Hyperspectral Images as Function-Valued Mappings, Their Self-similarity and a Class of Fractal Transforms

Edward R. Vrscay[1], D. Otero[1], and D. La Torre[2]

[1] Department of Applied Mathematics, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
[2] Department of Economics, Business and Statistics, University of Milan, Milan, Italy
{ervrscay,dotero}@uwaterloo.ca, davide.latorre@unimi.it

Abstract. A formulation of hyperspectral images as function-valued mappings is introduced, along with a set of simple models of affine self-similarity for digital hyperspectral images. As in the case of greyscale images, these models examine how well vector-valued image subblocks are approximated by other subblocks, as measured by the distribution of approximation errors. This set of models includes both same-scale and cross-scale modes of approximation, the latter of which provides the basis of a method of fractal transforms over hyperspectral images.

1 Introduction

In this paper, we introduce a simple model of affine self-similarity for hyperspectral images [6]. Our model may be viewed as an extension of earlier work [1,4,13] involving greyscale images. Here, as in [15], we view hyperspectral images mathematically as function-valued mappings: At each location or pixel x, the hyperspectral image $u(x)$ is a *function* which is supported on a domain which is appropriate to the application. For example, in the case of remote sensing, in which $u(x)$ is defined by a set of reflectance values measured at various wavelengths, this domain is a subset of the non-negative real line, $\mathbf{R}_+$. In the case of diffusion magnetic resonance images, where $u(x)$ may be defined by a set of probabilities of diffusion from x in various directions, the domain is S^2, the unit sphere in $\mathbf{R}^3$.

Our studies of image self-similarity have been motivated by the effectiveness of a number of *nonlocal image processing* methods, including nonlocal-means denoising [5], compression [10], restoration [18], superresolution [8,9] and fractal image coding [11,14]. Such effectiveness may be attributed to the fact that pixel blocks of a natural image can, in some way, be well approximated by a number of other blocks of the same image. A function-valued formalism allows this notion of *image self-similarity* to be extended to hyperspectral images and shows that they may be treated by various nonlocal methods.

Some very interesting questions still remain unanswered, e.g., whether it is better to perform such methods (i) along separate channels, (ii) separately at

M. Kamel and A. Campilho (Eds.): ICIAR 2013, LNCS 7950, pp. 225–234, 2013.

fixed pixels, (iii) with particular subgroups of channels or (iv) a combination of (i)-(iii). We continue to pursue these most interesting avenues of inquiry.

Finally, we mention that the work presented here complements the well-established notion that hyperspectral images generally demonstrate a high degree of correlation which can be exploited to accomplish various tasks [16], e.g., compression, compressed sensing, denoising.

2 Mathematical Formulation of Hyperspectral Images

It is important to begin with a brief account of the mathematical foundations of our study. The basic ingredients of our formalism are:

The base space X: The compact support of the hyperspectral images, with metric d_X. For convenience, $X = [0,1]^n$, where $n = 2$ or 3. In practical considerations, i.e., digitized hyperspectral images, X will be replaced by a discrete array, the *pixel space*, as described in Section 3.

The range or spectral space $L^2(\mathbf{R}_s)$: The space of square-integrable functions supported on a compact set $\mathbf{R}_s \subset \mathbf{R}_+$, where $\mathbf{R}_+ = \{y \in \mathbf{R} \mid y \geq 0\}$. $L^2(\mathbf{R_s})$ is a Hilbert space with inner product defined as follows,

$$\langle f, g \rangle = \int_{\mathbf{R}_s} f(t)\, g(t)\, dt, \quad \forall f, g \in L^2(\mathbf{R}_s). \tag{1}$$

This inner product defines a norm on $L^2(\mathbf{R}_s)$, to be denoted as $\|\cdot\|_{L^2(\mathbf{R}_s)}$.

Now let Y denote the set of all function-valued mappings from the base space X to $L^2(\mathbf{R}_s)$. Given a hyperspectral image $u \in Y$, its value $u(x)$ at a particular location $x \in X$ will be a function – more precisely, an element of the space $L^2(\mathbf{R}_s)$. Following the same prescription as in [15], the norm $\|\cdot\|_{L^2(\mathbf{R}_s)}$ arising from Eq. (1) may be used to define a metric d_Y over the space Y: The distance between two hyperspectral images $u, v \in Y$ will then be defined as

$$d_Y(u,v) = \left[\int_X \|u(x) - v(x)\|^2_{L^2(\mathbf{R}_s)}\, dx \right]^{1/2}. \tag{2}$$

Furthermore, the metric space (Y, d_Y) of hyperspectral images is complete.

3 A Simple Model for the Affine Self-similarity of a Digital Hyperspectral Image

The remainder of this paper will be concerned with a discrete version of the formulation in the previous section: digital hyperspectral images supported on an $N_1 \times N_2$-pixel array, M channels per pixel. Formally, a digital hyperspectral image I may be represented by a vector-valued image function, $u : X \to \mathbf{R}^M_+$, where $X = \{1, 2, \cdots, N_1\} \times \{1, 2, \cdots, N_2\}$ is the *base* or *pixel space* and R^M_+, the nonnegative orthant of $\mathbf{R}^M$, is the *spectral space*. At a pixel location $(i_1, i_2) \in$

X, the hyperspectral image function $u(i_1, i_2)$ is a non-negative M-vector with components $u_k(i_1, i_2)$, $1 \le k \le M$.

As in [1] let $\mathcal{R}^{(n)}$ denote a set of nonoverlapping $n \times n$-pixel subblocks R_i, such that $X = \cup_i R_i$, i.e., $\mathcal{R}^{(n)}$ forms a partition of X. We let $u(R_i)$ denote the portion of the hyperspectral image function u that is supported on subblock $R_i \subset X$. In this setting, $u(R_i)$ is an $n \times n \times M$ cube of nonnegative real numbers.

We now wish to examine how well a hyperspectral image subblock $u(R_i)$ is approximated by other image subblocks $u(R_j)$, $j \ne i$. For convenience, we denote an image subblock $u(R_i)$ being approximated as a *range block* and a subblock $u(R_j)$, $j \ne i$, approximating it as a *domain block*. In order to clarify the roles of these blocks, we shall denote the domain blocks as $u(D_j)$. This notation also allows for the possibility of considering the eight possible mappings (four rotations and four inversions) of the pixels in D_j to the pixels in R_i, in which case an additional index will be required, i.e., $u(D_j^{(k)})$, $1 \le k \le 8$. In this paper, for simplicity, we consider only the identity mapping, so that $D_j = R_j$.

At this point, it is tempting to consider the approximations of subblocks separately channel by channel, i.e., to treat the hyperspectral image u as a set of M greyscale images and evaluate the approximations over each greyscale image, along the lines done in [1]. This is certainly possible, but it defeats the purpose of examining any similarity/correlation across channels. As such, we shall define the distance between two image subblocks $u(R_i), u(D_j) \in \mathcal{R}^{(n)}$ to be

$$\Delta_{ij} = \frac{1}{n\sqrt{M}} \left[\sum_{i_1=I_1}^{I_1+n-1} \sum_{i_2=I_2}^{I_2+n-1} \sum_{k=1}^{M} [u_k(i_1, i_2) - u_k(i_1 + J_1, i_2 + J_2)]^2 \right]^{1/2}, \quad (3)$$

where the summation is understood to be performed over corresponding pixels of R_i and $D_j = R_j$. (Here it is assumed that subblock R_j is simply a translation of R_i by (J_1, J_2).) The RHS of Eq. (3) is, up to a factor, a discretized version of the distance between two image subblocks in Eq. (2). The factor $1/\sqrt{M}$ produces a root-mean-squared (RMS)/L^2 distance in the spectral space $\mathbf{R}_+^M$ at each pixel. The additional normalization factor $1/n$ produces a per-pixel error. This is helpful in the comparison of distributions for different range block sizes.

Case 1 Approximation: The distance Δ_{ij} in Eq. (3) is the error associated with the approximation

$$u(R_i) \approx u(D_j), \qquad \text{(Case 1)}. \quad (4)$$

As in [1], we consider this to be the simplest approximation scheme.

Case 2 Approximations with Spectral Shift: In [1], the Case 1 approximation for greyscale images was greatly improved by adding a constant greyscale value β to the approximating block $u(D_j)$. The optimal value of β is easily found to be the difference in the mean values of the greyscale blocks $u(R_i)$ and $u(D_j)$.

As shown in the next section, numerical experiments indicate that the approximation,

$$u(R_i) \approx u(D_j) + \beta, \qquad \text{(Case 2(a))}, \quad (5)$$

where $\beta \in \mathbf{R}$ is a constant, does not, in general, improve the Case 1 approximation in Eq. (4) significantly. It is too much to expect a constant shift β to accomodate the many channels of a hyperspectral image. A significantly greater improvement is achieved if we allow a separate shift for each channel. The resulting approximation has the form

$$u(R_i) \approx u(D_j) + \underline{\beta}\,, \qquad \text{(Case 2(b))}, \tag{6}$$

where $\underline{\beta} \in \mathbf{R}_+^M$ is an M-vector. Eq. (6) is to be interpreted as follows: At each pixel (i_1, i_2) of the range block R_i, we have the vector-valued approximation,

$$u_k(i_1, i_2) \approx u_k(j_1, j_2) + \beta_k, \quad 1 \le k \le M, \tag{7}$$

where (j_1, j_2) denotes the corresponding pixel location in the domain block D_j.

The values of β_k that minimize the approximation error associated with Eq. (6), are easily found to be

$$\beta_k = \overline{u_k(R_i)} - \overline{u_k(D_j)}, \quad 1 \le k \le M. \tag{8}$$

Here, $\overline{u_k(R_i)}$ denotes the mean value of the kth channel of the hyperspectral image supported on subblock R_i. In other words, the channels are treated separately in this approximation.

Case 3 Approximation with Affine Scaling + Spectral Shift: In [1], the Case 2 approximation for greyscale images was greatly improved in many cases by scaling the approximating block, i.e., $u(R_i) \approx \alpha u(R_j) + \beta$. With reference to Eq. (6), one could introduce a separate α_k scaling coefficient to accompany the constant β_k for each channel, essentially treating a hyperspectral image as M separate greyscale images. A more "hyperspectral" approach that captures the similarity between channels is to employ only one scaling coefficient α for all channels, i.e.,

$$u(R_i) \approx \alpha\, u(D_j) + \underline{\beta}, \qquad \text{(Case 3)}. \tag{9}$$

The optimal values for α and the β_k are derived in the Appendix.

In fact, we have found experimentally that when separate scaling coefficients α_k are employed for each channel, a histogram plot of the variances of the optimal $\underline{\alpha}$-vectors is strongly peaked at zero, implying that for many blocks the α_k values are close to each other. Consequently, employing a single α coefficient does not increase the approximation error significantly for many blocks.

For $m \in \{1, 2(a), 2(b), 3\}$, we let $\Delta_{ij}^{(Case\ m)}$ denote the error in approximating the range block $u(R_i)$ with the domain block $u(D_j)$ using the Case m approximation schemes defined above. Since more optimization parameters involved as we move from Case 1 to Case 3, it follows that

$$0 \le \Delta_{ij}^{(Case\ 3)} \le \Delta_{ij}^{(Case\ 2(b))} \le \Delta_{ij}^{(Case\ 2(a))} \le \Delta_{ij}^{(Case\ 1)}. \tag{10}$$

As m increases, the distributions $\Delta_{ij}^{(Case\ m)}$ should become more concentrated toward zero error. This effect was observed for greyscale images in [1].

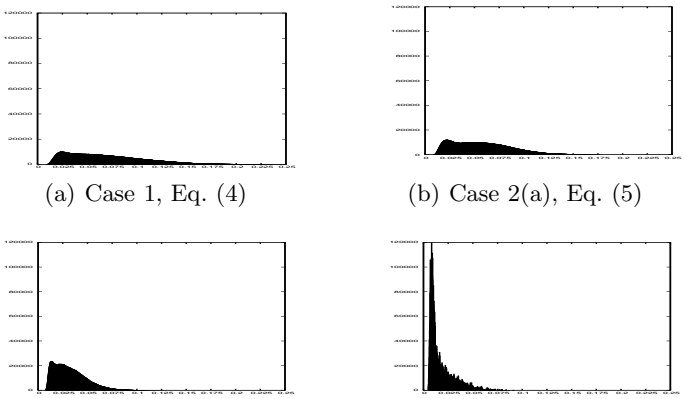
4 Results of Some Numerical Experiments

Here we report briefly on an examination of the self-similarity properties of two hyperspectral data sets. (Similar results have been obtained for a number of other images.)

1. A 33-channel hyperspectral image, "Scene 2," downloaded from the webpage of D.H. Foster, University of Manchester [12], to be referred to as the "HS fern image".
2. AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) image, "Yellowstone calibrated scene 0," a 224-channel image, available from the Jet Propulsion Laboratory site [2], to be referred to as the "Yellowstone AVIRIS" image.

In Figure 1 are presented histogram plots of the $\Delta_{ij}^{(Case\ m)}$ error distributions, $m = 1, 2(a), 2(b)$ and 3, for the 33-channel HS fern image. The four plots employ the same scaling of x- and y-axes so that comparisons can easily be made. Unlike the situation for natural greyscale images, as seen in [1], the Case 1 error distribution is quite flat and unimpressive. We also observe that the Case 2(a) error distribution is quite similar to that of Case 1, indicating that the use of a single constant β for each block does not improve the approximation significantly. The Case 2(b) error distribution demonstrates a much greater peaking toward zero error, showing that the use of a vector shift $\underline{\beta}$ improves the approximations significantly. A dramatic increase in near-zero peaking is demonstrated by the Case 3 error distribution, indicating a much improved approximation.

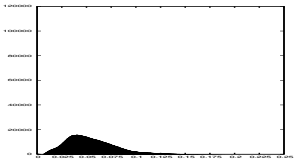
In Figure 2 are presented histogram approximations of the $\Delta_{ij}^{(Case\ m)}$ error distributions for the 224-channel AVIRIS image.



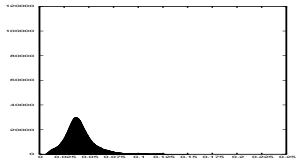
(a) Case 1, Eq. (4) (b) Case 2(a), Eq. (5)

(c) Case 2(b), Eq. (6) (d) Case 3, Eq. (9)

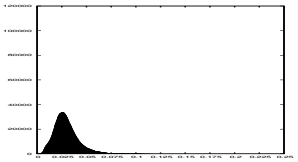
Fig. 1. Per-pixel error distributions $\Delta_{ij}^{(Case\ m)}$ for 33-channel hyperspectral fern image. In all cases, 8×8-pixel blocks R_i and D_j were used.



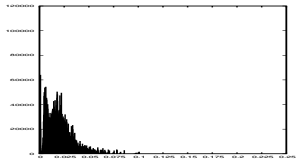
(a) Case 1, Eq. (4)



(b) Case 2(a), Eq. (5)



(c) Case 2(b), Eq. (6)

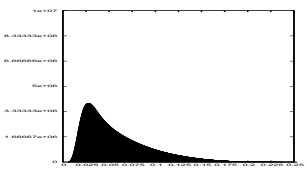


(d) Case 3, Eq. (9)

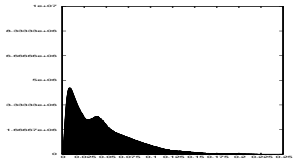
Fig. 2. Per-pixel error distributions $\Delta_{ij}^{(Case\ m)}$ for the 224-channel AVIRIS image. In all cases, 8×8-pixel blocks R_i and D_j were used.

Single-Pixel Self-similarity of Spectral Functions. The self-similarity study of greyscale images reported in [1] was limited to a particular partition size, namely, 8×8-pixel blocks. Experimentally, it is observed, as expected, that the Δ_{ij} error distributions for a given block size demonstrate increased peaking toward zero error as we move from Case 1 to Case 3. And for a given Case m, the $\Delta_{ij}^{(Case\ m)}$ error distributions shift toward zero error as the partition size n is decreased since it is generally easier to fit a lower number of data points.

Hyperspherical images, however, possess an additional degree of complexity since each pixel no longer supports a single greyscale value but rather an M-vector of spectral values. As such, it is feasible to look for "single-pixel" similarity. In Figure 3 are presented the Case 1 Δ_{ij} error distributions for the HS fern and Yellowstone AVIRIS images when the range and domain blocks R_j are single pixels. A significant amount of similarity is already demonstrated for this Case. Much of this is due to the high correlation of spectral functions supported on neighbouring pixels which is often graphically demonstrated by plotting these

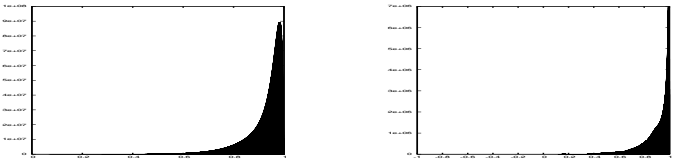


(a) Hyperspherical fern



(b) Yellowstone AVIRIS

Fig. 3. Case 1 per-pixel error distributions $\Delta_{ij}^{(Case\ 1)}$ for spectral functions supported on **single-pixel** blocks R_i.



(a) Hyperspherical fern (b) Yellowstone AVIRIS

Fig. 4. Pairwise correlations between **single-pixel** spectral functions

functions next to each other. Such high correlation will manifest itself in a principal components analysis (PCA) of hyperspectral images.

Correlation of "Single-Pixel" Spectral Functions. It is well known that the RMSE/L^2 distance is not a good measure of signal fidelity and image quality [17]. This has motivated a study of self-similarity [4] in terms of the "structural similarity" image quality measure. Here, as a complement to the RMS/L^2 distance-based single-pixel self-similarity results of Figure 3, we present, in Figure 4, plots of the correlations between single-pixel spectral functions of the hyperspherical fern and Yellowstone AVIRIS images. The correlation $C(\mathbf{x}, \mathbf{y})$ between two M-vectors $\mathbf{x}$ and $\mathbf{y}$ was computed with the usual formulas, i.e.,

$$C(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}{\sigma_{\mathbf{xy}}}, \qquad \sigma_{\mathbf{xy}} = \frac{1}{M-1}\sum_{k=1}^{M}(x_i - \bar{\mathbf{x}})(y_i - \bar{\mathbf{y}}). \tag{11}$$

The dramatic correlation demonstrated in these plots strongly suggests that single-pixel spectral functions are quite suitable for nonlocal methods of image processing. This may not be so clearly demonstrated by PCA or related methods.

Finally, we mention that our use of *nonoverlapping* range/domain blocks R_i, as opposed to the *sliding block* method usually employed in methods such as nonlocal means denoising, was motivated by a desire to reduce computational effort. The error distributions presented above may be viewed as a kind of "lower bound" in the characterization of self-similarity: Given the high correlation between spectral functions associated with neighbouring pixels, it is expected that the Δ-error distributions produced by sliding blocks will exhibit far more peaking than the distributions shown above. Furthermore, the single-pixel case, which clearly employs sliding blocks, yields "upper bounds" for self-similarity, since single spectral functions should be, in general, easiest to approximate.

Clearly, there is also the possibility of approximating spectral functions within *subsets* of channels. Once again, our results, based upon the entire M-channel spectrum, may be viewed as providing "lower bounds" to self-similarity.

5 A Class of Fractal Transforms on Hyperspectral Images

Fractal transforms of greyscale and color images [11,14] are based on an additional "cross-scale" self-similarity property, denoted as "Case 4" in [1].

Very briefly, a greyscale (i.e., $M = 1$-channel) image subblock $u(R_i)$ supported on an $n \times n$-pixel range block $R_i \subset X$ is approximated by a geometrically-contracted (i.e., pixel decimated) and affine greyscale-modified copy of an image subblock $u(D_j)$ that is supported on a *larger*, $m \times m$-pixel *domain block* $D_j \subset X$, where $m > n$. (Typically, $m = 2n$.) We refer the reader to [1] for more details regarding the greyscale fractal transform.

The formalism over greyscale images extends naturally to our M-channel digital hyperspectral image model of Section 3. In addition to the partition $\mathcal{R}^{(n)}$ of the support X formed by $n \times n$-pixel blocks, R_i, we introduce an associated *domain pool* $\mathcal{D}^{(m)}$ of $m \times m$-pixel blocks, $D_j \subset X$, with $m = 2n$. This set of blocks need not be overlapping but we normally require that $\cup_j D_j = X$.

Given an M-channel digital hyperspectral image u, an associated *fractal transform operator* T may be defined as follows: For each image block $u(R_i)$, $R_i \in \mathcal{R}^{(n)}$, we choose from $\mathcal{D}^{(m)}$ a domain block $u(D_{j(i)})$ to produce an approximation of the form,

$$u(R_i) \;\approx\; (Tu)(R_i) \;:=\; \alpha_i\, u(D_{j(i)})' + \underline{\beta_i} \qquad 1 \le i \le N_R \quad \text{(Case 4)}. \tag{12}$$

Here, N_R denotes the cardinality of the set $\mathcal{R}^{(n)}$ and the prime denotes an appropriate $2n \times 2n \to n \times n$ pixel decimation operation. The set of range-domain assignments $(i, j(i))$, $1 \le i \le N_R$, and affine transformation parameters $(\alpha_i, \underline{\beta_i})$ define a fractal transform T. Numerical experiments show that the *cross-scale* Case 4 error distributions are very similar to their *same scale* Case 2(b) counterparts of Eq. (6), as observed for (single-channel) greyscale images in [1].

Here we state without proof that under appropriate conditions on the α_i and the contraction factors $c_i = 1/4$ associated with the geometric transformations from domain to range blocks, the fractal transform T is contractive on the metric space (Y, d_Y) of hyperspectral images. From Banach's Fixed Point Theorem, these exists a unique fixed point of T, a hyperspectral image $\bar{u} \in Y$ such that $\bar{u} = T\bar{u}$. Moreover, $\bar{u}$ may be generated by the iteration procedure $v_{n+1} = Tv_n$, where $v_0 \in Y$ is any "seed" image, since $v_n \to \bar{u}$ as $n \to \infty$. (In fact, because we are working with discrete digital images, $v_K = \bar{u}$ for some finite integer K.)

The error associated with the approximation $u \approx Tu$ in Eq. (12) is $\|u - Tu\|_2$, where $\|\cdot\|_2$ denotes the norm on $L^2(X)$. It may be related to the distance between u and $\bar{u}$, the fixed point of T by means of the following consequence of Banach's Theorem, known in the fractal coding literature as the *Collage Theorem* [3],

$$\|u - \bar{u}\| \le \frac{1}{1 - c_T} \|u - Tu\|. \tag{13}$$

Here, c_T denotes the contraction factor of T.

Eq. (13) provides the basis of *fractal image coding* [11,14], in which an image u is approximated by the fixed point $\bar{u}$ of a fractal transform operator T. One tries to make the approximation error $\|u - \bar{u}\|$ by minimizing the so-called *collage error* $\|u - Tu\|$. With reference to Eq. (12), such *collage coding* is accomplished as follows: For each range block $u(R_i)$, choose the domain block $u(D_{j(i)})$ which *best approximates* $u(R_i)$, i.e., minimizes the approximation error $\|u(R_i) - (Tu)(R_i)\|_2$. This procedure minimizes the total collage error $\|u - Tu\|$.

Acknowledgements. We gratefully acknowledge that this research has been supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Discovery Grant (ERV).

References

1. Alexander, S.K., Vrscay, E.R., Tsurumi, S.: A simple, general model for the affine self-similarity of images. In: Campilho, A., Kamel, M.S. (eds.) ICIAR 2008. LNCS, vol. 5112, pp. 192–203. Springer, Heidelberg (2008)
2. AVIRIS hyperspectral image, Yellowstone calibrated scene 0, available from Information Processing Group website, Jet Propulsion Laboratory, California Institute of Technology, `http://compression.jpl.nasa.gov/hyperspectral/`
3. Barnsley, M.F.: Fractals Everywhere. Academic Press, New York (1988)
4. Brunet, D., Vrscay, E.R., Wang, Z.: Structural similarity-based affine approximation and self-similarity of images revisited. In: Kamel, M., Campilho, A. (eds.) ICIAR 2011, Part II. LNCS, vol. 6754, pp. 264–275. Springer, Heidelberg (2011)
5. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Modelling and Simulation 4, 490–530 (2005)
6. Chang, C.: Hyperspectral Data Exploitation, Theory and Applications. John Wiley and Sons, Hoboken (2007)
7. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Proc. 16, 2080–2095 (2007)
8. Ebrahimi, M., Vrscay, E.R.: Solving the inverse problem of image zooming using "Self-examples". In: Kamel, M.S., Campilho, A. (eds.) ICIAR 2007. LNCS, vol. 4633, pp. 117–130. Springer, Heidelberg (2007)
9. Elad, M., Datsenko, D.: Example-based regularization deployed to super-resolution reconstruction of a single image. The Computer Journal 50, 1–16 (2007)
10. Etemoglu, C., Cuperman, V.: Structured vector quantization using linear transforms. IEEE Trans. Sig. Proc. 51, 1625–1631 (2003)
11. Fisher, Y. (ed.): Fractal Image Compression: Theory and Application. Springer, New York (1995)
12. Hyperspectral images of natural scenes 2004, "Scene 2" (2004), `http://personalpages.manchester.ac.uk/staff/david.foster/default.htm`
13. 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, 470–507 (2009)
14. Lu, N.: Fractal Imaging. Academic Press, New York (1997)
15. Michailovich, O., La Torre, D., Vrscay, E.R.: Function-valued mappings, total variation and compressed sensing for diffusion MRI. In: Campilho, A., Kamel, M. (eds.) ICIAR 2012, Part II. LNCS, vol. 7325, pp. 286–295. Springer, Heidelberg (2012)
16. Motta, G., Rizzo, F., Storer, J.A. (eds.): Hyperspectral Data Compression. Springer Science+Business Media, Inc., New York (2006)
17. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? A new look at signal fidelity measures. IEEE Sig. Proc. Mag. 26, 98–117 (2009)
18. Zhang, D., Wang, Z.: Image information restoration based on long-range correlation. IEEE Trans. Cir. Syst. Video Tech. 12, 331–341 (2002)

Appendix

The approximation problem in Eq. (9) may be expressed in the form

$$y_{ik} \approx \alpha x_{ik} + \beta_k, \qquad 1 \le i \le N, \;\; 1 \le k \le M, \tag{14}$$

where $N = n^2$. We wish to find the parameters α and β_k which minimize the squared L^2 distance,

$$\Delta^2 = \sum_{i=1}^{N} \sum_{k=1}^{M} \left(y_{ik} - \alpha x_{ik} - \beta_k \right)^2 . \tag{15}$$

Imposition of the stationarity criterion $\dfrac{\partial \Delta^2}{\partial a} = 0$ yields the following linear equation in α and the β_k,

$$\left(\sum_{i=1}^{N} \sum_{k=1}^{M} x_{ik}^2 \right) \alpha + \sum_{k=1}^{M} \left(\sum_{i=1}^{N} x_{ik} \right) \beta_k = \sum_{i=1}^{N} \sum_{k=1}^{M} x_{ik} y_{ik} \, . \tag{16}$$

Imposition of the stationarity criteria $\dfrac{\partial \Delta^2}{\partial \beta_k} = 0$, $1 \le k \le M$, yields the following linear equations in α and the β_k,

$$\left(\sum_{i=1}^{N} x_{ik} \right) \alpha + n \beta_k = \sum_{i=1}^{N} y_{ik} \, , \qquad 1 \le k \le M. \tag{17}$$

Dividing both sides of the above equation by n yields the relations,

$$\beta_k = \bar{y}_k - \alpha \bar{x}_k, \qquad 1 \le k \le M, \tag{18}$$

where

$$\bar{x}_k = \frac{1}{N} \sum_{i=1}^{N} x_{ik}, \qquad \bar{y}_k = \frac{1}{N} \sum_{i=1}^{N} y_{ik}, \tag{19}$$

denote the means over each channel. Substitution of (18) into Eq. (17) yields the following solution for α:

$$\alpha = \frac{\sum_{k=1}^{M} \sum_{i=1}^{N} x_{ik} \left(y_{ik} - \bar{y}_k \right)}{\sum_{i=1}^{N} \sum_{k=1}^{M} x_{ik}^2 - n \sum_{k=1}^{M} \bar{x}_k^2} \, . \tag{20}$$

From this result, we may solve for the β_k using Eq. (18).