

# Structural Similarity-Based Affine Approximation and Self-Similarity of Images Revisited

Dominique Brunet<sup>1</sup>, Edward R. Vrscay<sup>1</sup>, and Zhou Wang<sup>2</sup>

<sup>1</sup> Department of Applied Mathematics, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

<sup>2</sup> Department of Electrical and Computer Engineering, Faculty of Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1  
dbrunet@uwaterloo.ca, ervrscay@uwaterloo.ca, zhouwang@ieee.org

**Abstract.** Numerical experiments indicate that images, in general, possess a considerable degree of affine self-similarity, that is, blocks are well approximated in root mean square error (RMSE) by a number of other blocks when affine greyscale transformations are employed. This has led to a simple  $L^2$ -based model of affine image self-similarity which includes the method of fractal image coding (cross-scale, affine greyscale similarity) and the nonlocal means denoising method (same-scale, translational similarity). We revisit this model in terms of the structural similarity (SSIM) image quality measure, first deriving the optimal affine coefficients for SSIM-based approximations, and then applying them to various test images. We show that the SSIM-based model of self-similarity removes the “unfair advantage” of low-variance blocks exhibited in  $L^2$ -based approximations. We also demonstrate experimentally that the local variance is the principal factor for self-similarity in natural images both in RMSE and in SSIM-based models.

**Key words:** self-similarity, structural similarity index, affine approximation, image model, non-local image processing

## 1 Introduction

The effectiveness of a good number of nonlocal image processing methods, including nonlocal-means denoising [1], restoration [2, 3], compression [4], super-resolution [5–7] and fractal image coding [8–10], is due to how well pixel-blocks of an image can, in some way, be approximated by other pixel blocks of the image. This property of natural images may be viewed as a form of *self-similarity*.

In [11], a simple model of affine self-similarity which includes a number of nonlocal image processing methods as special cases was introduced. (It was analyzed further in [12].) An image  $I$  will be represented by an image function  $u : X \rightarrow R_g$ , where  $R_g \subset \mathbf{R}$  denotes the *greyscale range*. Unless otherwise specified, we work with normalized images, i.e.,  $R_g = [0, 1]$ . The support  $X$  of an image function  $u$  is assumed to be an  $n_1 \times n_2$ -pixel array. Let  $\mathcal{R}$  be a set of

$n \times n$ -pixel subblocks  $R_i$ ,  $1 \leq i \leq N_R$  such that  $X = \cup_i R_i$ , i.e.,  $\mathcal{R}$  forms a covering of  $X$ . We let  $u(R_i)$  denote the portion of  $u$  that is supported on  $R_i$ .

We examine how well an image block  $u(R_i)$  is approximated by other image blocks  $u(R_j)$ ,  $j \neq i$ . Let us consider a block  $u(R_i)$  being approximated as the *range block* and a block  $u(R_j)$ ,  $j \neq i$ , approximating it as the *domain block*. In order to distinguish the roles of these blocks, we shall denote the domain blocks as  $u(D_j)$  with the understanding that  $D_j = R_j$ . For two pixel blocks  $R_i$  and  $D_j$ , the approximation of an image range block  $u(R_i)$  by a domain block  $u(D_j)$  may be written in the following general form,

$$u(R_i) \approx \alpha_{ij}u(D_j) + \beta_{ij}, \quad i \neq j. \quad (1)$$

The error associated with the approximation in (1) will be defined as

$$\Delta_{ij} = \min_{\alpha, \beta \in \Pi} \|u(R_i) - \alpha u(D_j) - \beta\|, \quad i \neq j, \quad (2)$$

where  $\|\cdot\|$  denotes the  $L^2(X)$  norm (or RMSE) and where  $\Pi \subset \mathbf{R}^2$  denotes the  $(\alpha, \beta)$  parameter space appropriate for each case.

The affine self-similarity model was comprised of four cases. The optimal parameters and associated errors for each case will be given. In what follows, we denote  $\mathbf{x} = u(R_i)$ ,  $\mathbf{y} = u(D_j)$  and  $N = n^2$ . The statistical measures  $s_{\mathbf{x}}$ ,  $s_{\mathbf{y}}$ , etc., are defined in (7) below.

**Case 1: Purely translational.** This is the strictest view of similarity: Two image subblocks  $u(R_i)$  and  $u(D_j)$  are considered to be “close,”  $u(R_i) \approx u(D_j)$ , if the  $L^2$  distance  $\|u(R_i) - u(D_j)\|$  is small. This is the basis of nonlocal means denoising. There is no optimization here:  $\alpha_{ij} = 1$ ,  $\beta_{ij} = 0$  and the approximation error is simply

$$\Delta_{ij}^{(Case\ 1)} = \|\mathbf{x} - \mathbf{y}\| = N^{-1/2} \sqrt{(N-1)[s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 - 2s_{\mathbf{xy}}] + [\bar{\mathbf{x}} - \bar{\mathbf{y}}]^2}. \quad (3)$$

**Case 2: Translational + greyscale shift.** This is a slightly relaxed definition of similarity. Two image subblocks are considered similar if they are close up to a greyscale shift, i.e.,  $u(R_i) \approx u(D_j) + \beta$ . This simple adjustment can improve the nonlocal means denoising method since more blocks are available in the averaging process. In this case,  $\alpha_{ij} = 1$  and we optimize over  $\beta_{ij}$ :

$$\beta_{ij} = \bar{\mathbf{x}} - \bar{\mathbf{y}}, \quad \Delta_{ij}^{(Case\ 2)} = \left[ \frac{N-1}{N} \right]^{1/2} [s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 - 2s_{\mathbf{xy}}]^{1/2}. \quad (4)$$

**Case 3: Affine transformation.** A further relaxation is afforded by allowing affine greyscale transformations, i.e.,  $u(R_i) \approx \alpha u(D_j) + \beta$ . This method has been employed in vector quantization [4]. We optimize over  $\alpha$  and  $\beta$ .

$$\alpha_{ij} = \frac{s_{\mathbf{xy}}}{s_{\mathbf{y}}^2}, \quad \beta_{ij} = \bar{\mathbf{x}} - \alpha_{ij}\bar{\mathbf{y}}, \quad \Delta_{ij}^{(Case\ 3)} = \left[ \frac{N-1}{N} \right]^{1/2} \left[ s_{\mathbf{x}}^2 - \frac{s_{\mathbf{xy}}^2}{s_{\mathbf{y}}^2} \right]^{1/2}. \quad (5)$$

**Case 4: Cross-scale affine transformation.**  $u(R_i) \approx \alpha u(w(D_j)) + \beta$ , where  $D_j$  is larger than  $R_i$  and where  $w$  is a contractive spatial transformation. This is the basis of fractal image coding. The optimization process and the error distributions for Case 4 are almost identical to those of Case 3. For this reason, this case will not be discussed in this paper.

*Note 1.* In both Cases 2 and 3, the means of the range block and optimally transformed range block are equal, i.e.,  $\bar{x} = \alpha \bar{y} + \beta$ .

Of particular interest in [11] were the distributions of  $L^2$  errors denoted as  $\Delta_{ij}^{(Case\ k)}$ , in approximating range blocks  $u(R_i)$  by all other domain blocks  $u(D_j)$ ,  $j \neq i$ , for the cases  $1 \leq k \leq 3$ . In order to reduce the computational cost, we employ nonoverlapping subblocks. Normally, one could consider eight affine spatial transformations that map a square spatial domain block  $D_j$  to a square range block  $R_i$ . In our computations, however, unless otherwise specified, we shall consider only the identity transformation, i.e., zero rotation.

In Fig. 1 are shown the Case 1-3  $\Delta$ -error distributions for all possible matches for the *Lena* and *Mandrill* images using  $8 \times 8$ -pixel blocks.

As we move from Case 1 to Case 3 above, the error in approximating a given range block  $u(R_i)$  by a given domain block  $u(D_j)$  will generally decrease, since more parameters are involved in the fitting. It was observed the Case 3  $\Delta$ -error distributions for images demonstrate significant peaking near zero error, indicating that *blocks of these images are generally very well approximated by other blocks under the action of an affine greyscale transformation.*

For a given Case  $k$ , the  $\Delta$ -error distributions of some images were observed to be more concentrated near zero approximation error than others. The former images were viewed as possessing greater degrees of self-similarity than the latter. A quantitative characterization of relative degrees of self-similarity was also considered in terms of the means and variances of the error distributions. To illustrate, for the seven well-known test images employed in the study, the degree of Case 3 self-similarity could be ordered as follows:

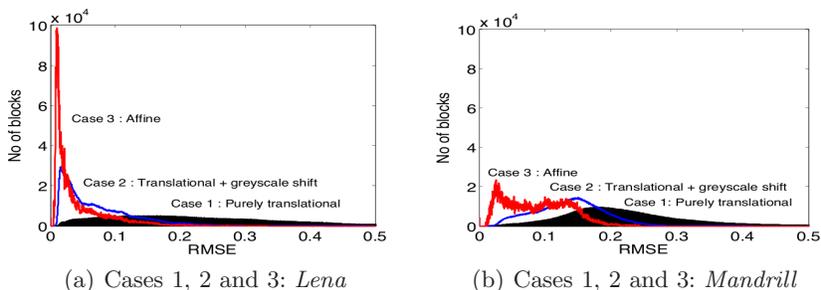
$$\begin{aligned} Lena \approx San\ Francisco > Peppers > Goldhill > \\ Boat > Barbara > Mandrill. \end{aligned}$$

It is important to note that the above model of self-similarity was based on the  $L^2$  distance measure since all  $\Delta$ -errors were in terms of root mean squared errors (RMSE) and the optimal greyscale coefficients  $\alpha$  and  $\beta$  were determined by minimizing the RMSE approximation error. Of course, this is not surprising since  $L^2$ -based distance measures, e.g., MSE, RMSE, PSNR, are the most widely used measures in image processing. However, it is well known [13] that  $L^2$ -based measures are not necessarily good measures of visual quality. In this paper, we re-examine the above self-similarity model in terms of the structural similarity (SSIM) image quality measure [14]. SSIM was proposed as an improved measure of assessing visual distortions between two images. The first step is to determine the formulas for optimal SSIM-based approximations of image range blocks by

domain blocks which correspond to Cases 1-3 above. We then present the distributions of SSIM measures between domain and range blocks for the *Lena* and *Mandrill* test images which, from above, lie on opposite ends of the  $L^2$ -based self-similarity spectrum.

It turns out that our SSIM-based results allow us to address the question, raised in [11], whether the self-similarity of an image is actually due to the *approximability* of its blocks which, in turn, is determined by their “flatness.” If range blocks of low standard deviation/variance are easier to approximate, then perhaps a truer measure of self-similarity (or lack thereof) may be obtained if their corresponding  $\Delta$  approximation errors are magnified appropriately to adjust for this “unfair advantage”. We shall show that the SSIM measure, because of its connection with a *normalized metric*, takes this “unfair advantage” into account, resulting in much less of the peaking near zero error demonstrated by RMSE approximation errors.

As shown in [11], the histogram distributions of the standard deviations  $s_{u(R_i)}$  of the  $8 \times 8$ -pixel range blocks of both are virtually identical to the Case 3  $\Delta$ -error distributions in Fig. 1. This is to be expected since the standard deviation of the image subblock  $u(R_i)$  is the RMSE associated with the approximation by its mean:  $u(R_i) \approx \bar{u}(R_i)$ . This is, in turn, a suboptimal form of the Case 3 approximation obtained by fixing the greyscale parameter  $\alpha = 0$ . The distribution of  $\alpha$  greyscale parameters is, however, found to be highly concentrated at zero [11], implying that in most cases, the standard deviation is a very good estimate of the Case 3  $\Delta$ -error.



**Fig. 1.** Case 1-3 RMS  $\Delta$ -error distributions for normalized *Lena* and *Mandrill* images over the interval  $[0, 0.5]$ . In all cases, nonoverlapping  $8 \times 8$ -pixel blocks  $R_i$  and  $D_j$  were used.

## 2 Structural Similarity and Its Use in Self-Similarity

As mentioned earlier, the structural similarity (SSIM) index [14] was proposed as an improved measure of assessing visual distortions between two images. If one of the images being compared is assumed to have perfect quality, the SSIM value

can also be interpreted as a perceptual quality measure of the second image. It is in this way that we employ it in our self-similarity study.

We express the SSIM between two blocks as a product of two components that measure (i) the similarities of their mean values and (ii) their correlation and contrast distortion. In what follows, in order to simplify the notation, we let  $\mathbf{x}, \mathbf{y} \in \mathbf{R}_+^N$  denote two non-negative  $N$ -dimensional signal/image blocks, e.g.,  $\mathbf{x} = (x_1, x_2, \dots, x_N)$ . The SSIM between  $\mathbf{x}$  and  $\mathbf{y}$  is defined as follows,

$$S(\mathbf{x}, \mathbf{y}) = S_1(\mathbf{x}, \mathbf{y})S_2(\mathbf{x}, \mathbf{y}) = \left[ \frac{2\bar{\mathbf{x}}\bar{\mathbf{y}} + \epsilon_1}{\bar{\mathbf{x}}^2 + \bar{\mathbf{y}}^2 + \epsilon_1} \right] \left[ \frac{2s_{\mathbf{x}\mathbf{y}} + \epsilon_2}{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 + \epsilon_2} \right], \quad (6)$$

where

$$\begin{aligned} \bar{\mathbf{x}} &= \frac{1}{N} \sum_{i=1}^N x_i, \quad \bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^N y_i, \\ s_{\mathbf{x}}^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{\mathbf{x}})^2, \quad s_{\mathbf{y}}^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{\mathbf{y}})^2, \\ s_{\mathbf{x}\mathbf{y}} &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{\mathbf{x}})(y_i - \bar{\mathbf{y}}). \end{aligned} \quad (7)$$

The small positive constants  $\epsilon_1, \epsilon_2 \ll 1$  are added for numerical stability along with an effort to accommodate the perception of the human visual system.

The component  $S_1$  in (6) measures the similarity of the mean values,  $\bar{\mathbf{x}}$  and  $\bar{\mathbf{y}}$  of, respectively,  $\mathbf{x}$  and  $\mathbf{y}$ . If  $\bar{\mathbf{x}} = \bar{\mathbf{y}}$ , then  $S_1(\mathbf{x}, \mathbf{y}) = 1$ , its maximum possible value. Its functional form was originally chosen in an effort to accommodate Weber's law of perception [14]. The component  $S_2$  in (6) follows the idea of divisive normalization [15]. Note that  $-1 \leq S(\mathbf{x}, \mathbf{y}) \leq 1$ , and  $S(\mathbf{x}, \mathbf{y}) = 1$  if and only if  $\mathbf{x} = \mathbf{y}$ . A negative value of  $S(\mathbf{x}, \mathbf{y})$  implies that  $\mathbf{x}$  and  $\mathbf{y}$  are negatively correlated.

## 2.1 Optimal SSIM-Based Affine Approximation

We now consider the approximation of an image range block  $u(R_i)$  by a domain block  $u(D_j)$  as written in (1) in terms of the structural similarity measure. The SSIM measure associated with the approximation in (1) will be defined as

$$S_{ij} = \max_{\alpha, \beta \in \Pi} S(u(R_i), \alpha u(D_j) + \beta), \quad i \neq j. \quad (8)$$

The optimal parameters and associated SSIM measures are given below, but only for the zero stability parameter case, i.e.,  $\epsilon_1 = \epsilon_2 = 0$ . Because of space restrictions, we omit the algebraic details and simply state the results. In what follows, we once again denote  $\mathbf{x} = u(R_i)$ ,  $\mathbf{y} = u(D_j)$  and  $N = n^2$ .

**Case 1: Purely translational.** There is no optimization in this case:  $\alpha_{ij} = 1$ ,  $\beta_{ij} = 0$  and the SSIM measure is simply

$$S_{ij}^{(\text{Case 1})} = S(\mathbf{x}, \mathbf{y}). \quad (9)$$

**Case 2: Translational + greyscale shift.** Here,  $\alpha_{ij} = 1$  and we optimize over  $\beta$ .

$$\beta_{ij} = \bar{\mathbf{x}} - \bar{\mathbf{y}} , \quad S_{ij}^{(\text{Case 2})} = S_2(\mathbf{x}, \mathbf{y}) = \frac{2s_{\mathbf{x}\mathbf{y}}}{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2} . \quad (10)$$

Note that the SSIM-optimal  $\beta$  parameter is identical to its  $L^2$  counterpart.

**Case 3: Affine greyscale transformation.** We optimize over  $\alpha$  and  $\beta$ .

$$\alpha_{ij} = \text{sign}(s_{\mathbf{x}\mathbf{y}}) \frac{s_{\mathbf{x}}}{s_{\mathbf{y}}} , \quad \beta_{ij} = \bar{\mathbf{x}} - \alpha_{ij}\bar{\mathbf{y}} , \quad S_{ij}^{(\text{Case 3})} = \frac{|s_{\mathbf{x}\mathbf{y}}|}{s_{\mathbf{x}}s_{\mathbf{y}}} , \quad (11)$$

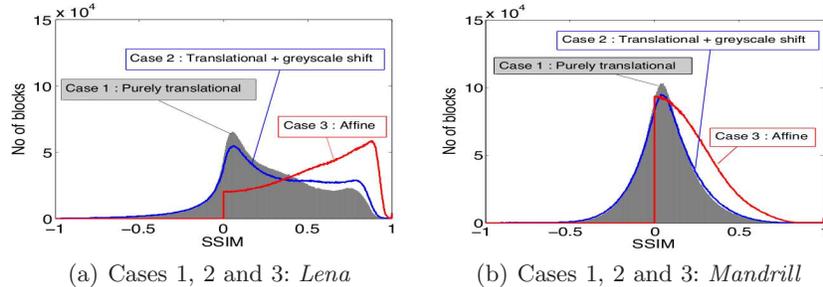
where  $\text{sign}(t) = 1$  if  $t > 0$ , 0 if  $t = 0$ , and  $-1$  if  $t < 0$ . In this case, the SSIM measure  $S_{ij}$  is the magnitude of the correlation between  $\mathbf{x}$  and  $\mathbf{y}$ .

*Note 2.* In Cases 2 and 3, the means of the range block and optimally transformed range block are equal, i.e.,  $\bar{\mathbf{x}} = \alpha\bar{\mathbf{y}} + \beta$ , as was the case for  $L^2$ -fitting.

Since more parameters are involved as we move from Case 1 to Case 3, the associated SSIM measures behave as follows,

$$S_{ij}^{(\text{Case 1})} \leq S_{ij}^{(\text{Case 2})} \leq S_{ij}^{(\text{Case 3})} . \quad (12)$$

In Fig. 2 are shown the Case 1-3 SSIM measure distributions over the interval  $[-1, 1]$  of the *Lena* and *Mandrill* images, once again using  $8 \times 8$ -pixel blocks.



**Fig. 2.** Case 1-3 SSIM measure distributions for normalized *Lena* and *Mandrill* images over  $[-1, 1]$ . In all cases, nonoverlapping  $8 \times 8$ -pixel blocks  $R_i$  and  $D_j$  were used.

Before commenting on these plots, we briefly discuss the issue of the stability parameters,  $\epsilon_1$  and  $\epsilon_2$  in (6). As proposed in [14], in all computations reported below the stability parameters employed were  $\epsilon_1 = 0.01^2$  and  $\epsilon_2 = 0.03^2$ . In the case  $\epsilon_1 = \epsilon_2 = 0$ , the Case 1 SSIM measure distributions of the *Lena* and *Mandrill* images are almost identical. The slightly nonzero values of the stability parameters will increase the SSIM values associated with domain-range pairs with low variance. Since the *Lena* image contains a higher proportion of such blocks, there is a slight increase of the distribution for  $S > 0$ .

The difference between the two distributions is more pronounced for Case 2. For the *Lena* image, the better domain-range block approximations yielded by the greyscale shift causes its SSIM measure distribution to increase over the region  $S \subset [0.5, 0.8]$ .

But the situation is most interesting for Case 3, i.e., affine greyscale approximation. For both images, there are no negative SSIM values. This follows from the positivity of  $S_{ij}$  in (11) which is made possible by the inclusion of the  $\alpha$  scaling factor. When the domain and range blocks are correlated, as opposed to anticorrelated, i.e.,  $s_{\mathbf{xy}} > 0$  then the optimal  $\alpha$  coefficient is positive, implying that  $S$  will be positive. When  $\alpha < 0$ , the domain and range blocks are anticorrelated – multiplying the domain block by a negative  $\alpha$  value will “undo” this anticorrelation to produce a roughly correlated block.

The SSIM distribution for the *Lena* image has a much stronger component near  $S = 1$ , indicating that many more blocks are well approximated in terms of the SSIM measure. Conversely, the SSIM measure for the *Mandrill* image is quite strongly peaked at  $S = 0$ . In summary, the SSIM measure corroborates the fact that the *Lena* image is more self-similar than the *Mandrill* image. That being said, despite the dramatic peaking of the RMS  $\Delta$ -error distribution of the *Lena* image at zero error – primarily due to a high proportion of low-variance blocks – its SSIM measure distribution does not demonstrate such peaking near  $S = 1$ . This will be explained in the following section.

## 2.2 Relation between Optimal $L^2$ - and SSIM-Based Greyscale Coefficients

At this point it is instructive to compare the affine greyscale transformations of the  $L^2$ - and SSIM-based approximations. Obviously, for Case 1, no comparison is necessary since no greyscale transformations are employed. For Case 2, the greyscale shift  $\beta = \bar{u}(R_i) - \bar{u}(D_j)$  is the same in both approximations. For Case 3, it is sufficient to compare the  $\alpha$  greyscale coefficients. Recall that for a given domain block  $\mathbf{x} = u(D_j)$  and range block  $\mathbf{y} = u(R_i)$ ,

$$\alpha_{L^2} = \frac{s_{\mathbf{xy}}}{s_{\mathbf{y}}^2} , \quad \alpha_{SSIM} = \text{sign}(s_{\mathbf{xy}}) \frac{s_{\mathbf{x}}}{s_{\mathbf{y}}} . \quad (13)$$

It follows that

$$\frac{\alpha_{SSIM}}{\alpha_{L^2}} = \frac{s_{\mathbf{x}} s_{\mathbf{y}}}{|s_{\mathbf{xy}}|} \geq 1 , \quad (14)$$

where the final inequality follows from (11).

This result implies that the SSIM-based affine approximation  $\alpha u(D_j) + \beta$  will have a higher variance than its  $L^2$ -based counterpart. Such a “contrast enhancement” was also derived for SSIM-based approximations using orthogonal bases [16].

Finally, note that the coefficients  $\alpha_{SSIM}$  and  $\alpha_{L^2}$  always have the same sign. Numerically, we find that their values generally do not differ greatly: A histogram plot of their ratios is strongly peaked at 1.

### 2.3 SSIM, Normalized Metrics and Image Self-Similarity vs. Image “Approximability”

The fact that  $S(\mathbf{x}, \mathbf{y}) = 1$  if and only  $\mathbf{x} = \mathbf{y}$  suggests that the function

$$T(\mathbf{x}, \mathbf{y}) = 1 - S(\mathbf{x}, \mathbf{y}) \quad , \quad \mathbf{x}, \mathbf{y} \in \mathbf{R}_+^N \quad , \quad (15)$$

could be considered a measure of the distance between  $\mathbf{x}$  and  $\mathbf{y}$ , since  $\mathbf{x} = \mathbf{y}$  implies that  $T(\mathbf{x}, \mathbf{y}) = 0$ . We now show that for Case 2 and Case 3, the function  $T(\mathbf{x}, \mathbf{y})$  may be expressed in terms of the  $L^2$  distance  $\|\mathbf{x} - \mathbf{y}\|$ .

First recall that for both Case 2 and Case 3 and for  $L^2$ - and SSIM-based approximations of a range block  $\mathbf{x} = u(R_i)$ , the mean of the best affine approximation  $\mathbf{y} = \alpha u(D_j) + \beta$  is equal to the mean of  $\mathbf{x}$ . As such, we consider the function  $T(\mathbf{x}, \mathbf{y})$  in the special case that  $\bar{\mathbf{x}} = \bar{\mathbf{y}}$ . This implies that  $S(\mathbf{x}, \mathbf{y}) = S_2(\mathbf{x}, \mathbf{y})$ , the second component of SSIM, and that

$$T(\mathbf{x}, \mathbf{y}) = 1 - \frac{2s_{\mathbf{x}\mathbf{y}} + \epsilon_2}{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 + \epsilon_2} = \frac{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 - 2s_{\mathbf{x}\mathbf{y}}}{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 + \epsilon_2} = \frac{1}{N-1} \frac{\|\mathbf{x} - \mathbf{y}\|^2}{s_{\mathbf{x}}^2 + s_{\mathbf{y}}^2 + \epsilon_2} \quad . \quad (16)$$

In other words, the function  $T(\mathbf{x}, \mathbf{y})$  is an inverse variance-weighted squared  $L^2$ -distance between  $\mathbf{x}$  and its optimal affine approximation  $\mathbf{y}$ . In fact, one can show (see [17]) that  $\sqrt{T(\mathbf{x}, \mathbf{y})}$  is indeed a metric when the means are matched.

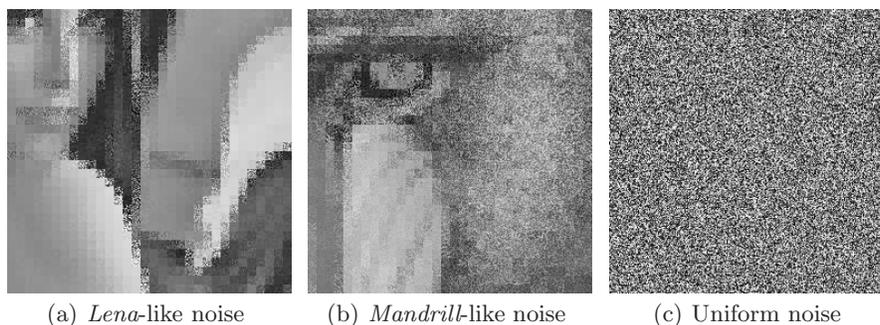
As mentioned earlier, lower-variance blocks are more easily approximated in the  $L^2$  sense than higher-variance blocks. Consequently, the Case 3  $\Delta$ -error distributions of images with a higher proportion of “flatter,” i.e., low variance, blocks will exhibit a greater degree of peaking near zero, particularly for Case 3. The structural similarity index compensates for this “flatness bias.” The question is whether this greater peaking should actually be interpreted as self-similarity. This is addressed in the next section.

## 3 Self-Similarity of Natural Images vs. Pure Noise Images

The presence of noise in an image will decrease the ability of its subblocks to be approximated by other subblocks. In [11] it was observed that as (independent, Gaussian) noise of increasing variance  $\sigma^2$  is added to an image, any near-zero peaking of its  $\Delta$ -error distribution becomes diminished. Moreover, a  $\chi$ -squared error distribution associated with the noise which peaks at  $\sigma$  eventually dominates the  $\Delta$ -error distribution. This peaking at  $\sigma$  is actually the basis of the block-variance method of estimating additive noise.

Naturally, the SSIM measure distributions will also be affected by the presence of noise. But instead of simply adding noise to natural images, we wish to study pure noise images. Synthesizing such kinds of images allows us to compare the  $\Delta$ -error distributions of natural images with a benchmark image that possesses no self-similarity. Indeed, for independent pure noise images there is no self-similarity between two blocks in the sense that the expectation of the covariance between them is zero. The only parameters affecting the self-similarity (in

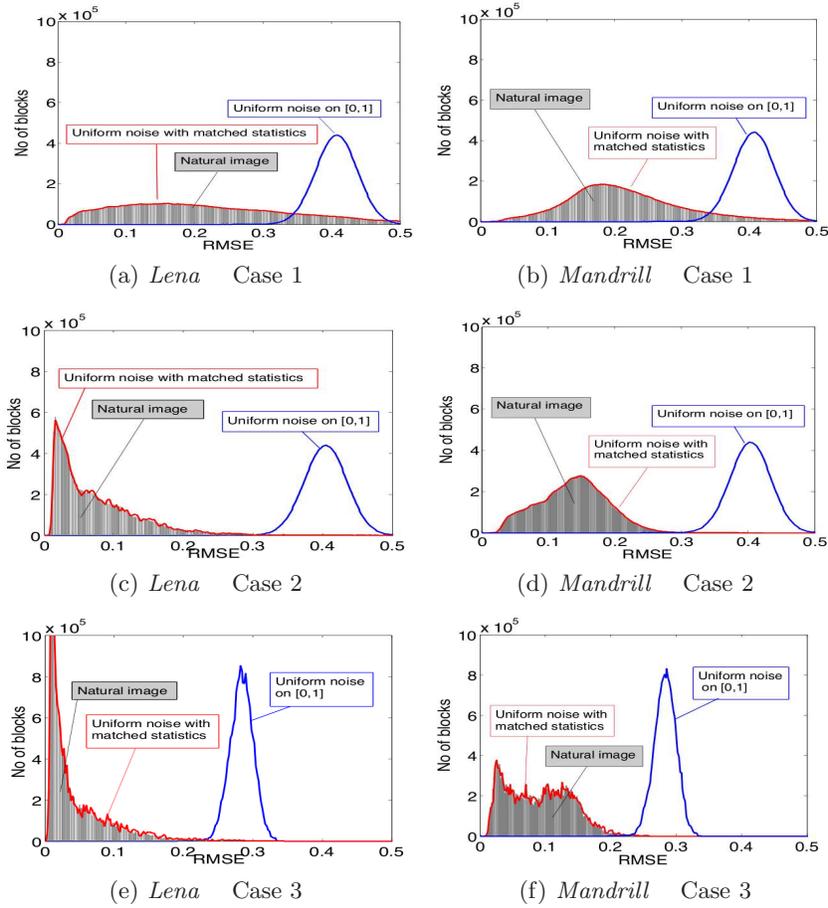
RMSE or in SSIM-sense) are the local mean and the local variance of the image. This leads to the following idea: Generate an image from a uniform distribution with the local mean and variance matched to the statistics of a natural image. In our experiments, we chose an uniform distribution, but the histograms would have been similar for Gaussian or Poisson probability distribution. In Fig. 3 are shown two examples of pure noise images for which the local statistics are matched to a natural image. Also shown is a pure noise image following i.i.d. uniform distribution on  $[0, 1]$ . Disjoint blocks were used to compute the local statistics to be consistent through the paper, but it is by no means necessary to generate pure noise images block by block.



**Fig. 3.** Images made of uniform noise with statistics matching the local mean and variance of natural images: (a) *Lena* (b) *Mandrill* (c) Noise image with each pixel value taken from an uniform distribution on  $[0, 1]$ .

In Fig. 4, we compare the RMS  $\Delta$ -error distribution of natural images with the RMS  $\Delta$ -error distribution of a pure noise image with the local statistics matched and of an pure noise image following an uniform distribution on  $[0, 1]$ . We observe that there is no more self-similarity for natural images than for pure noise images with matched statistics. Notice that *all* possible blocks were compared, whereas in non-local image processing only a limited number of (best) blocks are usually needed. So even if the best matches are generally more self-similar, on average, natural images are not more self-similar than pure noise images with matched statistics. *We conclude that low variance is the principal factor for self-similarity according to RMSE.*

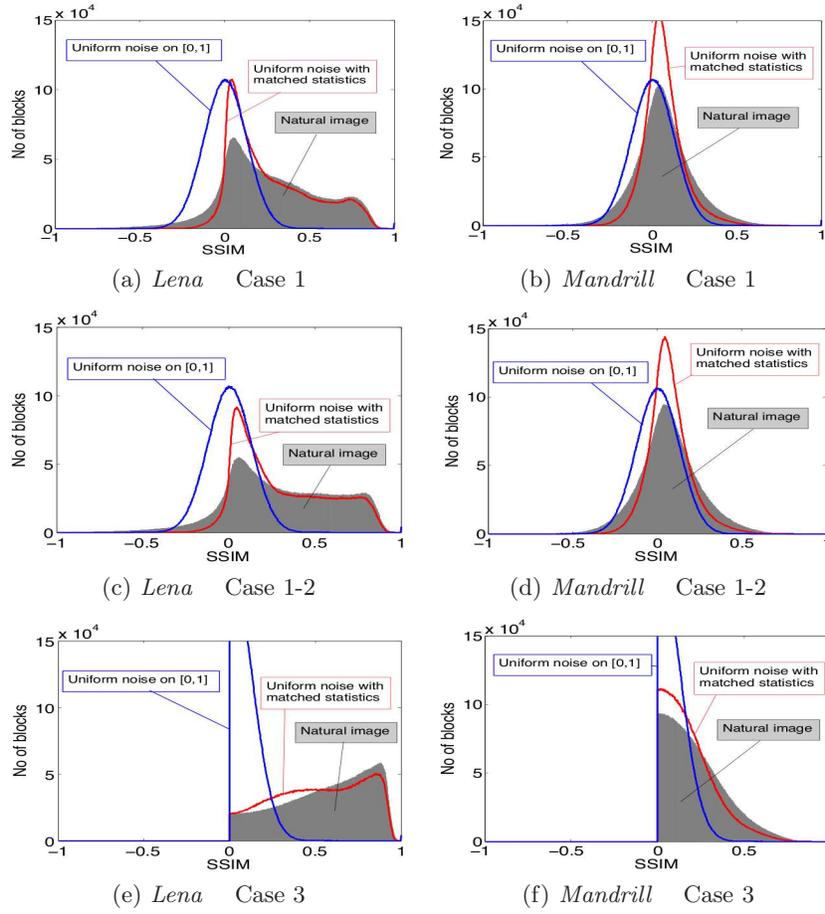
In order to correct this low variance bias, the same experiment was performed with the SSIM index for Case 1-3. The results are shown in Fig. 5. Now, we can see a difference between the SSIM measure distributions of natural images and pure noise images. We quantify the self-similarity of images by computing the center of gravity (the mean of the distribution) of the SSIM measure distributions. The results are shown in Table 1. Again, the local variance has a major influence on the self-similarity of images, but now we can see, as hoped, that natural images are more self-similar than pure noise images in the SSIM-sense.



**Fig. 4.** Comparison of RMS  $\Delta$ -error distribution of *Lena* and *Mandrill* for Case 1-3 (grey histogram) with the RMS  $\Delta$ -error distribution of pure noise images for which the local mean and local variance are matched (red) and with the RMS  $\Delta$ -error distribution of a i.i.d. uniform pure noise image on  $[0, 1]$  (blue).

To determine theoretically the distribution of the structural similarity between two blocks generated by a known probability distribution remains an open question. The difficulty here is the fact that rational functions are involved in the definition of the SSIM measure.

**Acknowledgements.** We gratefully acknowledge the generous support of this research by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the forms of Discovery Grants (ERV, ZW), a Strategic Grant (ZW), a collaborative research and development (CRD) grant (ERV, ZW) and a Post-graduate Scholarship (DB). ZW would also like to acknowledge partial support



**Fig. 5.** Comparison of SSIM measure distribution of *Lena* and *Mandrill* for Case 1-3 (grey histogram) with the SSIM measure distribution of pure noise images for which the local mean and local variance are matched (red) and the SSIM measure distribution of a i.i.d. uniform pure noise image on  $[0, 1]$  (blue).

**Table 1.** Mean of the SSIM measure distributions of natural images (NI), pure noise images with matched statistics (MN) and uniform pure noise image (UN) for Case 1-3 *Lena* and *Mandrill*.

	Case 1	Case 1	Case 2	Case 2	Case 3	Case 3
	<i>Lena</i>	<i>Mandrill</i>	<i>Lena</i>	<i>Mandrill</i>	<i>Lena</i>	<i>Mandrill</i>
NI	0.2719	0.0682	0.3091	0.0731	0.5578	0.2246
MN	0.2698	0.0684	0.3074	0.0735	0.5206	0.1896
UN	0.0057	0.0057	0.0057	0.0057	0.1003	0.1004

by the Province of Ontario Ministry of Research and Innovation in the form of an Early Researcher Award.

## References

1. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. *Multiscale Modelling and Simulation* **4** (2005) 490–530
2. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Processing* **16** (2007) 2080–2095
3. Zhang, D., Wang, Z.: Image information restoration based on long-range correlation. *IEEE Trans. Circuits and Systems for Video Tech.* **12** (2002) 331–341
4. Etemoglu, C., Cuperman, V.: Structured vector quantization using linear transforms. *IEEE Trans. Signal Processing* **51** (2003) 1625–1631
5. Ebrahimi, M., Vrscay, E.R.: Solving the inverse problem of image zooming using self-examples. In Kamel, M., Campilho, A., eds.: *Proc. Int. Conf. on Image Analysis and Recognition*. Volume 4633 of LNCS. Springer, Heidelberg (2007) 117–130
6. Elad, M., Datsenko, D.: Example-based regularization deployed to super-resolution reconstruction of a single image. *The Computer Journal* **50** (2007) 1–16
7. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. *IEEE Computer Graphics and Applications* **22** (2002) 56–65
8. Barnsley, M.F.: *Fractals Everywhere*. Academic Press, New York (1988)
9. Lu, N.: *Fractal Imaging*. Academic Press, New York (1997)
10. Ghazel, M., Freeman, G., Vrscay, E.R.: Fractal image denoising. *IEEE Trans. Image Processing* **12** (2003) 1560–1578
11. Alexander, S.K., Vrscay, E.R., Tsurumi, S.: A simple model for the affine self-similarity of images. In Kamel, M., Campilho, A., eds.: *Proc. Int. Conf. on Image Analysis and Recognition*. Volume 5112 of LNCS., Springer, Heidelberg (2008) 192–203
12. La Torre, D., Vrscay, E.R., Ebrahimi, M., Barnsley, M.F.: Measure-valued images, associated fractal transforms and the self-similarity of images. *SIAM J. Imaging Sci.* **2** (2009) 470–507
13. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine* **26**(1) (2009) 98–117
14. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Processing* **13**(4) (2004) 600–612
15. Wainwright, M.J., Schwartz, O., Simoncelli, E.P.: Natural image statistics and divisive normalization: Modeling nonlinearity and adaptation in cortical neurons. In Rao, R., et al., eds.: *Probabilistic Models of the Brain: Perception and Neural Function*, MIT Press, Cambridge (2002) 203–222
16. Brunet, D., Vrscay, E.R., Wang, Z.: Structural similarity-based approximation of signals and images using orthogonal bases. In Kamel, M., Campilho, A., eds.: *Proc. Int. Conf. on Image Analysis and Recognition*. Volume 6111 of LNCS., Springer, Heidelberg (2010) 11–22
17. Brunet, D., Vrscay, E.R., Wang, Z.: A class of image metrics based on the structural similarity quality index. In: *8th International Conference on Image Analysis and Recognition (ICIAR' 11)*, Burnaby, BC, Canada (June 2011)