the images examined, i.e., $D_i = R_i$. (Very similar results are obtained if we consider *all* possible 8×8-pixel blocks obtained from single-pixel shifts.)

First we examine the strict translational similarity of some test images, i.e., Case 1. In Figure 1 are plotted histogram distributions of the approximation errors $\Delta_{ij}^{(Case\ 1)}$, which are simply the $\mathcal{L}^2$ distances between subblocks $u(R_j)$ and $u(R_i)$, for the 512×512-pixel normalized test images *Lena* and *Mandrill*. At first glance, it would appear that the two images are quite translationally self-similar since both distributions peak in the interval $[0.1, 0.2]$, that of the *Mandrill* image being more pronounced.



(a) Case 1: *Lena* (b) Case 1: *Mandrill*

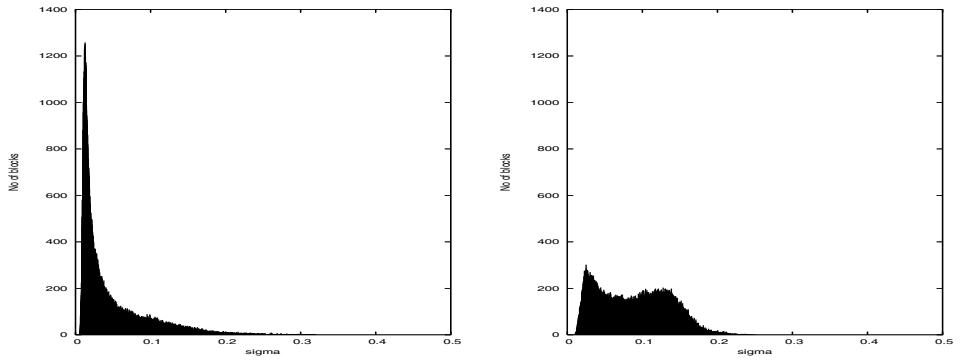
Figure 1. *Case* 1 *error distributions* $\Delta_{ij}^{(Case\ 1)} = \|u(R_j) - u(R_i)\|$, $i \neq j$, *for* 512×512*-pixel normalized* Lena *and* Mandrill *images, over the interval* $\Delta \in [0, 1]$. 8×8*-pixel blocks* R_i.



(a) Cases 1, 2 and 3: *Lena* (b) Cases 1, 2 and 3: *Mandrill*

Figure 2. *Same-scale RMS self-similarity error distributions—Cases* 1, 2, *and* 3*—for normalized* Lena *and* Mandrill *images, over the interval* $\Delta \in [0, 0.5]$. *Case* 1 *distributions from Figure* 1 *are shaded. In all cases,* 8×8*-pixel blocks* R_i *were used.*

In Figure 2 are presented the Δ-error distributions for *Lena* and *Mandrill* for all three cases. The reduction in approximation errors as one moves from Case 1 (shaded) to Cases 2 and 3 is clearly demonstrated, with the latter two distributions exhibiting greater peaking near zero. We observe that enormous improvements are achieved for the *Lena* image in going even from Case 1 to Case 2, where only the greyscale shift parameter β is employed. (Note that the distributions are plotted over the subinterval $[0, 0.5]$.) From these plots, we would conclude



(a) *Lena* (b) *Mandrill*

Figure 3. *Distributions of* $\sigma(u(R_i))$ *of* 8×8*-pixel blocks for normalized* Lena *and* Mandrill *images, over the interval* $[0, 0.5]$. *Note the similarity to Case* 3 *distributions of Figure* 2.

that the two images are more *affinely* self-similar than they are *translationally* self-similar, in particular for the *Lena* image.

In Figure 3 are plotted the histogram distributions of the standard deviations $\sigma(u(R_i))$ of the 8×8 range blocks for the *Lena* and *Mandrill* images. There is a noteworthy similarity between these distributions and the Case 3 distributions of Figure 1 which can be explained as follows. The standard deviation of an image block $\sigma(u(R_i))$ is the RMS error (RMSE) in approximating $u(R_i)$ by its mean value, $\bar{u}(R_i)$. This is equivalent to setting the greyscale parameter $\alpha = 0$ and optimizing over β in (13). Removing the condition $\alpha = 0$ will generally produce better approximations, implying that

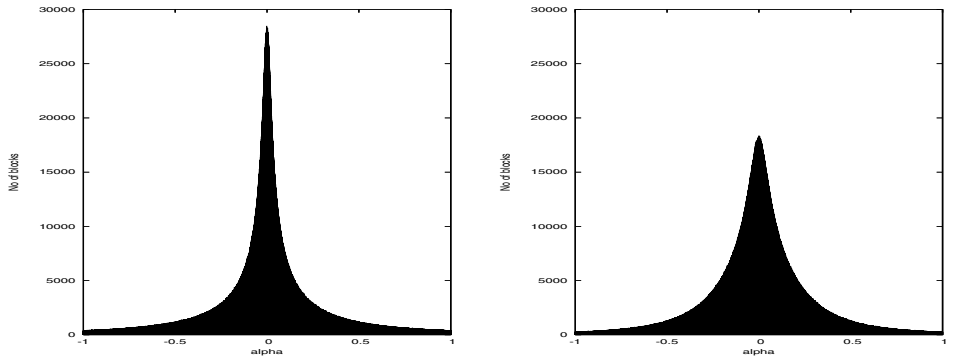
$$0 \leq \Delta_{ij}^{(Case\ 3)} \leq \sigma(u(R_i)). \tag{19}$$

As such, the Case 3 Δ-error distributions will be shifted perturbations of the block variance distributions. That being said, we observe that the distributions of α greyscale coefficients generally exhibit a significant peaking at zero, as shown in Figure 4. As such, the perturbations from the $\sigma(u(R_i))$ distributions to the Case 3 Δ-error distributions will be small.

These observations explain the difference in Case 2 and Case 3 Δ-error distributions between the *Lena* and *Mandrill* images. Figure 3 shows that the *Lena* image contains a significantly higher proportion of "flatter" image subblocks, i.e., blocks of low variance, than the *Mandrill* image. From (19), the Case 2 Δ-error distribution for *Lena* will be more concentrated near zero. Further improvement is expected with Case 3; cf. (18).

Of course, the above discussion suggests that the distribution of block variances is the most important factor in determining the degree of self-similarity of an image I, i.e., how well its subblocks may be approximated by other subblocks. The degree of *affine approximability* of an image is essentially determined by its degree of *flatness*. Once again, a benchmark example is the constant image $u = C$.

In [2], we examined the Δ_{ij} distributions for a number of test images. Some results are presented in Table 1: The entries have been arranged in a kind of "decreasing self-similarity" based upon increasing mean and, to some extent, increasing width. Estimates of the (natural logarithm) entropies of these distributions have also been presented in Table 1 (third



(a) *Lena* (b) *Mandrill*

Figure 4. *Distributions of Case* 3*, same-scale,* α *greyscale coefficients for normalized* Lena *and* Mandrill *images.*

Table 1

Columns 1–3: *Means, standard deviations, and entropies of collage error distributions for some standard test images. Columns* 4 *and* 5: *Means and standard deviations of* σ*-distributions of these images, to show their agreement with columns* 1 *and* 2*, respectively. Reprinted from* [*S. K. Alexander, E. R. Vrscay, and S. Tsurumi,* A simple, general model for the affine self-similarity of images*, in International Conference on Image Analysis and Recognition (ICIAR* 2008*), Lecture Notes in Comput. Sci.* 5112*, Springer-Verlag, Berlin, Heidelberg,* 2008*, pp.* 192–203 *(Table* 1*, p.* 202*)*]. *Copyright* © 2008 *Springer-Verlag. Reprinted with kind permission of Springer Science+Business Media. All rights reserved.*

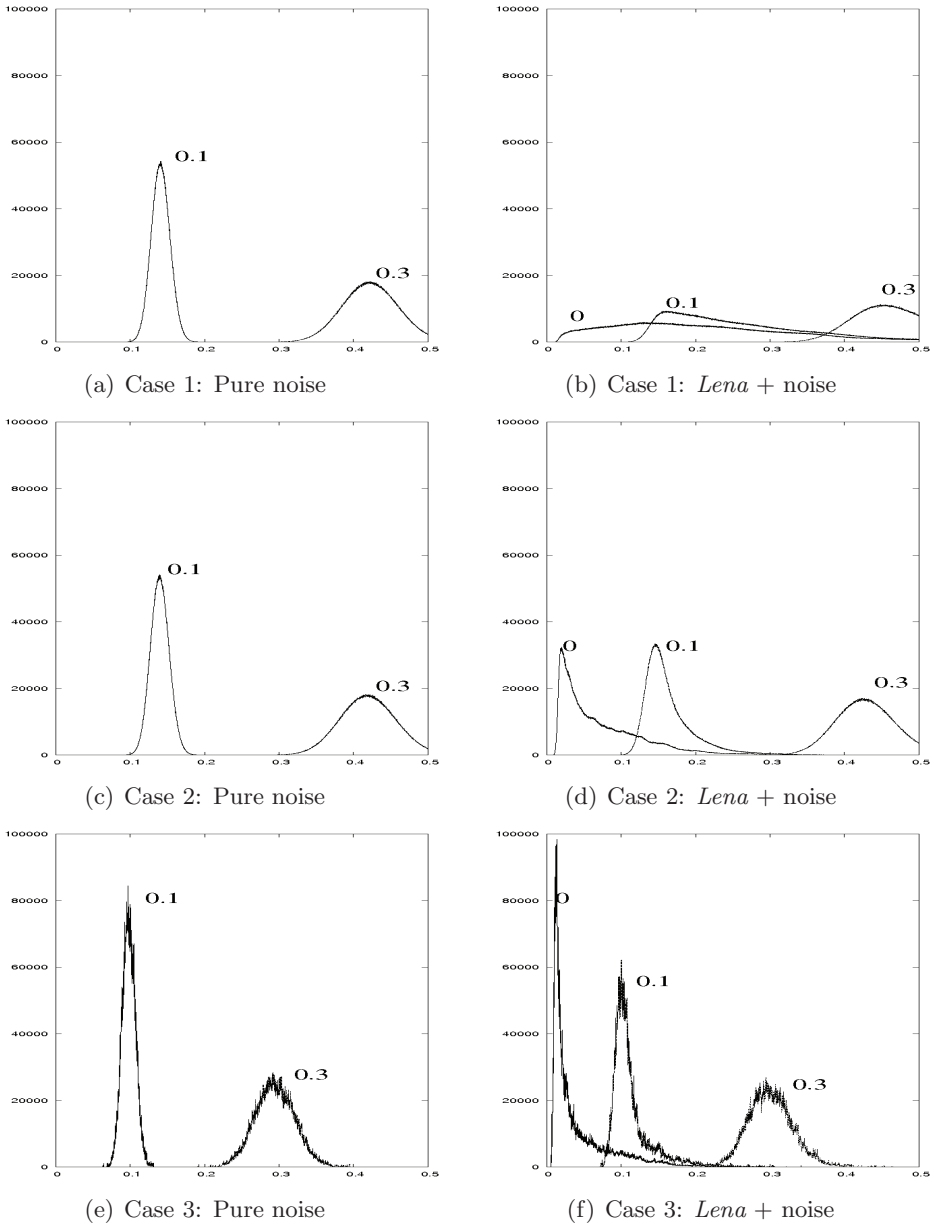
Image	Collage errors			Range block stddevs	
	mean	stddev	entropy	mean	stddev
Lena	0.043	0.044	2.26	0.046	0.046
San Francisco	0.046	0.057	2.01	0.048	0.059
Peppers	0.047	0.050	2.32	0.049	0.052
Goldhill	0.049	0.034	2.46	0.052	0.036
Boat	0.052	0.052	2.58	0.055	0.055
Barbara	0.060	0.049	2.69	0.064	0.051
Mandrill	0.089	0.048	2.85	0.089	0.048
Zelda	0.126	0.055	3.09	0.141	0.054

column)—note that with the exception of *San Francisco*, they increase as we proceed down the table. The *Lena* and *Mandrill* images can be viewed as lying roughly on opposite sides of a spectrum of distributions that vary in their mean values and variances.

In the final two columns of this table we present the estimates of the means and standard deviations of the distributions of standard deviations for these images (cf. Figure 3) to show their excellent agreement with those of the collage error distributions.

3.2. The effects of noise on Δ-error distributions. The presence of noise in an image will generally decrease the ability of its subblocks to be approximated by other subblocks. As such, we expect that the Δ-error distributions will be shifted away from zero error and possibly broadened as well. This was observed in [2] and will be discussed only briefly here.

In Figure 5 are shown the Δ-error distributions for the normalized *Lena* image for added Gaussian zero-mean noise $\mathcal{N}(0, \sigma^2)$ with several variances (right column). In the left column



(a) Case 1: Pure noise (b) Case 1: *Lena* + noise

(c) Case 2: Pure noise (d) Case 2: *Lena* + noise

(e) Case 3: Pure noise (f) Case 3: *Lena* + noise

Figure 5. *Δ-error distributions for constant image $u = 0.5$ (left) and normalized* Lena *image (right) plus independent Gaussian noise $\mathcal{N}(0, \sigma^2)$, Cases* 1–3. *For $\sigma = 0$, the Δ-distributions of the image $u = 0.5$ consist of a single peak at $\Delta = 0$.*

are shown the Δ-error distributions for the "pure noise" images $u = 0.5 + \mathcal{N}(0, \sigma^2)$ for comparison. (In these cases, the Δ-distributions of the noiseless $u = 0.5$ images consist of a single peak at $\Delta = 0$.) In all cases, the peaks of the distributions move outward with increasing σ. Moreover, the peaks for the noisy *Lena* roughly coincide with the peaks of their pure noise counterparts. For Case 3, it can be shown [2] that the Δ-distributions for pure noise images

peak at the standard deviation σ.

Finally we mention that for sufficiently low σ, i.e., $\sigma = 0.1$, the Δ-distributions of the pure noise images are sharper (and therefore higher) than those of the noisy *Lena* images.

3.3. Same-scale self-similarity and measure-valued image functions. The Δ-error distributions associated with an image function u give some idea of the overall self-similarity of an image. However, it may be useful to characterize the degree of self-similarity of the image at particular points or regions. Here we show how the space of measure-valued image functions (Y, d_Y) introduced in section 2 can be useful for this purpose.

In what follows, we again assume that the image function u is an $n_1 \times n_2$-pixel array with greyscale range $\mathbb{R}_g = [0, 1]$. We also assume that the range and domain blocks come from a common pool $\mathcal{R}$. We now consider all possible range/domain block pairs (R_i, R_j), along with the *isometric* affine transformations $w_{ij}^{(k)}$, $1 \le k \le 8$, that map R_j to R_i. For simplicity of notation, we omit the k index. (In most of the examples presented in this section, we employ only translations, i.e., $k = 1$.)

For a given same-scale scheme, i.e., Case q, where $q \in \{1, 2, 3\}$, we first compute all possible appropriate approximation errors $\Delta_{ij}^{(Case\ q)}$ (cf. (13)) and let $\phi_{ij}^{(Case\ q)}(t)$ denote the greyscale maps associated with these errors. (Recall that the greyscale maps assume different forms in the three cases.) For each range/domain pairing (R_i, R_j), we then assign a weighting function p_{ij} which is normalized as follows:

$$\sum_{j=1}^{N_R} p_{ij} = 1, \quad 1 \le i \le N_R. \tag{20}$$

An obvious question is the choice of the weighting parameters p_{ij}. It would seem natural to employ higher weights for those domain blocks D_j that yield lower approximation errors Δ_{ij}. Here we consider a weighting scheme that is similar in form to the one used in [7] for NL-means denoising:

$$p_{ij} = \frac{1}{Z_i} \exp\left(-\frac{\Delta_{ij}^P}{h^P} \right), \tag{21}$$

where $P > 0$, $h > 0$, and $Z_i = \sum_j \exp(-\Delta_{ij}^P / h^P)$ is the normalization factor. In practice, P is either 1 or 2. As for the adjustable parameter h, note the following:

1. In the limit $h \to 0$, the p_{ij}(s) with the smallest error Δ_{ij} will be selected.
2. In the limit $h \to \infty$, all p_{ij} become equal—in other words, all parent/domain pairings are employed equally.

For a given Case q and a set of prescribed weights $\{p_{ij}\}$ we may define an operator $T^{(Case\ q)}$, the action of which on an image function $u(x)$ is given by

$$v(x) = (T^{(Case\ q)} u)(x) = \sum_{j=1}^{N_R} p_{ij} \phi_{ij}^{(Case\ q)}(u(w_{ij}^{-1}(x))), \qquad x \in R_i. \tag{22}$$

In other words, the value $v(x)$ at an $x \in R_i$ is replaced by a weighted sum of modified pixel values—we shall refer to these values as the *preimages* of the value $v(x)$—from all other blocks R_j.

We state the following trivial result without proof.

Lemma 3.1.

$$T^{(Case\ q)} : \mathcal{L}^2(X) \to \mathcal{L}^2(X), \qquad q \in \{1, 2, 3\}. \tag{23}$$

A number of other properties of T can be established (e.g., Lipschitz, hence continuous), but they are not important to the discussion at hand.

The following result establishes that the distance $\|T^{(Case\ q)}u - u\|$ reflects the total Case q self-similarity, or lack thereof, of the image.

Lemma 3.2.

$$\|T^{(Case\ q)}u - u\| \le \sum_{i,j}^{N_R} p_{ij}\Delta_{ij}^{(Case\ q)}. \tag{24}$$

Proof. For notational convenience, we omit the superscript "$(Case\ q)$." In what follows, we also let $\|u\|_{R_i}$ denote the quantity $\|u|_{R_i}\|$, i.e., the norm of the restriction of u on the set R_i. Because the blocks R_i are nonoverlapping, it follows that

$$\|Tu - u\| = \sum_i \|Tu - u\|_{R_i}. \tag{25}$$

Then

$$\begin{aligned}
\|Tu - u\|_{R_i} &= \left\| \sum_j p_{ij}\phi_{ij} \circ u \circ w_{ij}^{-1} - u \right\|_{R_i} \\
&= \left\| \sum_j p_{ij}(\phi_{ij} \circ u \circ w_{ij}^{-1} - u) \right\|_{R_i} \\
&\le \sum_j p_{ij} \|\phi_{ij} \circ u \circ w_{ij}^{-1} - u\|_{R_i} \\
&= \sum_j p_{ij}\Delta_{ij}. \qquad (26)
\end{aligned}$$

From (25), the desired result follows. ∎

For the special "benchmark" self-similar image, $u = C$, for which $\Delta_{ij} = 0$ for all i, j, we note that $\|Tu - u\| = 0$. In other words, u is a fixed point of the operator T for Cases 1–3. It would be interesting to investigate the distances $\|Tu - u\|$ for various images, various weighting schemes (i.e., $h = 0$, $h = \infty$), and also in the presence of noise, but such an investigation is beyond the scope of this paper.

The T operators introduced above essentially "collapse" all preimages of an image function value $u(x)$ onto a single value. It may be interesting to examine the *range of values* that are assumed by these preimages. It is here that the idea of a measure-valued image function will be useful.

For this purpose, we associate with the image function $u(x)$ a corresponding measure-valued image function $\mu(x) \in (Y, d_Y)$ (cf. section 2) as follows:

$$\mu(x) = \delta_{u(x)}, \qquad x \in X. \tag{27}$$

Here, δ_t denotes a unit point mass measure at $t \in \mathbb{R}_g$. We now define a measure-valued image $\nu = M\mu \in Y$ as follows: For any measurable set $S \subset \mathbb{R}_g = [0,1]$ and any $x \in R_i$, we define

$$\begin{aligned} \nu(x)^{(Case\ q)}(S) &= (M^{(Case\ q)}\mu)(x)(S) \\ &= \sum_{j=1} p_{ij}\mu(w_{ij}^{-1}(x))(\phi_{ij}^{-1(Case\ q)}(S)), \qquad q \in \{1,2,3\}. \end{aligned} \tag{28}$$

The measure $\nu^{(Case\ q)}$, $q \in \{1,2,3\}$, is designed to reflect the Case q self-similarity of the image function u: Given a range block $R_i \subset X$, then at each point $x \in R_i$, we keep track of all greyscale values of the image function that are mapped to x by a domain/range mapping w_{ij} and modified by the corresponding Case q greyscale map $\phi_{ij}(t)$. These values are then weighted and "assembled" to define the probability measure ν at x, i.e., $\nu(x)$.

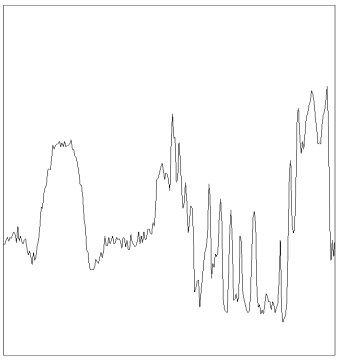
In Figure 6, we show all *contributions* to (essentially support of) the measures $\nu(x)^{(Case\ q)}$, $q = 1,2,3$, for the row of pixels $u(256,j)$, $j = 1,\ldots,256$, in the *Lena* image. These are the pixels that join the midpoint of the left edge of the image to the center of the image; cf. Figure 12. Their values are plotted in Figure 4(a). The range blocks were nonoverlapping 8×8-pixel blocks. Only range blocks from the top left quarter of the image were used in the fitting. Note that no weighting has been employed here.

In Case 1, for which the greyscale map $\phi(t)$ is the identity mapping, the corresponding pixel values from all blocks contribute to each pixel value on the line—no actual fitting of domain block pixels to corresponding range block pixels is accomplished by the greyscale maps. The periodicity of the pattern arises from the 8-pixel size of the range blocks.

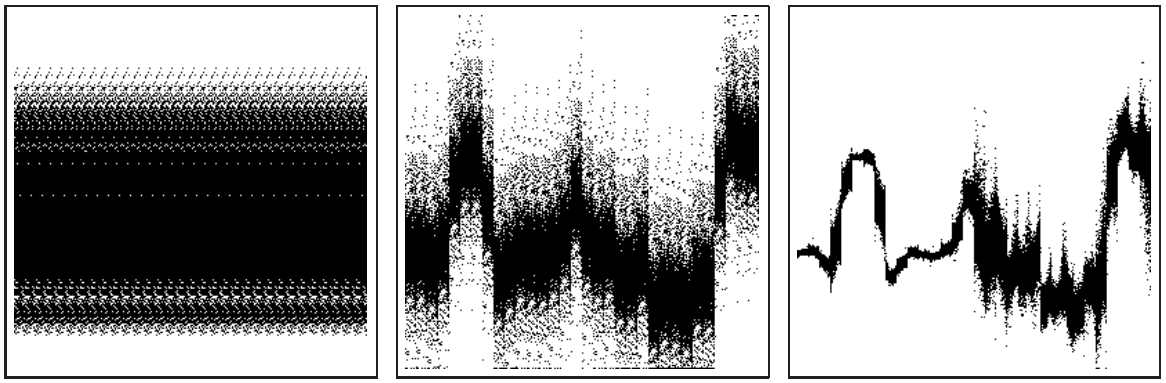
In Case 2, some fitting of domain block to range block pixels is accomplished by the greyscale shift map. As a result, the domain greyscale values have been moved somewhat toward the range greyscale values $u(256,j)$.

In Case 3, a much greater degree of fitting of domain block to range block pixels is accomplished by the affine greyscale map. As a result, even further shrinkage of domain greyscale values toward the range greyscale values $u(256,j)$ is accomplished.

In Figure 7 are shown pictorial representations of the measures $\nu(x)^{(Case\ q)}$, $q = 1,2,3$. For each Case q, we have used three values of the weight parameter h in (21), along with $P = 2$. In these figures, darker regions have higher associated measures. As expected from the previous figure, the measures become more concentrated about the actual greyscale values $u(256,j)$, $1 \le j \le 256$, of the *Lena* image as we move from Case 1 to Case 3. In all three cases, the weight parameter value $h = 0.01$ effectively concentrates the measures close to the *Lena* image values, even for the Case 1 measure. However, as h is increased to 0.1, the Case 1 measure becomes quite diffuse and, at $h = 1.0$, quite unrelated to the *Lena* image values. For Cases 2 and 3, there is virtually no change between $h = 0.1$ and $h = 1.0$ and even $h = \infty$ (equal probabilities, not shown in the figure).



(a) *Lena* (normalized) greyscale values $u(256, j)$, $1 \le j \le 256$



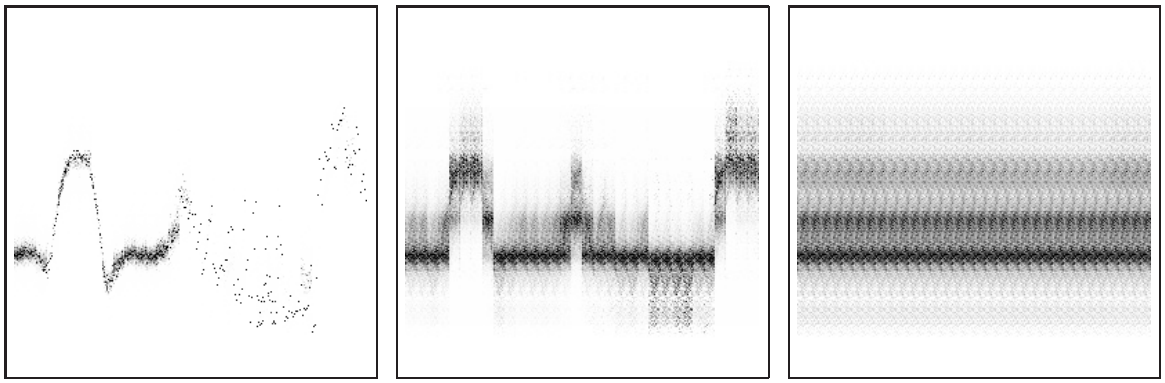
(b) Case 1 (c) Case 2 (d) Case 3

Figure 6. (a) *Greyscale values of the normalized* Lena *image for the row of pixels* $u(256, j)$, $j = 1, \ldots, 256$. (b)–(d) *Pictorial representations of the contributions to the measures* $\nu(x)^{(Case\ q)}$, $q = 1, 2, 3$, *respectively, for this row of pixels. The measures are arranged in vertical columns and occupy* 256 *bins that range in value from* 0 *(bottom) to* 1 *(top). In each case, the leftmost measure corresponds to* $j = 1$, *and the rightmost to* $j = 256$.

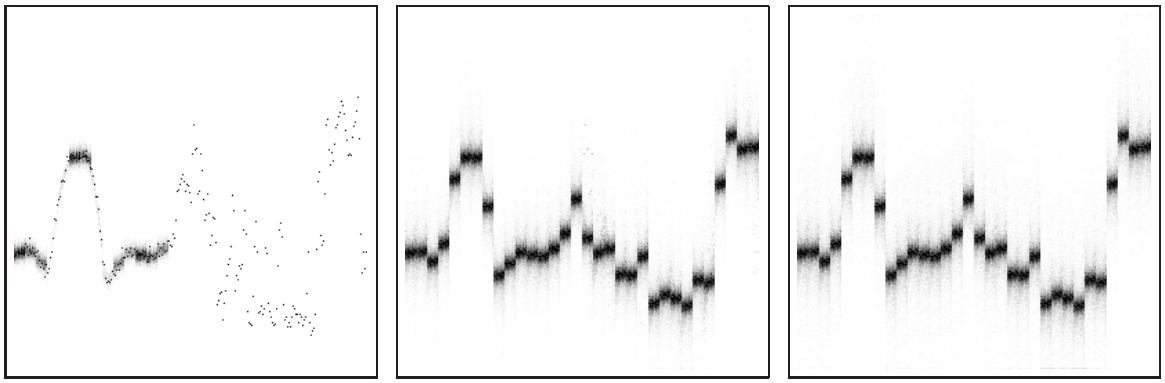
3.4. Effects of noise on measures and a simple denoising method. In the first column of Figure 8 are shown pictorial representations of the measures $\nu^{(Case\ q)}$ associated with the pixels $u(256, j)$, $1 \le j \le 256$, of the normalized *Lena* image with added noise $\mathcal{N}(0, \sigma^2)$, where $\sigma = 0.1$. These pictures should be compared to their noiseless counterparts in the middle column of Figure 7. As expected, the measures for the noisy case are more diffuse, although the Case 3 measure appears to be as concentrated for the noisy case as for the noiseless one.

A natural question is whether the means $\bar{x}$ of these measures estimate the greyscale values of the noiseless *Lena* image. In the middle column of Figure 8 are plotted the mean values of these measures for the three cases $q = 1, 2, 3$. The RMSE with respect to the noiseless (normalized) image values $u(256, j)$, $1 \le j \le 256$, is given for each case. As expected, the approximations are poorest in regions of high variance/oscillation, for example, $130 \le x \le 200$.

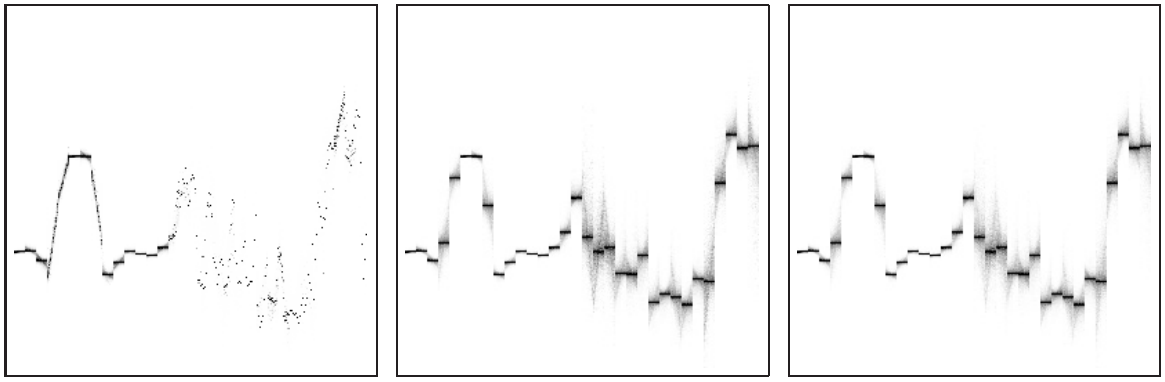
In the right column of Figure 8 are plotted the variances $\sigma(x)$ of the corresponding $\nu(x)^{(Case\ q)}$ measures. In accordance with the plots of the first column, there is a dramatic reduction in variance as we move from Case 2 to Case 3.



(a) Case 1, $h = 0.01$ (b) Case 1, $h = 0.1$ (c) Case 1, $h = 1.0$



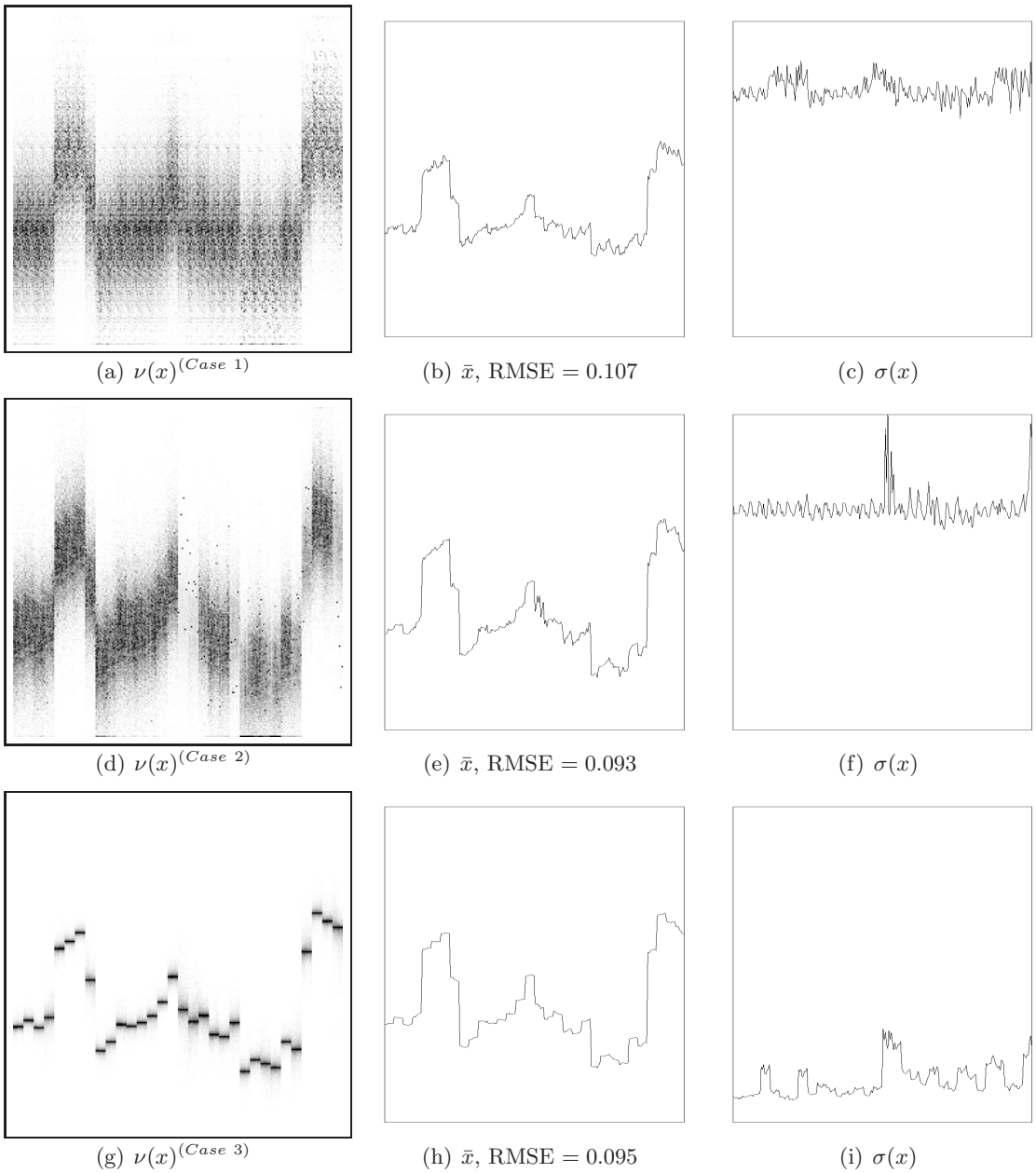
(d) Case 2, $h = 0.01$ (e) Case 2, $h = 0.1$ (f) Case 2, $h = 1.0$



(g) Case 3, $h = 0.01$ (h) Case 3, $h = 0.1$ (i) Case 3, $h = 1.0$

Figure 7. *Pictorial representations of measures* $\nu(x)^{(Case\ q)}$, $q = 1, 2, 3$, *for the* $u(256, j)$, $1 \le j \le 256$, *row of pixels for the* Lena *image. Three values of the weighting parameter* h *in* (21) *have been used, with* $P = 2$.

Note that, regardless of the comparatively high diffusivity of the Case 1 and Case 2 measures, the accuracy of their means—in terms of RMSE—is virtually identical to that of Case 3. In essence, all three same-scale similarity methods—Cases 1, 2, and 3—appear to perform denoising almost equally well in this example. This is generally true for the case of coarser approximations. The three methods perform differently when both the value of h and



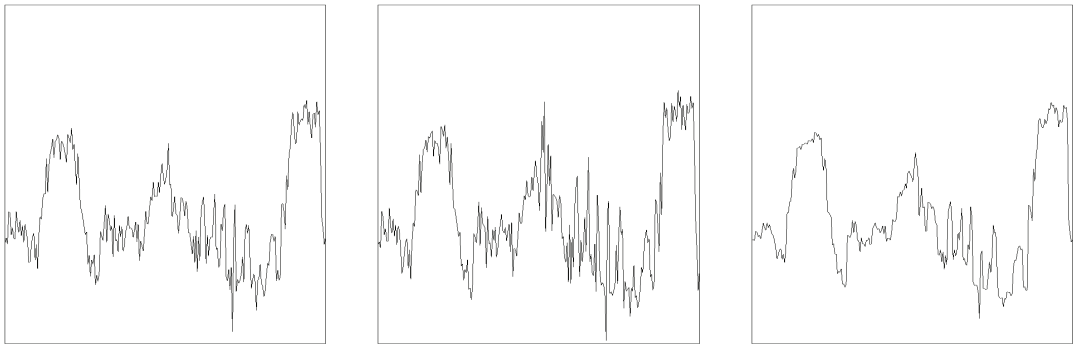
(a) $\nu(x)^{(Case\ 1)}$ (b) $\bar{x}$, RMSE $= 0.107$ (c) $\sigma(x)$

(d) $\nu(x)^{(Case\ 2)}$ (e) $\bar{x}$, RMSE $= 0.093$ (f) $\sigma(x)$

(g) $\nu(x)^{(Case\ 3)}$ (h) $\bar{x}$, RMSE $= 0.095$ (i) $\sigma(x)$

Figure 8. *Left column: Measures* $\nu(x)^{(Case\ q)}$ *for the normalized* Lena *image with added Gaussian noise with* $\sigma = 0.1$. *In all cases,* $h = 0.1$. *Vertical axis is* $[0, 1]$. *Middle column: Mean values* $\bar{x}$ *of the* $\nu(x)^{(Case\ q)}$ *measures. Vertical axis* $[0, 1]$. *These results represent denoised values of the noisy* Lena *image and should be compared to the noiseless image values in Figure* 6(a). *Right column: Variances* $\sigma(x)$ *of the corresponding* $\nu(x)^{(Case\ q)}$ *measures. Vertical axis* $[0, 0.15]$.

the range block size are decreased to produce finer approximations. For example, Figure 9 shows the mean value plots that result from $h = 0.05$ and 4×4-pixel range blocks. In this case, there is an improvement in accuracy as we move from Case 1 to Case 3.



(a) Case 1, RMSE = 0.077 (b) Case 2, RMSE 0.067 (c) Case 3, RMSE 0.062

Figure 9. *Mean values $\bar{x}$ of the $\nu(x)^{(Case\ q)}$ measures obtained from the noisy* Lena *image, with parameter* $h = 0.05$.

The above analysis, however, involves only a single (half) row of pixels in the *Lena* image. When we examine the results of Case 1–3 denoising on a larger portion of the image, as shown in Figure 10 (upper left quarter of the *Lena* image), the differences between the methods, at least in terms of RMSE values, are more pronounced.

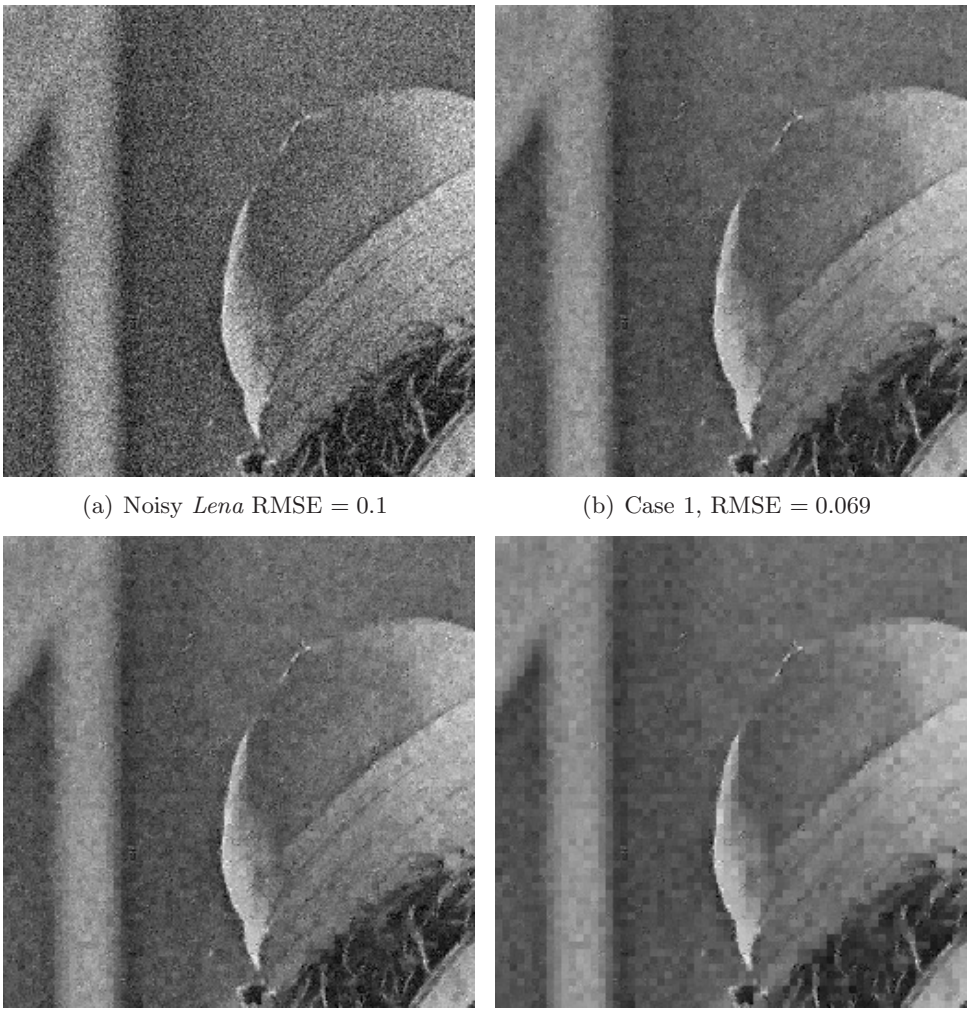
That being said, these few results are not meant to be interpreted as a complete analysis of this approximation problem. A more detailed investigation, which is beyond the scope of this paper, is clearly needed.

In the next section, we explain the effectiveness of this very simple method of denoising and relate it to the recently developed "nonlocal means" denoising method [7].

3.5. "Nonlocal-means denoising" as a translational self-similarity method. A standard technique for the reduction of additive white noise is to average over multiple samples. This is the basis of the very effective "nonlocal-means denoising algorithm" [7], where *the multiple samples are provided by the image itself.* Very briefly, each pixel $u(i)$ of a noisy image is replaced by a convex combination of other pixel values $u(j)$ from the image. The weights λ_{ij} of this averaging procedure depend upon the similarity between neighborhoods N_i and N_j centered about pixels i and j, respectively. Neighborhoods N_k that do not approximate N_i very well, i.e., with high $\mathcal{L}^2$ error $\|N_i - N_k\|$, are assigned low weights. Moreover, the weights λ_{ij} have roughly the same form as those used in (21), with $P = 2$.

The NL-means method may be viewed as a variation of the same-scale self-similarity operator $T^{(Case\ 1)}$ of (22) and therefore the block-based measure-based denoising method described in the previous section. In either procedure, an appropriately defined measure $\nu(x)$ acts as an intermediate step. Because it operates on individual pixels (based on Case 1 similarity between neighborhoods), the NL-means method will yield better results, but of course at greater computational expense.

It is remarkable that the NL-means method, which relies on the translational (Case 1) self-similarity of an image, works so well. Because of the translational symmetry requirement, only a few blocks will generally contribute significantly to the denoising of a given pixel. Nevertheless, as we have shown in the previous section, some improvement is possible with the simple inclusion of greyscale shifts (Case 2), which require minimal effort to compute.



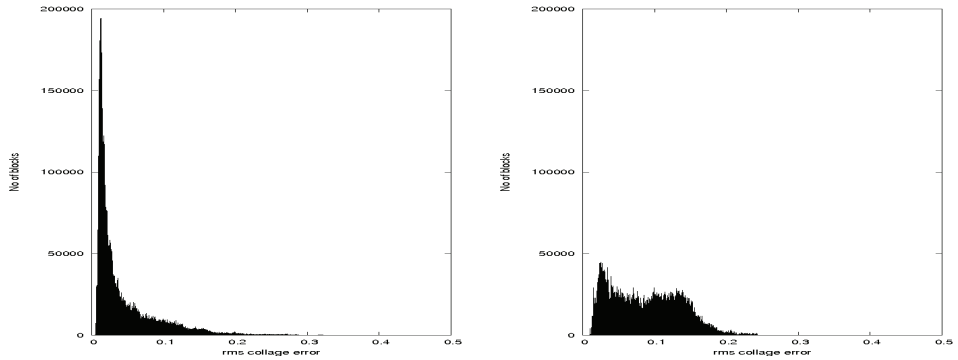
(a) Noisy *Lena* RMSE = 0.1 (b) Case 1, RMSE = 0.069

(c) Case 2, RMSE = 0.054 (d) Case 3, RMSE = 0.040

Figure 10. *Results of Case* 1–3 *denoising of upper quarter of* Lena *image.*

3.6. Case 4: Two-scale self-similarity and fractal image coding. The Δ-error distributions for Case 4, cross-scale self-similarity matching are generally quite similar to their Case 3, same-scale counterparts. In Figure 11(a) are presented histogram plots of the approximation errors Δ_{ij} for the *Lena* and *Mandrill* test images, once again for range block partitions $\mathcal{R}$ formed by the set of all 8×8 nonoverlapping pixel blocks of the images ($64^2 = 4096$ in total). For each image, the domain pool $\mathcal{D}$ was formed from the set of $32^2 = 1024$ 16×16 nonoverlapping pixel blocks. In addition, for each range/domain block pairing R_j/D_i, we considered all eight square-to-square contractions, for a total of 33,554,432 collage errors. The histogram distributions are very similar to the Case 3 same-scale Δ-error distributions presented in Figure 2.

The highly peaked nature of the Case 4 Δ-error distributions for many images indicates that range blocks R_i of an image are generally well approximated by a significant number of



(a) Case 4: *Lena* (b) Case 4: *Mandrill*

Figure 11. *Histogram distributions of two-scale collage errors* Δ_{ijk} *for all* 16×16 *to* 8×8 *pixel domain/range block pairings (and* 8 *square-to-square contractions) for* Lena *and* Mandrill *images.*

larger domain blocks D_j from the image. It is often the case that there are several "near-optimal" domain blocks, i.e., blocks for which the collage error is very close to the minimum value achieved by the optimal domain block. We shall discuss the exploitation of this feature in the next section.

The effect of additive, zero-mean Gaussian noise $\mathcal{N}(0,\sigma^2)$ on Case 4, cross-scale Δ-error distributions is the same as for the Case 3, same-scale case: As the variance σ^2 is increased, the Δ-error distributions are pushed away from the zero-error axis. Moreover, these distributions will peak at σ. Because of the marked similarity between these two cases, we omit the presentation of graphical evidence—it is sufficient to consult the bottom row of Figure 5.

Case 4, cross-scale self-similarity forms the basis of fractal image coding [6, 12, 24]. Given a "target" image u, each image subblock $u(R_i)$ is approximated by a geometrically contracted and affine greyscale-modified copy of a larger subblock $u(D_j)$. In traditional fractal coding, the range/domain assignments $(i, j(i))$ and associated greyscale parameters (α_i, β_i) define a *fractal transform* operator T. We may write (11) as follows:

$$u(x) \approx (Tu)(x) = \alpha_i u(w_{i,j(i)}^{-1}(x)) + \beta_i, \quad x \in R_i, \quad 1 \le i \le N_R. \tag{29}$$

Under appropriate conditions [14] involving the α_i and the contraction factors of the spatial maps $w_{i,j(i)}$, T is contractive in $\mathcal{L}^2(X)$. From Banach's fixed point theorem, this implies the existence of a unique fixed point function $\bar{u} = T\bar{u}$. Furthermore, $\bar{u}$ may be generated by iteration: Starting with any seed image u_0, let $u_{n+1} = Tu_n$. Then $u_n \to \bar{u}$ as $n \to \infty$. (In the discrete case, convergence is achieved after a finite number of iterations.) Unless all greyscale parameters β_i are zero, which is never the case in practice, the fixed point $\bar{u}$ is nonzero. And from (29), $\bar{u}$ is an approximation to the target image u, to be discussed in more detail below.

The fractal transform operator T in (29) is a *nonlocal* operator since blocks of an image function are replaced by modified copies of blocks from elsewhere in the image. The fixed point $\bar{u}$ of T is affinely self-similar since the approximation in (29) becomes an equality:

$$\bar{u}(x) = (T\bar{u})(x) = \alpha_i \bar{u}(w_{i,j(i)}^{-1}(x)) + \beta_i, \quad x \in R_i, \quad 1 \le i \le N_R. \tag{30}$$

The mathematical basis for fractal-based approximation is provided by the so-called *collage theorem* [5], which is a simple consequence of Banach's theorem:

$$\|u - \bar{u}\| \;\le\; \frac{1}{1 - c_T}\|Tu - u\|, \tag{31}$$

where c_T is the contraction factor of T. Given a set of range blocks $\mathcal{R}$ and a *domain pool* $\mathcal{D}$, one tries to to make the approximation error $\|u - \bar{u}\|$ as small as possible by minimizing the *collage error* $\|u - Tu\|$. From (29) this is generally done as follows: For each range block R_i, we search the domain pool $\mathcal{D}$ for the block $D_{j(i)}$ that yields the lowest approximation error Δ_{ij} in (13). (In most applications, the range of the α_{ij} greyscale parameters is restricted to $[-\alpha_{\max}, \alpha_{\max}]$, where $\alpha_{\max}$ is typically 1 in value, in order to guarantee contractivity of T. Such a "clamping" procedure has been employed in the calculations reported in this paper.)

Of course, if the size of the domain pool $\mathcal{D}$ is increased, it may be possible to decrease the collage error. However, an increase in the size of $\mathcal{D}$ implies a greater computational cost because of the searching involved. This is a well-known problem in fractal coding [12, 6, 24]. The procedure outlined above, referred to as *collage coding* in the fractal coding literature, was first proposed by Jacquin [20] and forms the basis of most, if not all, *block-based fractal coding* procedures.

In Figure 12 (lower left) is shown the fixed point approximation $\bar{u}$ to the standard 512×512-pixel *Lena* image (8 bits/pixel) obtained from collage coding, once again using a partition $\mathcal{R}$ of 8×8 nonoverlapping pixel blocks and a domain pool $\mathcal{D}$ of $32^2 = 1024$ 16×16 nonoverlapping pixel blocks. (This choice of domain pool is clearly not optimal.) This image was obtained by starting with the seed image $u_0(x) = 255$ (plain white image) and iterating $u_{n+1} = Tu_n$ to $n = 15$. Iterates u_1, u_2, and u_3 are also shown in this figure.

3.7. Multiparent block fractal coding. Historically, most fractal image coding research focused on its compression capabilities—obtaining acceptable accuracy with the smallest possible domain pool in order to minimize (i) search times and (ii) storage of the fractal code. The fact that range blocks R_i of an image are, in general, well approximated by a good number of domain blocks D_j does not seem to have been emphasized or exploited. Consequently, investigations generally focused on the results yielded by optimal domain blocks of the pool and not on the possible use of suboptimal ones.

More recently, however, the redundancy of good domain/range pairings has been exploited [1] in order to perform image denoising. Such *multiparent* block fractal coding schemes can be viewed as cross-scale analogues of the same-scale NL-means denoising scheme: The use of several domain blocks for each range block, followed by downsampling, corresponds to averaging over multiple samples, resulting in a reduction of noise variance.

In this section, we describe a simple multiparent block fractal coding scheme that results from a modification of the block-based coding method outlined in the previous section. As in the same-scale case, this multiparent scheme lends itself nicely to a measure-based formalism.

We consider the same (square) range and domain block pools, $\mathcal{R}$ and $\mathcal{D}$, respectively, used in the previous section. For each range block R_i, we compute the Δ_{ij} approximation errors associated with *all* domain blocks D_j; cf. (13). (Recall that for each range/domain pairing (R_i, D_j) there are eight spatial contraction/decimation maps $w_{ij}^k : D_j \to R_i$, $1 \le k \le 8$.

Figure 12. *Starting at upper left and moving clockwise: The iterates u_1, u_2, and u_3 along with the fixed point $\bar{u}$ of the fractal transform operator T designed to approximate the standard* 512×512 *(8 bits/pixel)* Lena *image. The "seed" image was $u_0(x) = 255$ (plain white). The fractal transform T was obtained by "collage coding" using* 4096 8×8 *nonoverlapping pixel range blocks. The domain pool consisted of the set of* 1024 *nonoverlapping* 16×16 *pixel blocks.*

Once again, for simplicity of notation, we shall omit the k index.) The optimal greyscale map minimizing the error Δ_{ij} will be denoted as

$$\phi_{ij}(t) = \alpha_{ij} t + \beta_{ij}. \tag{32}$$

For this pairing we also assign a weight p_{ij}, normalized so that

$$\sum_{j}^{N_R} p_{ij} = 1, \qquad 1 \le i \le N_R. \tag{33}$$

The resulting multiparent block transform operator T is then defined as

$$(Tu)(x) = \sum_j p_{ij}\alpha_{ij}u(w_{ij}^{-1}(x)) + \beta_{ij}, \quad x \in R_i. \tag{34}$$

This definition represents a generalization of the fractal transform operator of (29) in the previous section—not only one but several, perhaps all, domain blocks $D_j \in \mathcal{D}$ can contribute to the modification of $u(x)$ for $x \in R_i$. It may also be viewed as the Case 4, cross-scale analogue of the operators $T^{(Case\ q)}$ defined in (22).

Once again, an obvious question is the choice of the weighting parameters p_{ij}. It would once again seem natural to employ higher weights for those domain blocks D_j that yield lower approximation errors Δ_{ij}. Several weighting schemes were examined in [1] for the purpose of denoising. Here, for consistency, we shall employ the same weighting scheme as for multiparent schemes in Cases 1–3; cf. (21).

In principle, we could now proceed as in section 3.3 and define an associated operator $M^{(Case\ 4)}$ on the space Y of measure-valued image functions. And from this result we could construct Case 4 analogues of the results in Figure 7. However, the operator $M^{(Case\ 4)}$ will be shown to possess an additional property, namely, contractivity in (Y, d_Y). Unlike Cases 1–3, this will also allow us to iterate M to perform denoising.

In the next section, we set up the mathematical basis for the existence and contractivity of this *fractal transform operator* on (Y, d_Y).

4. A fractal transform operator M on (Y, d_Y). In this section, we construct and analyze a fractal transform operator M on the space (Y, d_Y) of measure-valued functions. For the benefit of the inexperienced reader, the appendix contains an overview of some main results from the theory of "iterated function systems" [4, 3], which provides the basis for our treatment.

We now list the ingredients for a fractal transform operator in the space Y. The reader will note that they form a kind of blending of IFS-based methods on measures (IFSP; see appendix) and functions (IFSM; see appendix). For simplicity, we assume that $X = [0, 1]$. The extension to $[0, 1]^n$ is straightforward. We have the following:

1. a set of N one-to-one, affine contraction maps $w_i : X \to X$, $w_i(x) = s_i x + a_i$, with the condition that $\cup_{i=1}^N w_i(X) = X$;
2. a set of N greyscale maps $\phi_i : \mathbb{R}_g \to \mathbb{R}_g$, assumed to be Lipschitz, i.e., for each i, there exists an $\alpha_i \geq 0$ such that

$$|\phi_i(t_1) - \phi_i(t_2)| \leq \alpha_i |t_1 - t_2| \quad \forall t_1, t_2 \in \mathbb{R}_g; \tag{35}$$

3. for each $x \in X$, a set of probabilities $p_i(x)$, $i = 1, \ldots, N$, with the following properties:
   (a) $p_i(x)$ are measurable,
   (b) $p_i(x) = 0$ if $x \notin w_i(X)$, and
   (c) $\sum_i^N p_i(x) = 1$ for all $x \in X$.

The action of the fractal transform operator $M : Y \to Y$ defined by the above is as follows: For a $\mu \in Y$ and any subset $S \subset [0, 1]$,

$$\nu(x)(S) = (M\mu(x))(S) = \sum_{i=1}^N p_i(x)\mu(w_i^{-1}(x))(\phi_i^{-1}(S)). \tag{36}$$

Theorem 4.1. *Let $p_i = \sup_{x\in X} p_i(x)$. Then for $\mu_1, \mu_2 \in Y$,*

$$d_Y(M\mu_1, M\mu_2) \le \left(\sum_{i=1}^{n} |s_i|\alpha_i p_i\right) d_Y(\mu_1, \mu_2). \tag{37}$$

Proof. Computing, we have

$$\begin{aligned}
d_Y(M\mu_1, M\mu_2) &= \int_X d_H(M\mu_1(x), M\mu_2(x))d\mu_{\mathcal{L}} \\
&= \int_X d_H\left(\sum_{i=1}^{N} p_i(x)\mu_1(w_i^{-1}(x)) \circ \phi_i^{-1}, \sum_{i=1}^{N} p_i(x)\mu_2(w_i^{-1}(x)) \circ \phi_i^{-1}\right) d\mu_{\mathcal{L}} \\
&\le \int_X \sum_{i=1}^{n} p_i(x) d_H(\mu_1(w_i^{-1}(x)) \circ \phi_i^{-1}, \mu_2(w_i^{-1}(x)) \circ \phi_i^{-1}) d\mu_{\mathcal{L}} \\
&\le \int_X \sum_{i=1}^{n} \alpha_i p_i(x) d_H(\mu_1(w_i^{-1}(x)), \mu_2(w_i^{-1}(x))) d\mu_{\mathcal{L}} \\
&\le \int_X \left(\sum_{i=1}^{n} |s_i|\alpha_i p_i\right) d_H(\mu_1(x), \mu_2(x)) d\mu_{\mathcal{L}} \\
&= \left(\sum_{i=1}^{n} |s_i|\alpha_i p_i\right) d_Y(\mu_1, \mu_2). \qquad \blacksquare
\end{aligned}$$

Corollary 4.2. *Let $p_i = \sup_{x\in X} p_i(x)$. Then M is a contraction on (Y, d_Y) if*

$$\sum_{i=1}^{n} |s_i|\alpha_i p_i < 1. \tag{38}$$

Consequently there exists a measure-valued mapping $\bar{\mu} \in Y$ such that $\bar{\mu} = M\bar{\mu}$.

We have the following examples.

Example 1. The fractal transform M is defined by the following two-IFS-map system on $X = [0, 1]$:

$$\begin{aligned}
w_1(x) &= \frac{1}{2}x, \quad \phi_1(t) = \frac{1}{2}t, \\
w_2(x) &= \frac{1}{2}x + \frac{1}{2}, \quad \phi_2(t) = \frac{1}{2}t + \frac{1}{2}.
\end{aligned}$$

The sets $w_1(X)$ and $w_2(X)$ overlap at the single point $x = \frac{1}{2}$, so we let

$$\begin{aligned}
p_1(x) = 1, \quad p_2(x) = 0, \quad x \in \left[0, \frac{1}{2}\right), \\
p_1(x) = 0, \quad p_2(x) = 1, \quad x \in \left(\frac{1}{2}, 1\right], \\
p_1\left(\frac{1}{2}\right) = p_2\left(\frac{1}{2}\right) = \frac{1}{2}.
\end{aligned}$$

It is easy to confirm that M is contractive. Its fixed point $\bar{\mu}$ is given by

$$\bar{\mu}(x) = \delta(t - x), \quad x \in [0, 1], \tag{39}$$

where $\delta(s)$ denotes the "Dirac delta function" at $s \in [0, 1]$.

Example 2. A "perturbation" of the fractal transform M in Example 1 is produced by adding the following IFS and associated greyscale maps:

$$w_3(x) = \frac{1}{2}x, \quad \phi_3(t) = \frac{1}{2}t + 0.1.$$

The sets $w_1(X)$ and $w_3(X)$ overlap over the entire subinterval $[0, \frac{1}{2}]$, so we let

$$p_1(x) = p_3(x) = \frac{1}{2}, \quad p_2(x) = 0, \quad x \in \left[0, \frac{1}{2}\right),$$
$$p_1(x) = p_3(x) = 0, \quad p_2(x) = 1, \quad x \in \left(\frac{1}{2}, 1\right],$$
$$p_1\left(\frac{1}{2}\right) = p_2\left(\frac{1}{2}\right) = p_3\left(\frac{1}{2}\right) = \frac{1}{3}.$$

Once again, it is easy to confirm that M is contractive. Its fixed point $\bar{\mu}(x)$ is sketched in Figure 13. The darkness of a point is proportional to the measure $\bar{\mu}(S_1, S_2)$ of the region in $[0, 1]^2$ represented by the point.

Note that the overlapping of the w_1 and w_3 maps over $[0, \frac{1}{2}]$ is responsible for the self-similar "splitting" of the measures $\bar{\mu}(x)$ over this interval, since ϕ_3 produces an upward shift in the greyscale direction. Since $w_2(x)$ maps the support $[0, 1]$ of the entire measure-valued function onto $[\frac{1}{2}, 1]$, the self-similarity of the measure over $[0, \frac{1}{2}]$ is carried over to $[\frac{1}{2}, 1]$.

4.1. Continuity of fixed point measures with respect to probabilities. In this section we establish the stability of fixed point measures of the fractal transforms introduced above with respect to changes in the probabilities. Consider the fixed points μ and μ^* of

$$(M\mu)_x(S) = \sum_{i=1}^{N} p_i(x)\mu_{w_i^{-1}(x)}(\phi_i^{-1}(S)), \tag{40}$$

$$(M\mu)_x(S) = \sum_{i=1}^{N} p_i^*(x)\mu_{w_i^{-1}(x)}(\phi_i^{-1}(S)). \tag{41}$$

Theorem 4.3. *Let $\tilde{p}_i = \sup_{x \in X} p_i(x)$ and suppose that M is a contraction. Then there exists a constant $C > 0$ such that*

$$d_Y(\mu, \mu^*) \leq \frac{C}{1 - \sum_{i=1}^{N} \tilde{p}_i \alpha_i |s_i|} \max_{1 \leq i \leq N} d_1(p_i, p_i^*). \tag{42}$$

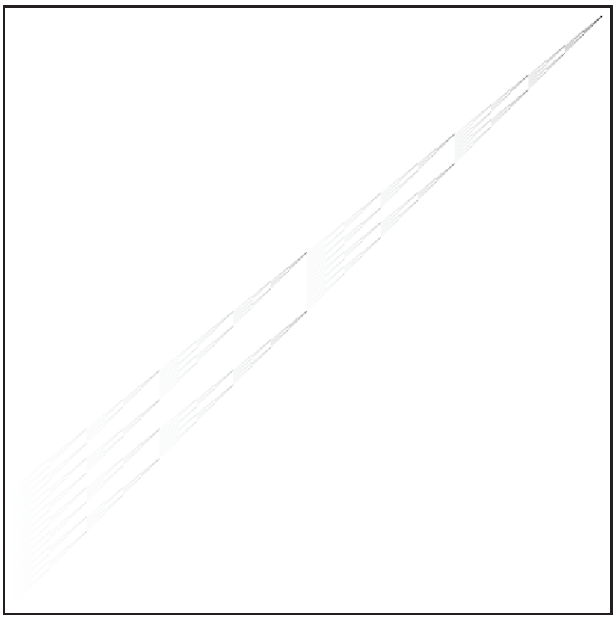


Figure 13. *A sketch of the invariant measure $\bar{\mu}(x)$ for the three-IFS-map fractal transform in Example 2, $x \in X = [0,1]$, $y \in \mathbb{R}_g = [0,1]$.*

Proof. Computing, we have

$$\begin{aligned} d_H(\mu_x, \mu_x^*) &= d_H\left(\sum_{i=1}^N p_i(x)\mu_{w_i^{-1}(x)} \circ \phi_i^{-1}, \sum_{i=1}^N p_i^*(x)\mu^*_{w_i^{-1}(x)} \circ \phi_i^{-1}\right) \\ &\le d_H\left(\sum_{i=1}^N p_i(x)\mu_{w_i^{-1}(x)} \circ \phi_i^{-1}, \sum_{i=1}^N p_i(x)\mu^*_{w_i^{-1}(x)} \circ \phi_i^{-1}\right) \\ &\quad + d_H\left(\sum_{i=1}^N p_i(x)\mu^*_{w_i^{-1}(x)} \circ \phi_i^{-1}, \sum_{i=1}^N p_i^*(x)\mu^*_{w_i^{-1}(x)} \circ \phi_i^{-1}\right) \\ &= \sum_{i=1}^N p_i(x)\alpha_i d_H(\mu_{w_i^{-1}(x)}, \mu^*_{w_i^{-1}(x)}) \\ &\quad + C|p_i(x) - p_i^*(x)|. \end{aligned}$$

This implies that

$$d_Y(\mu, \mu^*) \le \left(\sum_{i=1}^N \tilde{p}_i \alpha_i |s_i|\right) d_Y(\mu, \mu^*) + C \max_{1 \le i \le N} d_1(p_i, p_i^*), \tag{43}$$

which, upon rearrangement, yields the desired result. ∎

4.2. Moment relations induced by the fractal transform operator. In this section we show that the moments of measures in the space (Y, d_Y) also satisfy recursion relations when the greyscale maps ϕ_i are affine. We now consider the local or x-dependent moments of a

measure $\mu(x) \in Y$, defined as follows:

$$g_n(x) = \int_{\mathbb{R}_g} s^n d\mu_x(s), \quad n = 0, 1, 2, \ldots, \tag{44}$$

where we use the notation $\mu_x = \mu(x)$ in the Lebesgue integral for simplicity. By definition, $g_0(x) = 1$ for $x \in X$. Obviously the functions g_m are measurable on X (since $\mu(x)$ are measurable) and bounded so that $g_m \in \mathcal{L}^1(X, \mathcal{L})$. We now derive the relations between the moments of a measure $\mu \in Y$ and the moments of $\nu = M\mu$, where M is the fractal transform operator defined in (36).

Let h_n denote the moments of $\nu = M\mu$. Computing, we have

$$\begin{aligned} h_n(x) &= \int_{\mathbb{R}_g} s^n d(M\mu)_x(s) \\ &= \int_{\mathbb{R}_g} s^n d\left(\sum_{i=1}^N p_i(x) \mu_{w_i^{-1}(x)} \circ \phi_i^{-1} \right)(s) \\ &= \int_{\mathbb{R}_g} \sum_{i=1}^N p_i(x) s^n d(\mu_{w_i^{-1}(x)} \circ \phi_i^{-1})(s) \\ &= \int_{\mathbb{R}_g} \sum_{i=1}^N p_i(x) [\phi_i(s)]^n d(\mu_{w_i^{-1}(x)})(s). \end{aligned}$$

For affine greyscale maps of the form $\phi(s) = \alpha_i s + \beta_i$, we have

$$\begin{aligned} h_n(x) &= \int_{\mathbb{R}_g} \sum_{i=1}^N p_i(x) (\alpha_i + s\beta_i)^n d(\mu_{w_i^{-1}(x)})(s) \\ &= \sum_{i=1}^N \sum_{j=0}^n p_i(x) \int_{\mathbb{R}_g} c_{nj} (\alpha_i s)^j \beta_i^{n-j} d(\mu_{w_i^{-1}(x)})(s) \\ &= \sum_{i=1}^N \sum_{j=0}^n p_i(x) c_{nj} \alpha_i^j \beta_i^{n-j} \int_{\mathbb{R}_g} s^j d(\mu_{w_i^{-1}(x)})(s) \\ &= \sum_{j=0}^n \left[\sum_{i=1}^N p_i(x) c_{nj} \alpha_i^j \beta_i^{m-j} \right] g_j(w_i^{-1}(x)), \end{aligned} \tag{45}$$

where

$$c_{nj} = \begin{pmatrix} n \\ j \end{pmatrix}. \tag{46}$$

The reader may compare the above result to that of (67) for the IFSP case. The place-dependent moments $h_n(x)$ are related to the moments g_n evaluated at the preimages $w_i^{-1}(x)$. And it is the greyscale $\phi(s)$ maps that now "mix" the measures, as opposed to the spatial IFS maps $w_i(x)$ in (67).

In the special case that $\mu = \bar{\mu} = M\bar{\mu}$, the fixed point of M, then $h_n(x) = g_n(x)$, and we have

$$g_n(x) = \sum_{j=0}^{n} \left[\sum_{i=1}^{N} p_i(x) c_{nj} \alpha_i^j \beta_i^{n-j} \right] g_j(w_i^{-1}(x)). \tag{47}$$

In other words, the moments $g_n(x)$ satisfy recursion relations that involve moments of all orders up to n evaluated at preimages $w_i^{-1}(x)$. Note that this does not yield a rearrangement, analogous to (69), which will permit a simple recursive computation of the moments $g_n(x)$. Nevertheless, the moment *functions* can be computed recursively, as we now show.

First, note that for the particular case $n = 1$, the moment function $g_1(x)$ is a solution of the fixed point equation

$$g_1(x) = \sum_{i=1}^{N} p_i(x)[\alpha_i g_1(w_i^{-1}(x)) + \beta_i]. \tag{48}$$

(Note that $g_1(x)$ is the expectation value of $\mu(x)$.) But this implies that g_1 is the unique fixed point in $\mathcal{L}^1(X)$ of the contractive place-dependent IFSM operator defined by

$$(Mu)(x) = \sum_{i=1}^{N} p_i(x)[\alpha_i u(w_i^{-1}(x)) + \beta_i]. \tag{49}$$

This provides a method for computing g_1—at least approximately—by means of Banach's theorem. Starting at any $u_0 \in \mathcal{L}^1$ the sequence $M^n u = M(M^{n-1})u_0$ converges to g_1 as $n \to +\infty$.

Higher order moments can be computed in a similar recursive manner. To illustrate, consider the case $n = 2$. From (47), the moment $g_2(x)$ satisfies the fixed point equation

$$g_2(x) = \sum_{i=1}^{N} p_i(x)[\alpha_i^2 g_2(w_i^{-1}(x)) + 2\alpha_i\beta_i g_1(x) + \beta_i^2]. \tag{50}$$

In other words, g_2 is the fixed point of a contractive IFSM operator (cf. (75)), with condensation function

$$b(x) = \sum_{i=1}^{N} p_i(x)[2\alpha_i\beta_i g_1(x) + \beta_i^2]. \tag{51}$$

From a knowledge of g_1, the moment function g_2 may be computed via iteration. The process may now be iterated to produce g_3, etc.

Finally, note that g_1 and g_2 determine the pointwise variance $\sigma^2(x)$ of the measure $\bar{\mu}(x)$:

$$\begin{aligned} \sigma^2(x) &= \int_{\mathbb{R}_g} (s - g_1(s))^2 d\bar{\mu}_x(s) \\ &= g_2(x) - [g_1(x)]^2. \end{aligned} \tag{52}$$

Table 2

Optimal affine greyscale maps for each parent/child pairing R_i/D_j in the $(2,4)$ domain/range block IFSM described in the text. The final column lists the probabilities defined by (21) for the parameter value $h = 1$.

R_j	D_i	k	$\alpha_{ij}^{(k)}$	$\beta_{ij}^{(k)}$	$p_{ij}^{(k)}$
1	1	1	0.66193	−0.00497	0.30880
1	1	2	−0.65050	0.79469	0.19120
1	2	1	−0.65050	0.79469	0.19120
1	2	2	0.66193	−0.00497	0.30880
2	1	1	0.28561	0.64623	0.36206
2	1	2	−0.25802	0.97746	0.13794
2	2	1	−0.25802	0.97746	0.13794
2	2	2	0.28561	0.64623	0.36206
3	1	1	−0.25802	0.97746	0.13794
3	1	2	0.28561	0.64623	0.36206
3	2	1	0.28561	0.64623	0.36206
3	2	2	−0.25802	0.97746	0.13794
4	1	1	−0.65050	0.79469	0.19120
4	1	2	0.66193	−0.00497	0.30880
4	2	1	0.66193	−0.00497	0.30880
4	2	2	−0.65050	0.79469	0.19120

4.3. Multiparent block fractal coding and associated measure-valued image functions.

We now return to the multiparent block fractal coding scheme introduced in section 3.7. Associated with the block transform operator T in (34) will be an operator $M : Y \to Y$ defined as follows: For a $\mu \in Y$, $x \in R_i$, and any subset $S \subset [0,1]$, we define $\nu = M\mu$ as follows:

$$\nu(x)(S) = (M\mu(x))(S) = \sum_{j=1} p_{ij}\mu(w_{ij}^{-1}(x))(\phi_{ij}^{-1}(S)). \tag{53}$$

Note that *any* choice of the weights p_{ij} and the affine maps ϕ_{ij} subject to the conditions given above will produce a contractive fractal transform operator $M : Y \to Y$ with unique fixed point measure $\bar{\mu}$. However, we shall consider the weighting scheme of (21).

We illustrate with the following simple one-dimensional example: the target function

$$u(x) = 0.8\sin(\pi x) + 0.1, \quad x \in [0,1]. \tag{54}$$

Two domain blocks D_i and four range blocks R_j will be used, i.e.,

$$D_1 = [0, 0.5],\ D_2 = [0.5, 1],$$
$$R_1 = [0, 0.25],\ R_2 = [0.25, -0.5],\ R_3 = [0.5, 0.75],\ R_4 = [0.75, 1].$$

We also consider both possible geometric contractions $w_{ij}^{(k)}$ of each interval D_i to a smaller interval R_j: orientation preserving ($k = 1$) and orientation flipping ($k = 2$). The optimal greyscale maps $\phi_{ij}^{(k)}$ that minimize the collage errors $\Delta_{ij}^{(k)}$ are shown in Table 2. In the final column of this table are listed the probabilities $p_{ij}^{(k)}$ that are obtained from the exponential formula in (21) using the parameter values $P = 1$ and $h = 1$. The probabilities with the

highest value (0.36206) correspond to the best parent/child pairings that would be selected by standard fractal coding. Because of the geometrical symmetry of the target function, many of the probabilities and corresponding greyscale maps are equal. (For example, the approximation of R_1 by D_1 with zero flip should be the same as its approximation by D_2 with a flip.)

In Figure 14 we present pictorial representations of the measure-valued fixed points $\bar{\mu}(x)$ for four choices of probabilities $p_{ij}^{(k)}$ as defined by the exponential formula in (21): (i) $h \approx 0$, (ii) $h = 0.1$, (iii) $h = 1.0$, and (iv) $h = 10^4$, the latter essentially representing an infinite value of h. The $h \approx 0$ case corresponds to standard fractal coding, where only the best parent/child pairings are chosen. In this case, for each $x \in [0,1]$, the measure is concentrated around, or situated exclusively at, a greyscale value $y \in [0,1]$ that corresponds to the value $\bar{u}(x)$ of the standard fractal coding IFSM attractor. As h increases from 0, the probability measures at each x value become more "diffuse," starting essentially as Dirac unit masses on the IFSM attractor function when $h = 0$. The value $h = 10^4$ yields equal probabilities for all domain/range pairings to well over six decimal digits, i.e., $h \approx \infty$.

We now return to the *Lena* image study of sections 3.3 and 3.4 and provide the corresponding Case 4 results for measures and denoising. In Figure 15 are shown pictorial representations of the Case 4 measures $\nu(x) = (M\nu)(x)$ for the $u(256, j)$, $1 \le j \le 256$, row of pixels of the *Lena* image. These pictures complement the Case 1–3 results of Figure 7.

In Figure 16(a) is shown the pictorial representation of the Case 4 measure for the same row of pixels of the noisy normalized *Lena* image with added zero-mean Gaussian noise $\sigma = 0.1$. In 16(b) are plotted the mean values $\bar{x}$ of $\nu(x)^{(Case4)}$. And in 16(c) is plotted the variance $\sigma(x)$ of $\nu(x)^{(Case\ 4)}$. In all cases, $h = 0.1$. These figures should be compared with their Case 1–3 counterparts in Figure 8.

Now recall that the Case 4 cross-scale fractal image transform operator T and associated measure-valued transform operator M may be contractive, in which case there exist respective fixed points $\bar{u}$ and $\bar{\mu}$ of these operators. A natural question is whether the additional application of these operators will result in further denoising. In Figure 17 are shown the results of applying the Case 4 operator M to the measure $\nu(x)$ of Figure 16. On the left is plotted the mean-value curve associated with the measure $M\nu = M^2\mu$. Note that its overall shape is almost identical to the mean-value curve of ν in Figure 16. The RMSE, however, is actually slightly larger, indicating that no additional denoising has been accomplished. On the right is plotted the variance of the measure $M\nu$, which generally demonstrates lower values from those of the measure ν. The variance of $M\nu$ also demonstrates some rather repetitive self-similar features, which is worthy of further investigation. Additional applications of the M operator leave the measures, and therefore the associated plots in Figure 17, unchanged to several decimal digits of accuracy. As a result, we may consider these plots to represent the mean values and variances of the *fixed point measure* $\bar{\nu}(x)$ of M.

In earlier studies of fractal denoising, e.g., [17, 18], the first step was to compute the fractal transform T of a noisy image $\tilde{u}$. The fixed point $\bar{u}$ of T was then considered to provide an approximation of the corresponding noiseless image u. However, it is possible to improve this result by starting with T and perturbing it in an attempt to produce an estimate of the fractal transform T' of the noiseless image u and not $\tilde{u}$. The fixed point $\bar{u}'$ of this transform

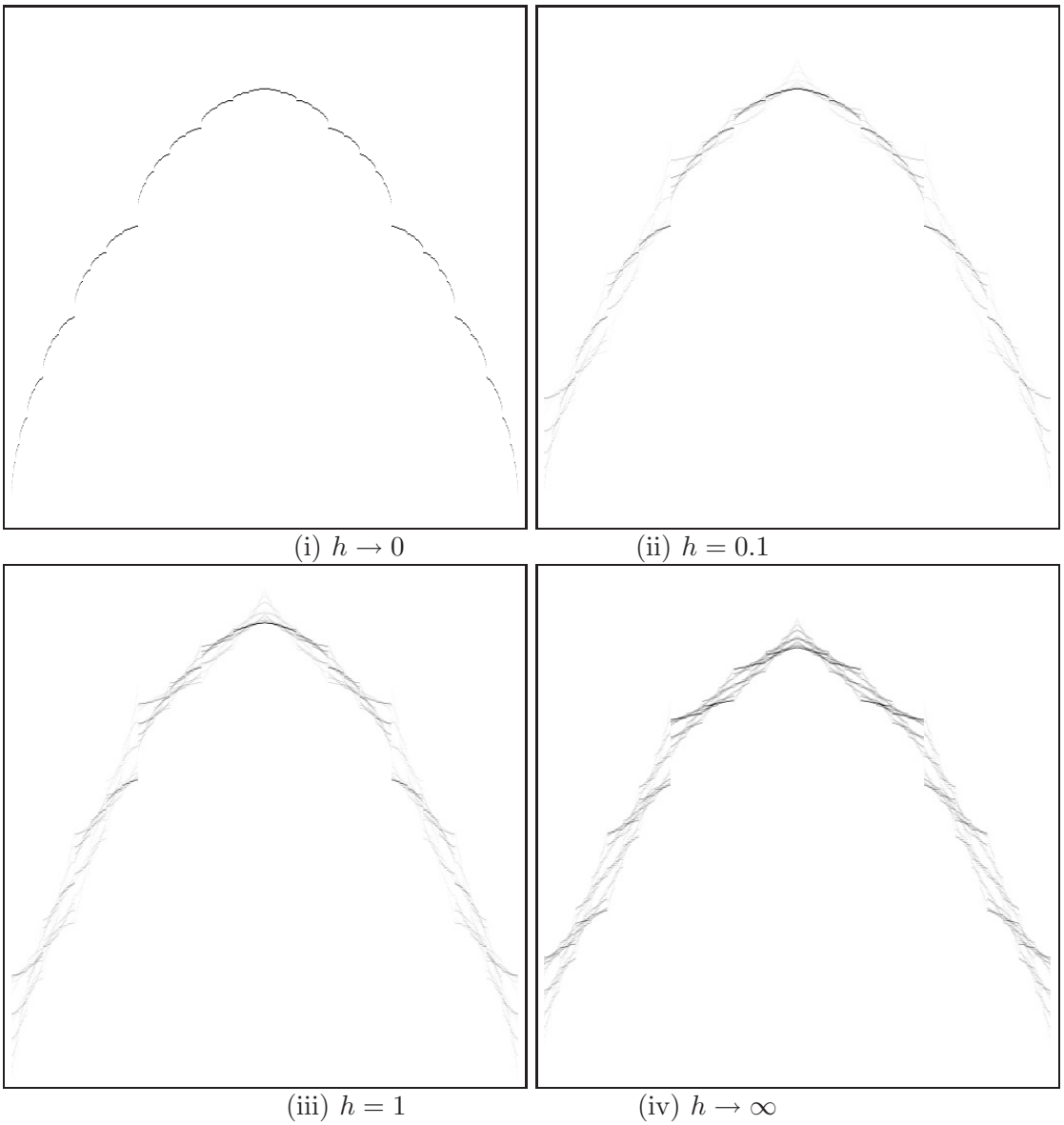
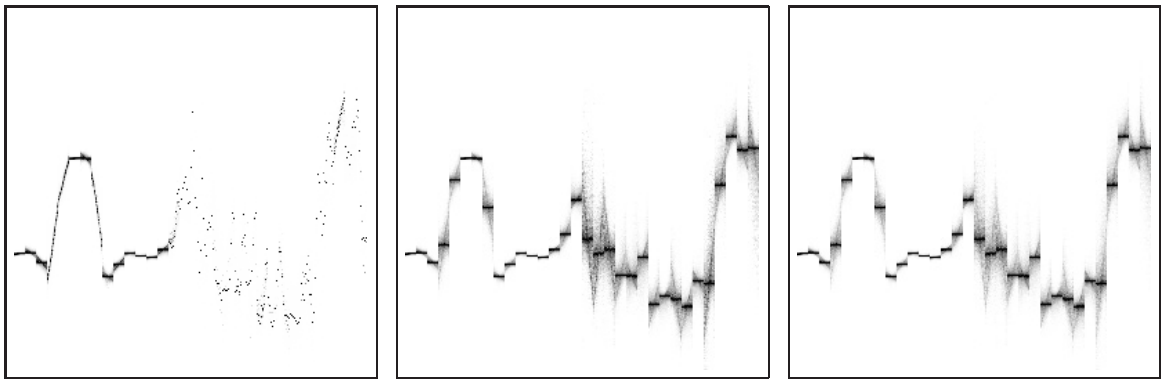


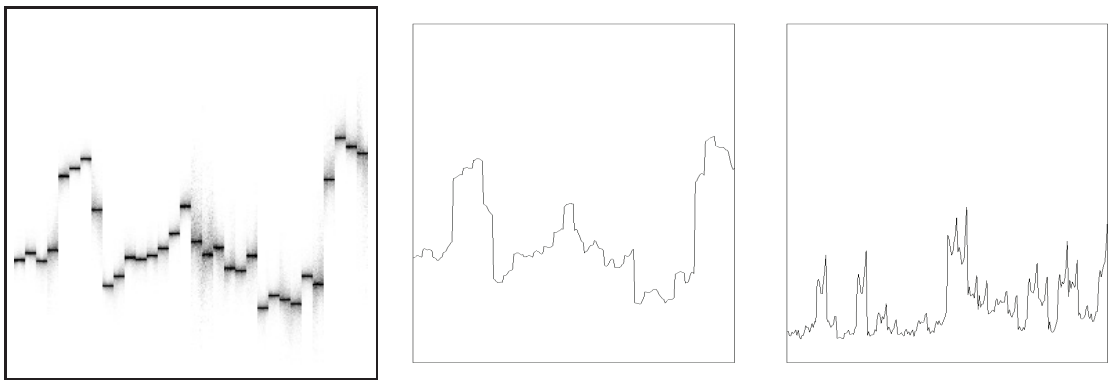
Figure 14. *Pictorial representations of the measure-valued fixed point images of the* $(2, 4)$ *domain/range block IFSM described in the text. The region is* $X \times \mathbb{R}_g = [0, 1] \times [0, 1]$. *At each* $x \in [0, 1]$, *the measure is sketched vertically. Four sets of probabilities* $p_i^{(k)}$, *as defined by the exponential formula in* (21), *were used.* (i) $h \approx 0$, *corresponding to standard fractal coding;* (ii) $h = 0.1$; (iii) $h = 1.0$, *the probabilities listed in Table* 1; (iv) $h = 10^4$ *(essentially infinity) yielding equal probabilities for all domain/range pairs.*

generally provided a better approximation to u. In all of these methods, the focus of the denoising procedure was to produce a fixed point approximation to the noiseless image. In contrast, the same-scale (Cases 1–3) as well as the cross-scale, fractal-based (Case 4) denoising methods outlined in this paper produce approximations of a noiseless image u by means of a transformation of the noisy image $\bar{u}$.



(a) Case 4, $h = 0.01$ (b) Case 4, $h = 0.1$ (c) Case 4, $h = 1.0$

Figure 15. *Pictorial representations of measures* $\nu(x)^{(Case\ 4)}$ *for the* $u(256, j)$, $1 \le j \le 256$, *row of pixels of the* Lena *image. Three values of the weighting parameter* h *in* (21) *have been used, with* $P = 2$. *These pictures complement the Case* 1–3 *results of Figure* 7. 8×8*-pixel range blocks,* 16×16*-pixel domain blocks.*

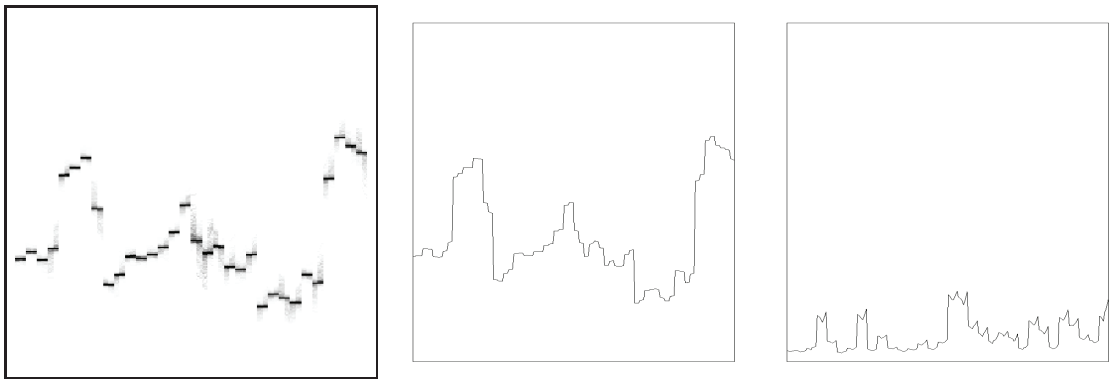


(a) $\nu(x)^{(Case\ 4)}$ (b) $\bar{x}$, RMSE $= 0.092$ (c) $\sigma(x)$

Figure 16. (a) *Pictorial representation of measure* $\nu(x)^{(Case\ 4)}$ *for the* $u(256, j)$, $1 \le j \le 256$, *row of pixels for the normalized* Lena *image plus Gaussian noise* $\sigma = 0.1$. (b) *Mean values* $\bar{x}$ *of* $\nu(x)^{(Case4)}$. (c) *Variance* $\sigma(x)$ *of this measure—vertical scale* $[0, 0.15]$. *In all cases,* $h = 0.1$. *These figures complement the Case* 1–3 *results of Figure* 8. 8×8*-pixel range blocks,* 16×16*-pixel domain blocks.*

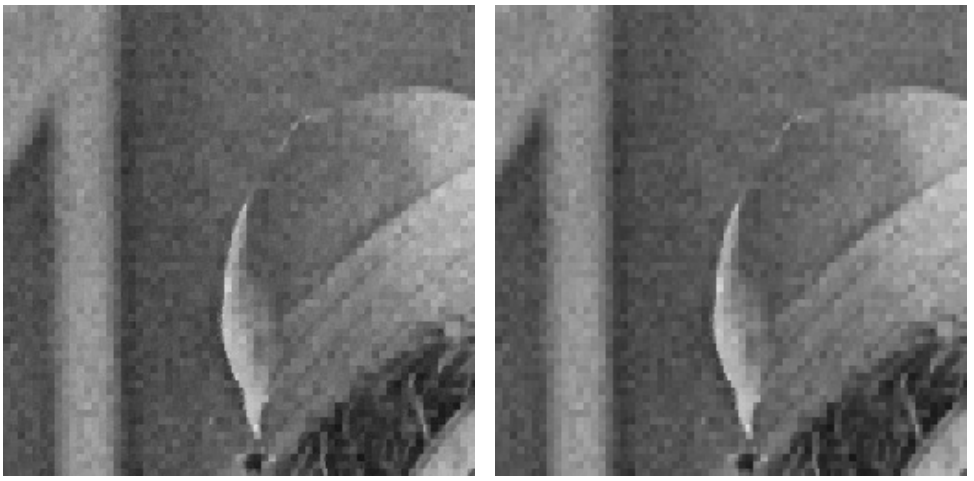
When h is decreased to 0.05 and 4×4-pixel range blocks are used, the resulting mean-value estimate of the denoised image is very similar to the Case 3 result of Figure 9, with the same RMSE value of 0.062. As such, we shall not show the result. Instead we show, in Figure 18, the results of applying the Case 4 method to the entire upper quarter of the *Lena* image. A comparison of these results with those of Case 1–3 denoising in Figure 10 shows that the cross-scale, Case 4 denoising procedure seems to afford no significant improvement over same-scale, Case 3 denoising, either visually or in terms of RMSE. A more exhaustive investigation is required, however.

5. Concluding remarks. In this paper we have constructed an appropriate complete metric space (Y, d_Y) of measure-valued images $\mu : [0, 1]^n \to \mathcal{M}([0, 1])$. Furthermore, we have shown how such a formalism is applicable to NL image processing. In particular, it was shown



(a) $\bar{\nu}^{(Case\ 4)}(x)$ (b) $\bar{x}$, RMSE $= 0.094$ (c) $\sigma(x)$

Figure 17. (a) *Pictorial representation of invariant measure* $M\nu(x)^{(Case\ 4)} = M^2\mu$ *for the* $u(256, j)$, $1 \le j \le 256$, *row of pixels for the normalized* Lena *image plus Gaussian noise* $\sigma = 0.1$. (b) *Mean values* $\bar{x}$ *of* $\nu(x)^{(Case4)}$. (c) *Variance* $\sigma(x)$ *of this measure—vertical scale* $[0, 0.15]$. *In all cases,* $h = 0.1$.



(a) $\nu^{(Case\ 4)} = M\mu$, RMSE $= 0.039$ (b) $\bar{\nu}^{(Case\ 4)}$, RMSE $= 0.039$

Figure 18. *Results of Case* 4 *denoising of upper quarter of* Lena *image. Left: One application of Case* 4 *fractal transform operator* M. *Right: Fixed point of* M. *These figures complement those of Figure* 10.

how the space (Y, d_Y) can be used with a simple model of self-similarity of images that includes both same-scale as well as cross-scale similarity.

In the same-scale case, we focused particular attention on the method of NL-means denoising. The measure-valued image approach may be viewed as a kind of intermediate step which accounts for all greyscale values, as well as their weights, that contribute to the estimate of a denoised greyscale value.

In the cross-scale case, we showed how the space (Y, d_Y) may be associated with a multi-parent method of block fractal coding. For this purpose, a method of fractal transforms was formulated over this space. Under suitable conditions, the fractal transform operator $M : Y \to Y$ is contractive, implying the existence of a unique fixed point measure $\bar{\mu} = T\bar{\mu}$.

The pointwise moments of this measure were also shown to satisfy a set of *place-dependent* recursion relations.

Generally, images are locally self-similar to a high degree: The image supported over a given range block R_i is generally well approximated by images supported on a good number of domain blocks D_j after spatial contraction and greyscale modification. We have shown, by means of a number of examples, how the measure-valued image formalism can be used to capture such self-similarities throughout an image. We are currently exploring in more detail the role of the variance function $\sigma(x)$ in characterizing self-similarity. As well, we are exploring the use of the measure-valued image formalism in other NL image processing algorithms such as zooming.

6. Appendix. Essentials of iterated function systems.

6.1. Basic results involving contractive maps.

For the benefit of the reader, we mention some important results for contractive maps. They provide the basis for fractal-based approximation methods.

Theorem 6.1 (Banach). *Let (X, d) be a complete metric space. Also let $T : X \to X$ be a contraction mapping with contraction factor $c \in [0, 1)$; i.e., for all $x, y \in X$, $d(Tx, Ty) \leq cd(x, y)$. Then there exists a unique $\bar{x} \in X$ such that $\bar{x} = T\bar{x}$. Moreover, for any $x \in X$, $d(T^n x, \bar{x}) \to 0$ as $n \to \infty$.*

A simple triangle inequality along with Banach's theorem yields the following result.

Theorem 6.2 (see "collage theorem" [5, 3]). *Let (X, d) be a complete metric space and let $T : X \to X$ be a contraction mapping with contraction factor $c \in [0, 1)$. Then for any $x \in X$,*

$$d(x, \bar{x}) \leq \frac{1}{1-c} d(x, Tx), \tag{55}$$

where $\bar{x}$ is the fixed point of T.

Another manipulation of the triangle inequality involving x, Tx, and $\bar{x}$ yields the following interesting result.

Theorem 6.3 ("anti-collage theorem" [25]). *Assume the conditions of the collage theorem above. Then for any $x \in Y$,*

$$d(x, \bar{x}) \geq \frac{1}{1+c} d(x, Tx). \tag{56}$$

Theorem 6.4 ("continuity of fixed points" [8]). *Let (Y, d_Y) be a complete metric space and T_1, T_2 be two contractive mappings with contraction factors c_1 and c_2 and fixed points y_1^* and y_2^*, respectively. Then*

$$d_Y(y_1^*, y_2^*) \leq \frac{1}{1-c} d_{Y,\sup}(T_1, T_2), \tag{57}$$

where

$$d_{Y,\sup}(T_1, T_2) = \sup_{x \in X} d(T_1(x), T_2(y)) \tag{58}$$

and $c = \min\{c_1, c_2\}$.

6.2. Iterated function systems and invariant measures. In what follows, (X, d) denotes a compact metric "base space," typically $[0,1]^n$. Let $\mathbf{w} = \{w_1, \ldots, w_N\}$ be a set of 1-1 contraction maps $w_i : X \to X$, to be referred to as an N-map IFS. Let $c_i \in [0,1)$ denote the contraction factors of the w_i and define $c = \max_{1 \le i \le N} c_i$. Note that $c \in [0,1)$.

Now let $\mathcal{H}(X)$ denote the set of nonempty compact subsets of X and h denote the Hausdorff metric. Then $(\mathcal{H}, h)$ is a complete metric space [11]. Associated with the IFS maps w_i is a set-valued mapping $\hat{\mathbf{w}} : \mathcal{H}(X) \to \mathcal{H}(X)$, the action of which is defined to be

$$\hat{\mathbf{w}}(S) = \bigcup_{i=1}^{N} w_i(S), \qquad S \in \mathcal{H}(X), \tag{59}$$

where $w_i(S) := \{w_i(x), x \in S\}$ is the image of S under w_i, $i = 1, 2, \ldots, N$.

Theorem 6.5 (see [19]). *$\hat{\mathbf{w}}$ is a contraction mapping on $(\mathcal{H}(X), h)$:*

$$h(\mathbf{w}(A), \mathbf{w}(B)) \le ch(A, B), \quad A, B \in \mathcal{H}(X). \tag{60}$$

Corollary 6.6. *There exists a unique set $A \in \mathcal{H}(X)$ such that $\mathbf{w}(A) = A$, the so-called* attractor *of the IFS $\mathbf{w}$. Moreover, for any $S \in \mathcal{H}(X)$, $h(\mathbf{w}^n(S), A) \to 0$ as $n \to \infty$.*

Now let $\mathcal{M}(X)$ denote the set of Borel probability measures on X and d_H denote the Monge–Kantorovich metric on this set:

$$d_H(\mu, \nu) = \sup_{f \in Lip_1(X, \mathbb{R})} \left[\int_X f(x) d\mu - \int_X f(x) d\nu \right], \tag{61}$$

where

$$Lip_1(X, \mathbb{R}) = \{f : X \to \mathbb{R} \mid |f(x_1) - f(x_2)| \le d(x_1, x_2) \ \ \forall x_1, x_2 \in X\}. \tag{62}$$

For $i \le i \le N$, let $0 < p_i < 1$ be a partition of unity associated with the IFS maps w_i so that $\sum_{i=1}^{N} p_i = 1$. Associated with this IFSP $(\mathbf{w}, \mathbf{p})$ is the so-called *Markov operator*, $M : \mathcal{M}(X) \to \mathcal{M}(X)$, the action of which is

$$\nu(S) = (M\mu)(S) = \sum_{i=1}^{N} p_i \mu(w_i^{-1}(S)) \quad \forall S \in \mathcal{H}(X). \tag{63}$$

(Here, $w_i^{-1}(S) = \{y \in X \mid w_i(y) \in S\}$.)

Theorem 6.7 (see [19]). *M is a contraction mapping on $(\mathcal{M}(X), d_H)$:*

$$d_H(M\mu, M\nu) \le c d_H(\mu, \nu), \quad \mu, \nu \in \mathcal{M}(X). \tag{64}$$

Corollary 6.8. *There exists a unique measure $\bar{\mu} \in \mathcal{M}(X)$, the so-called* invariant measure *of the IFSP $(\mathbf{w}, \mathbf{p})$, such that $\bar{\mu} = M\bar{\mu}$. Moreover, for any $\mu \in \mathcal{M}(X)$, $d_H(M^n\mu, \bar{\mu}) \to 0$ as $n \to \infty$.*

The reader is referred to the book by Barnsley [3] for more detailed discussions as well as numerous examples.

6.3. Moment relations for measures and IFSP. In applications, it is most convenient to employ affine IFS maps. In this case, the moments of the invariant measure $\bar{\mu}$ of the Markov operator M satisfy a set of relations that allow them to be computed recursively [4, 3, 13]. We illustrate with the one-dimensional case, i.e., $X = [0, 1]$. The extension to higher dimensions is quite straightforward.

The affine IFS maps will be denoted as

$$w_i(x) = s_i x + a_i, \quad i = 1, 2, \ldots, N. \tag{65}$$

The moments of a probability measure $\mu \in \mathcal{M}(X)$ are defined as

$$g_n = \int_X x^n d\mu, \quad n = 0, 1, 2, \ldots. \tag{66}$$

By definition, $g_0 = 1$. Now let $\nu = M\mu$. Then, from (63), the moments of ν are given by

$$\begin{aligned} h_n &= \int_X x^n d(M\mu)(x) \\ &= \int_X x^n d\left(\sum_{i=1}^N p_i \mu \circ w_i^{-1} \right)(x) \\ &= \int_X \sum_{i=1}^N p_i x^n d(\mu \circ w_i^{-1})(x) \\ &= \int_X \sum_{i=1}^N p_i [w_i(x)]^n d\mu(x) \\ &= \int_X \sum_{i=1}^N p_i [s_i x + a_i]^n d\mu(x). \end{aligned}$$

Expansion of the binomial followed by an interchange of summation and integration yields the result

$$h_n = \sum_{j=0}^n \left(\begin{array}{c} n \\ j \end{array} \right) \left[\sum_{i=1}^N p_i s_i^j a_i^{n-j} \right] g_j. \tag{67}$$

If we let

$$\mathbf{g} = (g_0, g_1, \ldots)^T, \quad \mathbf{h} = (h_0, h_1, \ldots)^T \tag{68}$$

denote the (infinite) moment vectors of μ and ν, respectively, then the Markov operator M is seen to induce a linear mapping $\mathbf{h} = A\mathbf{g}$, where A is represented by a lower triangular matrix. This was originally pointed out in [13].

In the special case that $\mu = \bar{\mu} = M\bar{\mu}$, the invariant measure of the IFSP, then $h_n = g_n$ and (67) can be rearranged to yield

$$\left(1 - \sum_{i=1}^N p_i s_i^n \right) g_n = \sum_{j=0}^{n-1} \left(\begin{array}{c} n \\ j \end{array} \right) \left[\sum_{i=1}^N p_i s_i^j a_i^{n-j} \right] g_j. \tag{69}$$

This result, originally derived in [4], shows that the moments g_n of the invariant measure $\bar{\mu}$ may be computed recursively, starting with $g_0 = 1$.

6.4. IFS on function spaces. For the moment, we consider a general function space $\mathcal{F}(X)$ supported on X. The essential components of a *fractal transform operator* are as follows:

1. a set of N one-to-one contraction maps $w_i : X \to X$ with the condition that $\cup_{i=1}^{N} w_i(X) = X$;
2. a set of associated *greyscale maps* $\phi_i : \mathbb{R} \to \mathbb{R}$ that are assumed to be Lipschitz on $\mathbb{R}$; i.e., for each ϕ_i there exists a $K_i \geq 0$ such that

$$|\phi_i(t_1) - \phi_i(t_2)| \leq K_i |t_1 - t_2| \quad \forall\, t_1, t_2 \in \mathbb{R}. \tag{70}$$

The above two sets of maps are said to comprise an "iterated function system with greyscale maps" (IFSM) $(\mathbf{w}, \Phi)$ [14]. For each $x \in X$, this IFSM produces one or more *fractal components* defined as

$$g_i(x) = \begin{cases} \phi_i(u(w_i^{-1}(x))) & \text{if } \ x \in w_i(X), \\ 0 & \text{otherwise.} \end{cases} \tag{71}$$

If several fractal components exist for an $x \in X$, then they are combined with an operation that is suitable for the space in which we are working (see [14] for more details and examples of the various function spaces that can be considered). The natural operation in $\mathcal{L}^p(X)$ is the summation operation: For $u \in \mathcal{L}^p(X)$,

$$v(x) = (Tu)(x) = \sum_{i=1}^{N} g_i(x). \tag{72}$$

Theorem 6.9. *Let $(\mathbf{w}, \Phi)$ be an IFSM, with spatial contractions w_i and Lipschitz greyscale maps ϕ_i. Then for $p \geq 1$ and $u, v \in \mathcal{L}^p(X)$,*

$$\| Tu - Tv \| \leq \sum_{i=1}^{N} c_i^{1/p} K_i \, \| u - v \| \,. \tag{73}$$

Corollary 6.10. *If $c = \sum_{i=1}^{N} c_i K_i^p < 1$, then T is contractive in $\mathcal{L}^p(X)$ with fixed point $\bar{u} \in \mathcal{L}^p(X)$. The fixed point equation*

$$\bar{u}(x) = (T\bar{u})(x) = \sum_{i=1}^{N} \phi_i(\bar{u}(w_i^{-1}))(x) \tag{74}$$

indicates that $\bar{u}$ is "self-similar," i.e., that it can be written as a sum of spatially contracted and greyscale-modified copies of itself.

It is also convenient to define IFSM operators with *condensation functions*. For example, given a set of IFS maps w_i, associated constants α_i, and condensation function $b(x)$, $x \in X$, define the action of the associated operator T as follows: For $u \in \mathcal{L}(X)$,

$$v(x) = (Tu)(x) = b(x) + \sum_{i=1}^{N} \alpha_i u(w_i^{-1}(x)). \tag{75}$$

Acknowledgments. This work was started during a research visit by DLT to the Department of Applied Mathematics of the University of Waterloo. DLT thanks ERV for this opportunity. Finally, the authors would like to thank the referees as well as the Associate Editor for comments, criticisms, and suggestions that greatly enhanced the quality of this paper.

REFERENCES

[1] S. K. Alexander, *Multiscale Methods in Image Modelling and Image Processing*, Ph.D. thesis, Department of Applied Mathematics, University of Waterloo, Waterloo, ON, 2005.
[2] S. K. Alexander, E. R. Vrscay, and S. Tsurumi, *A simple, general model for the affine self-similarity of images*, in International Conference on Image Analysis and Recognition (ICIAR 2008), Lecture Notes in Comput. Sci. 5112, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 192–203.
[3] M. F. Barnsley, *Fractals Everywhere*, Academic Press, New York, 1989.
[4] M. F. Barnsley and S. Demko, *Iterated function systems and the global construction of fractals*, Proc. Roy. Soc. London Ser. A, 399 (1985), pp. 243–275.
[5] M. F. Barnsley, V. Ervin, D. Hardin, and J. Lancaster, *Solution of an inverse problem for fractals and other sets*, Proc. Nat. Acad. Sci. USA, 83 (1985), pp. 1975–1977.
[6] M. F. Barnsley and L. P. Hurd, *Fractal Image Compression*, A.K. Peters, Wellesley, MA, 1993.
[7] A. Buades, B. Coll, and J. M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul., 4 (2005), pp. 490–530.
[8] P. Centore and E. R. Vrscay, *Continuity of fixed points for attractors and invariant measures for iterated function systems*, Canad. Math. Bull., 37 (1994), pp. 315–329.
[9] M. Ebrahimi and E. R. Vrscay, *Solving the inverse problem of image zooming using "self-examples,"* in Image Analysis and Recognition, Lecture Notes in Comput. Sci. 4633, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 117–130.
[10] M. Elad and D. Datsenko, *Example-based regularization deployed to super-resolution reconstruction of a single image*, Comput. J., 50 (2007), pp. 1–16.
[11] K. Falconer, *The Geometry of Fractal Sets*, Cambridge University Press, Cambridge, UK, 1985.
[12] Y. Fisher, *Fractal Image Compression, Theory and Application*, Springer-Verlag, New York, 1995.
[13] B. Forte and E. R. Vrscay, *Solving the inverse problem for measures using iterated function systems: A new approach*, Adv. in Appl. Probab., 27 (1995), pp. 800–820.
[14] B. Forte and E. R. Vrscay, *Theory of generalized fractal transforms*, in Fractal Image Encoding and Analysis, NATO ASI Ser. F 159, Y. Fisher, ed., Springer-Verlag, Berlin, 1998, pp. 145–168.
[15] B. Forte and E. R. Vrscay, *Inverse problem methods for generalized fractal transforms*, in Fractal Image Encoding and Analysis, NATO ASI Ser. F 159, Y. Fisher, ed., Springer-Verlag, Berlin, 1998, pp. 169–199.
[16] W. T. Freeman, T. R. Jones, and E. C. Pasztor, *Example-based super-resolution*, IEEE Comput. Graph. Appl., 22 (2002), pp. 56–65.
[17] M. Ghazel, G. Freeman, and E. R. Vrscay, *Fractal image denoising*, IEEE Trans. Image Process., 12 (2003), pp. 1560–1578.
[18] M. Ghazel, G. Freeman, and E. R. Vrscay, *Fractal-wavelet image denoising revisited*, IEEE Trans. Image Process., 15 (2006), pp. 2669–2675.
[19] J. Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J., 30 (1981), pp. 713–747.
[20] A. Jacquin, *Image coding based on a fractal theory of iterated contractive image transformations*, IEEE Trans. Image Process., 1 (1992), pp. 18–30.
[21] M. Kisielewicz, *Differential Inclusions and Optimal Control*, Math. Appl. 44, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.
[22] H. Kunze and E. R. Vrscay, *Solving inverse problems for ordinary differential equations using the Picard contraction mapping*, Inverse Problems, 15 (1999), pp. 745–770.

[23] D. La Torre, F. Mendivil, and E. R. Vrscay, *Iterated function systems on multifunctions*, in Math Everywhere: Deterministic and Stochastic Modelling in Biomedicine, Economics and Industry, G. Aletti, M. Burger, A. Micheletti, and D. Morale, eds., Springer-Verlag, Berlin, 2007, pp. 125–138.
[24] N. Lu, *Fractal Imaging*, Academic Press, New York, 2003.
[25] E. R. Vrscay and D. Saupe, *Can one break the "collage barrier" in fractal image coding?*, in Fractals: Theory and Applications in Engineering, M. Dekking, J. Levy-Vehel, E. Lutton, and C. Tricot, eds., Springer-Verlag, London, 1999, pp. 307–323.
[26] D. Zhang and Z. Wang, *Image information restoration based on long-range correlation*, IEEE Trans. Cir. Syst. Video Tech., 12 (2002), pp. 331–341.