Very High Quality Image Restoration by Combining Wavelets and Curvelets

J.-L. Starck [a], D.L. Donoho [b] and E.J. Candès [c]

[a] DAPNIA/SEI-SAP, CEA/Saclay, 91191 Gif sur Yvette, France
[b] Statistics Department, Stanford University,
Sequoia Hall, Stanford, CA 94305 USA
[c] Applied and Computational Mathematics, Mail Code 217-50,
California Institute of Technology, Pasadena, CA 91125, USA

ABSTRACT

We outline digital implementations of two newly developed multiscale representation systems, namely, the ridgelet and curvelet transforms. We apply these digital transforms to the problem of restoring an image from noisy data and compare our results with those obtained via well established methods based on the thresholding of wavelet coefficients. We develop a methodology to combine wavelets together these new systems to perform noise removal by exploiting all these systems simultaneously. The results of the combined reconstruction exhibits clear advantages over any individual system alone. For example, the residual error contains essentially no visually intelligible structure: no structure is lost in the reconstruction.

KEYWORDS

Wavelet, ridgelet, curvelet, multiscale methods, filtering, image restoration.

1. INTRODUCTION

Wavelets and related multiscale representations pervade all areas of signal processing. The recent inclusion of wavelet algorithms in JPEG 2000 –the new still-picture compression standard– testifies to this lasting and significant impact. A series of recent papers,[1,2] however, argued that wavelets and related classical multiresolution ideas are playing with a limited dictionary made up of roughly isotropic elements occurring at all scales and locations. We view as a limitation the facts that those dictionaries do not exhibit highly anisotropic elements and that there is only a fixed number of directional elements, independent of scale. Despite the success of the classical wavelet viewpoint, there are objects, e.g. images that do not exhibit isotropic scaling and thus call for other kinds of multiscale representation. In short, the theme of this line of research is to show that classical multiresolution ideas only address a portion of the whole range of interesting multiscale phenomena and that there is an opportunity to develop a whole new range of multiscale transforms.

Following on this theme, Candès and Donoho introduced new multiscale systems like curvelets[2] and ridgelets[3] which are very different from wavelet-like systems. Curvelets and ridgelets take the the form of basis elements which exhibit very high directional sensitivity and are highly anisotropic. In two-dimensions, for instance, curvelets are localized along curves, in three dimensions along sheets, etc. Continuing at this informal level of discussion we will rely on an example to illustrate the fundamental difference between the wavelet and ridgelet approaches –postponing the mathematical description of these new systems.

Consider an image which contains a vertical band embedded in white noise with relatively large amplitude. Figure 1 (top left) represents such an image. The parameters are as follows: the pixel width of the band is 20 and the SNR is set to be 0.1. Note that it is not possible to distinguish the band by eye. The wavelet transform (undecimated wavelet transform) is also incapable of detecting the presence of this object; roughly speaking, wavelet coefficients correspond to averages over approximately isotropic neighborhoods (at different scales) and those wavelets clearly do not correlate very well with the the very elongated structure (pattern) of the object to be detected.

We now turn our attention towards procedures of a very different nature which are based on line measurements. To be more specific, consider an *ideal* procedure which consists in integrating the image intensity over columns; that

Further author information: Send correspondence to jstarck@cea.fr

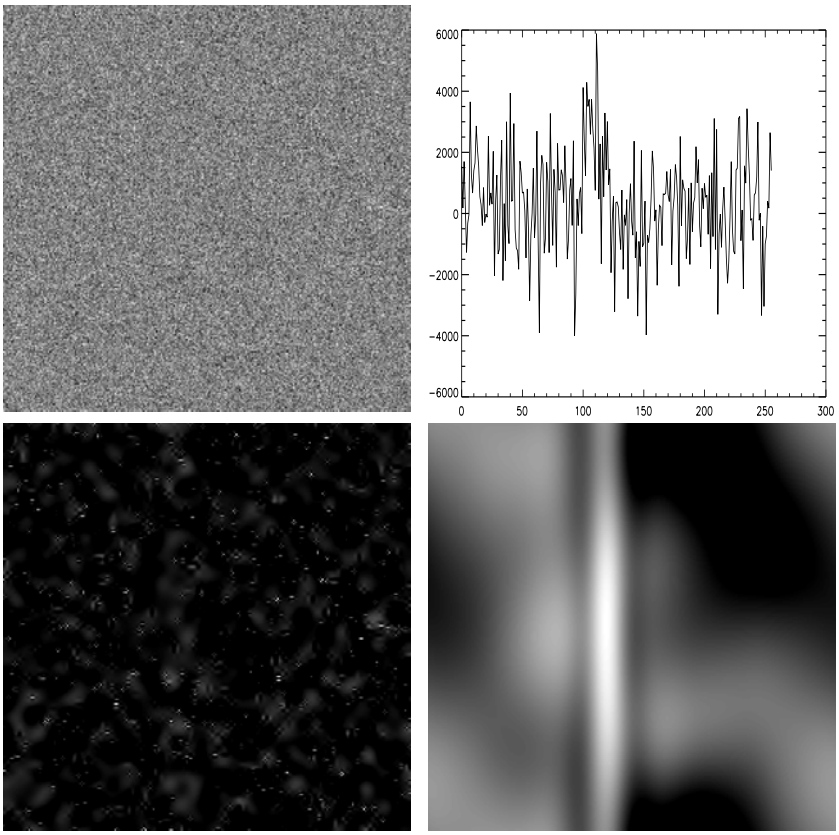

Figure 1. Top, original image containing lines and gaussians. Bottom left, reconstructed image for the à trous wavelet coefficient, bottom right, reconstructed image from the ridgelet coefficients.

is, along the orientation of our object. We use the adjective "ideal" to emphasize the important fact that this method of integration requires a priori knowledge about the structure of our object. This method of analysis gives of course an improved signal to noise ratio for our linear functional better correlate the object in question, see the top right panel of Figure 1.

This example will make our point. Unlike wavelet transforms, the ridgelet transform processes data by first computing integrals over lines with all kinds of orientations and locations. We will explain in the next section how the ridgelet transform further processes those line integrals. For now, we apply naive thresholding of the ridgelet coefficients and "invert" the ridgelet transform; the bottom right panel of Figure 1 shows the reconstructed image. The qualitative difference with the wavelet approach is striking. We observe that this method allows the detection of our object even in situations where the noise level (standard deviation of the white noise) is five times superior to the object intensity.

The contrasting behavior between the ridgelet and the wavelet transforms will be one of the main themes of this paper which is organized as follows. We first briefly review some basic ideas about ridgelet and curvelet representations in the continuum. In parallel to a previous article,[4] Section 2 rapidly outlines a possible implementation strategy and applies our digital curvelet transforms to the problem of removing noise from image data. The numerical results reported here seem very encouraging. We finally develop an approach which combines both the wavelet and curvelet transforms and search for a decomposition which is a solution of an optimization problem in this joint representation. We find that denoising experiments give results of rather unexpected high quality.

2. THE CURVELET TRANSFORM

2.1. The Ridgelet Transform

The two-dimensional continuous ridgelet transform in $\mathbf{R}^2$ can be defined as follows.[3] We pick a smooth univariate function $\psi : \mathbf{R} \to \mathbf{R}$ with sufficient decay and satisfying the admissibility condition

$$\int |\hat{\psi}(\xi)|^2/|\xi|^2 \, d\xi < \infty, \tag{1}$$

which holds if, say, ψ has a vanishing mean $\int \psi(t)dt = 0$. We will suppose a special normalization about ψ so that $\int_0^\infty |\hat{\psi}(\xi)|^2 \xi^{-2} d\xi = 1$.

For each $a > 0$, each $b \in \mathbf{R}$ and each $\theta \in [0, 2\pi)$, we define the bivariate *ridgelet* $\psi_{a,b,\theta} : \mathbf{R}^2 \to \mathbf{R}^2$ by

$$\psi_{a,b,\theta}(x) = a^{-1/2} \cdot \psi((x_1 \cos\theta + x_2 \sin\theta - b)/a); \tag{2}$$

A ridgelet is constant along lines $x_1 \cos\theta + x_2 \sin\theta = const.$ Transverse to these ridges it is a wavelet.

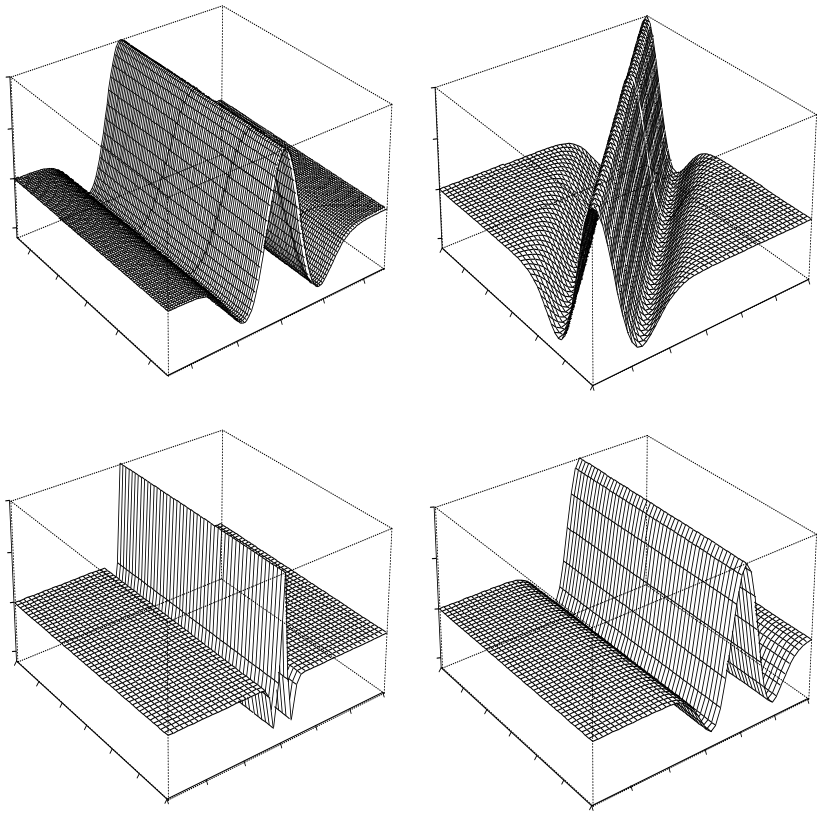

Figure 2. A Few Ridgelets

Figure 2.1 graphs a few ridgelets with different parameter values. The top right, bottom left and right panels are obtained after simple geometric manipulations of the upper left ridgelet, namely rotation, rescaling, and shifting.

Given an integrable bivariate function $f(x)$, we define its ridgelet coefficients by

$$\mathcal{R}_f(a, b, \theta) = \int \overline{\psi}_{a,b,\theta}(x) f(x) dx.$$

We have the exact reconstruction formula

$$f(x) = \int_0^{2\pi} \int_{-\infty}^{\infty} \int_0^{\infty} \mathcal{R}_f(a, b, \theta) \psi_{a,b,\theta}(x) \frac{da}{a^3} db \frac{d\theta}{4\pi} \tag{3}$$

valid a.e. for functions which are both integrable and square integrable.

Ridgelet analysis may be construed as wavelet analysis in the Radon domain. Recall that the Radon transform of an object f is the collection of line integrals indexed by $(\theta, t) \in [0, 2\pi) \times \mathbf{R}$ given by

$$Rf(\theta, t) = \int f(x_1, x_2)\delta(x_1 \cos\theta + x_2 \sin\theta - t)\, dx_1 dx_2, \tag{4}$$

where δ is the Dirac distribution. Then the ridgelet transform is precisely the application of a 1-dimensional wavelet transform to the slices of the Radon transform where the angular variable θ is constant and t is varying.

This viewpoint strongly suggests developing approximate Radon transforms for digital data. This subject has received a considerable attention over the last decades as the Radon transform naturally appears as a fundamental tool in many fields of scientific investigation. Our implementation follows a widely used approach in the literature of medical imaging and is based on discrete fast Fourier transforms. The key component is to obtain approximate digital samples from the Fourier transform on a polar grid, i.e. along lines going through the origin in the frequency plane. Because of space limitations, we will not detail our approach and refer to the reader to.[4] As a compromise Figure 3 (left) represents the flowgraph of the ridgelet transform.

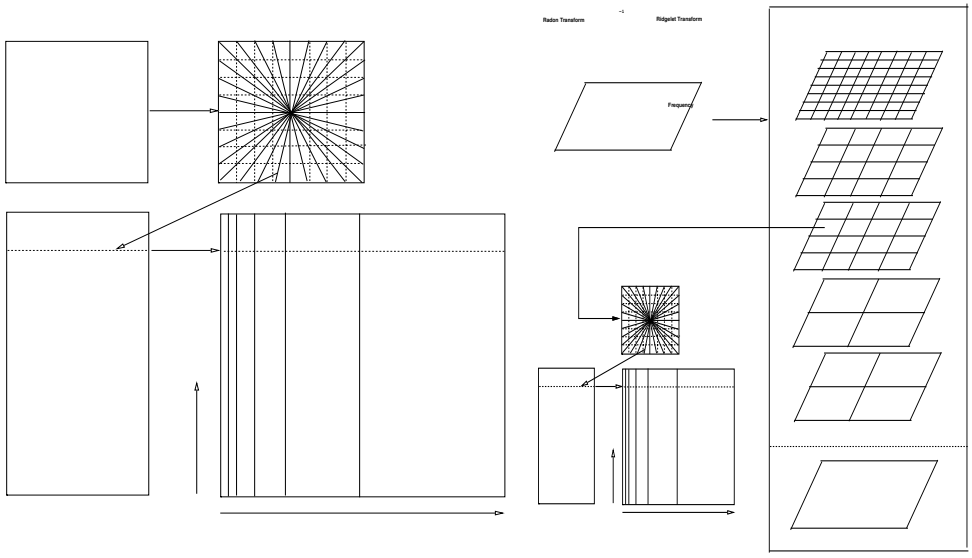

Figure 3. Left, ridgelet transform flowgraph. Each of the $2n$ radial lines in the Fourier domain is processed separately. The 1-D inverse FFT is calculated along each radial line followed by a 1-D nonorthogonal wavelet transform. In practice, the one-dimensional wavelet coefficients are directly calculated in the Fourier space. Right, Curvelet transform flowgraph. The figure illustrates the decomposition of the original image into subbands followed by the spatial partitioning of each subband. The ridgelet transform is then applied to each block.

The ridgelet transform of a digital array of size $n \times n$ is an array of size $2n \times 2n$ and hence introduces a redundancy factor equal to 4.

Local Ridgelet Transforms

Speaking in engineering terms, one might say that the ridgelet transform is well adapted for picking linear structures of about the size of the image. However, interesting linear structures, e.g line segments, may occur at a wide range of scales. Following upon a well-established tradition in time-frequency analysis, there is an opportunity to develop a pyramid of ridgelet transforms. We may indeed apply classical ideas such as recursive dyadic partitioning, and thereby construct dictionaries of windowed ridgelets, renormalized and transported to a wide range of scales and locations.

To make things more explicit we consider the situation at a fixed scale. The image is decomposed into smoothly overlapping blocks of sidelength b pixels in such a way that the overlap between two vertically adjacent blocks is a

rectangular array of size b by $b/2$; we use overlap to avoid blocking artifacts. For an n by n image, we count $2n/b$ such blocks in each direction. The partitioning introduces redundancy, as a pixel belongs to 4 neighboring blocks. More details on a possible implementation of the digital ridgelet transform can be found in.[4] Taking the ridgelet transform of these smoothly localized data is what we call the local ridgelet transform.

2.2. Digital Curvelet Transform

The idea of the curvelet transform[2] is first to decompose the image into subbands, i.e. to separate the object out into a series of disjoint scales. Each scale is then analyzed by means of a local ridgelet transform. Roughly speaking, different levels of the multiscale ridgelet pyramid are used to represent different subbands of a filter bank output. The key point is that there is a very special relationship between the depth of the multiscale pyramid and the index of the dyadic subbands; the sidelength of the localizing windows is doubled *at every other* dyadic subband, hence maintaining the fundamental property of the curvelet transform which says that elements of length about $2^{-j/2}$ serve for the analysis and synthesis of the j-th subband $[2^j, 2^{j+1}]$.

In our opinion the "à trous" subband filtering algorithm is especially well-adapted to the needs of the digital curvelet transform. The algorithm decomposes an n by n image I as a superposition of the form

$$I(x,y) = c_J(x,y) + \sum_{j=1}^{J} w_j(x,y),$$

where c_J is a coarse or smooth version of the original image I and w_j represents 'the details of I' at scale 2^{-j}, see[5] for more information. Thus, the algorithm outputs $J+1$ subband arrays of size $n \times n$. (The indexing is such that, here, $j = 1$ corresponds to the finest scale (high frequencies).)

As a side comment, we will also note that the coarse description of the image c_J is not processed. We used the default value $B_{min} = 16$ pixels in our implementation. Figure 3 (right) gives an overview of the organization of the algorithm.

This implementation of the curvelet transform is redundant. The redundancy factor is equal to $16J+1$ whenever J scales are employed. Finally, the method enjoys exact reconstruction and stability, because each step of the transform is both invertible and stable.

2.3. Filtering by the curvelet transform

In our first example, a Gaussian white noise with a standard deviation fixed to 20 was added to a subimage of the classical `Lenna` image (512 by 512). We employed several methods to filter the noisy image:

1. Thresholding of the Curvelet transform.
2. Bi-orthogonal wavelet de-noising methods using the Dauchechies-Antonini 7/9 filters (FWT-7/9) and hard thresholding.
3. Undecimated wavelet de-noising using the Dauchechies-Antonini 7/9 filters (UWT-7/9) and hard thresholding.

Our experiments are reported on Figure 4. The undecimated wavelet transform approach outputs a PSNR comparable to that obtained via the curvelet transform, but the curvelet reconstruction does not contain the quantity of disturbing artifacts along edges that one sees in wavelet reconstructions. An examination of the details of the restored images (Figure 4) is instructive. One notices that the decimated wavelet transform exhibits distortions of the boundaries and suffers substantial loss of important detail. The undecimated wavelet transform gives better boundaries, but completely omits to reconstruct certain ridges in the hatband. In addition, it exhibits numerous small-scale embedded blemishes; setting higher thresholds to avoid these blemishes would cause even more of the intrinsic structure to be missed.

Further results are visible at the following URL: `http://www-stat.stanford.edu/~jstarck`.

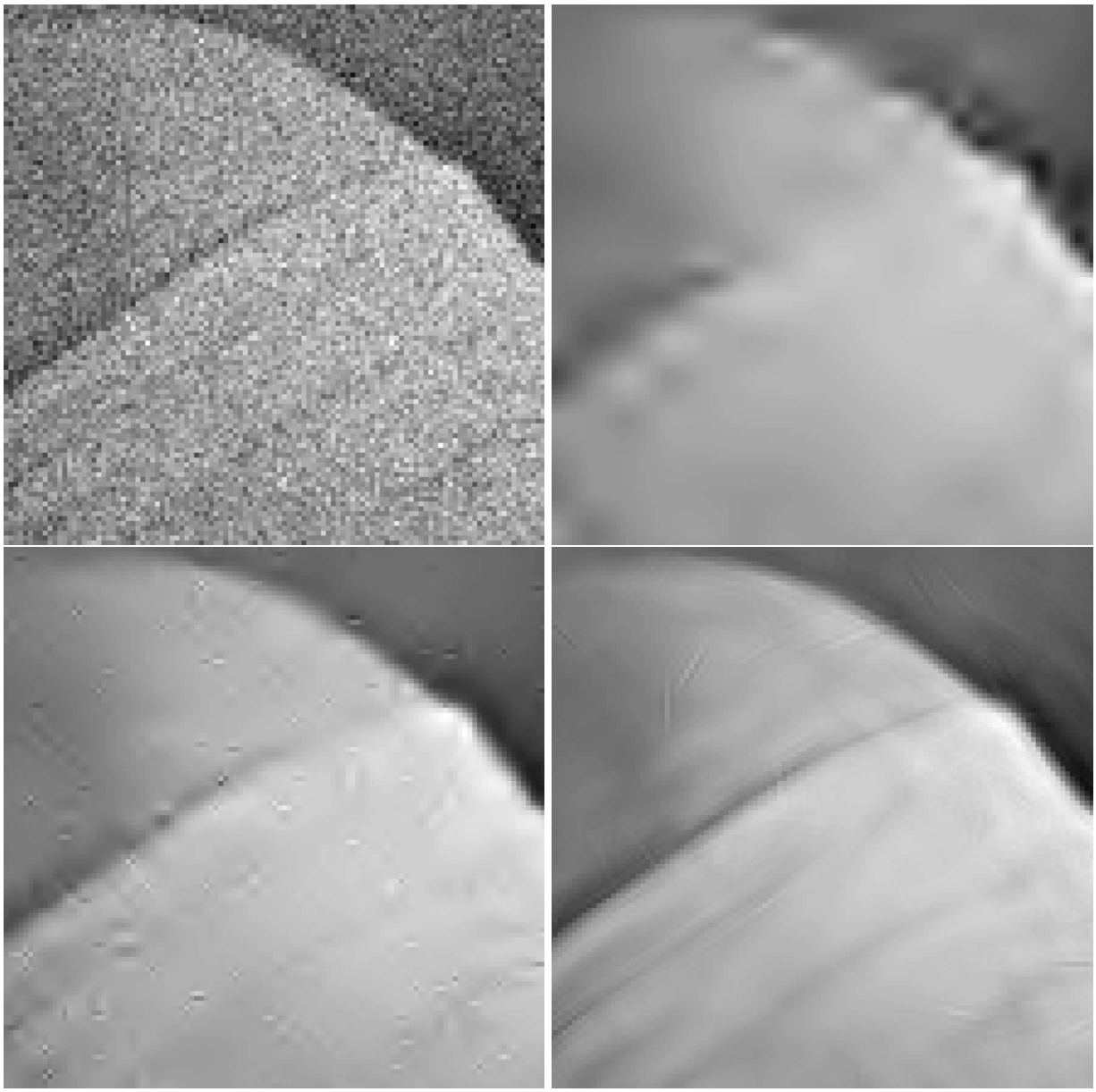

Figure 4. The figure displays the noisy image (top left), and the restored images after denoising by means of the DWT (top right), UWT (bottom left), and the curvelet transform (bottom right). The diagonal lines of the hat have been recovered with much greater fidelity in the curvelet approach.

3. THE COMBINED APPROACH

Although the results obtained by simply thresholding curvelet expansion are encouraging, there is of course ample room for further improvement. A quick inspection of the residual images for both the wavelet and curvelet transforms shown on figure 5 reveals the existence of very different features. For instance, wavelets do not restore long edges with high fidelity while curvelets are challenged with small features such as `Lenna`'s eyes. Loosely speaking, each transform has its own area of expertise and this complementarity may be of great potential. This section will develop a denoising strategy based on the idea of combining both transforms.

The idea of combining several representations is not new.[6–8] There are well-known general principles for synthesizing a signal s as a "sparse" linear combination $s = \sum_\gamma a_\gamma \varphi_\gamma$ of elements taken from a of dictionary of waveforms $\mathcal{D} = (\varphi_\gamma)_{\gamma \in \Gamma}$. Two of the most frequently discussed approaches are the Matching Pursuit (MP)[9] and the Basis Pur-

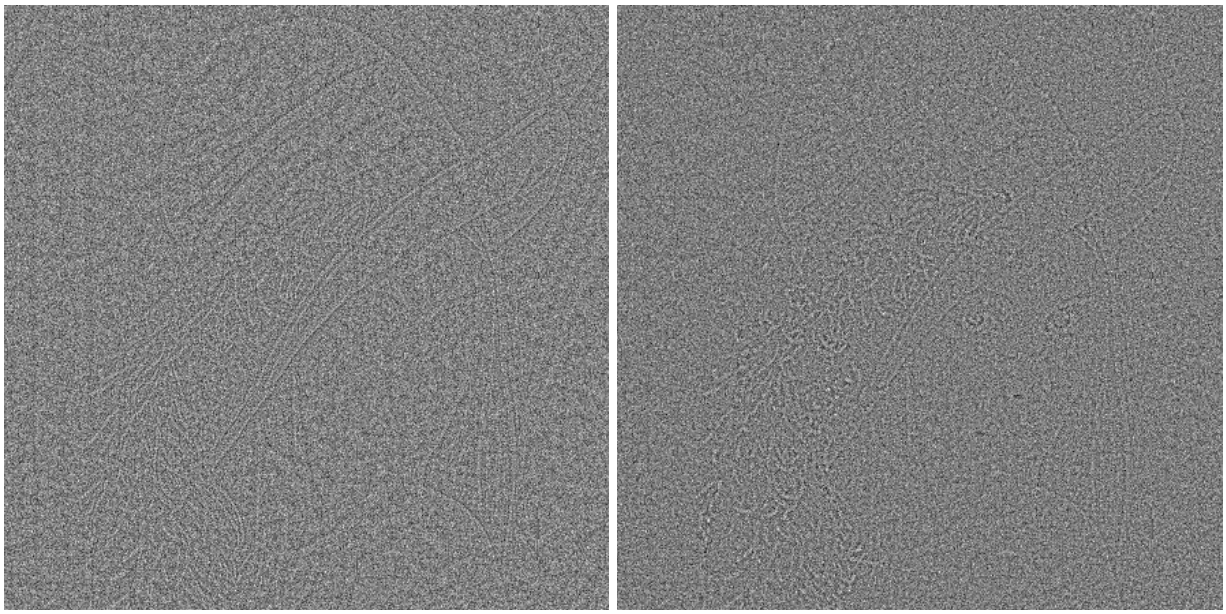

Figure 5. Residual for thresholding of the undecimated wavelet transform and thresholding of the curvelet transform.

suit (BP).[6] MP is a greedy procedure that operates in a stepwise fashion by sequentially adding elements that best correlate with the current residual. In stark contrast, BP is a global procedure which synthesizes an approximation $\tilde{s}$ to s by minimizing a functional of the type

$$\|s - \tilde{s}\|_{\ell_2}^2 + \lambda \cdot \|\alpha\|_{\ell_1}, \quad \tilde{s} = \Phi\alpha. \tag{5}$$

In many cases, BP or MP synthesis algorithms are computationally very expensive. We propose here an alternative approach for the sole purpose of image restoration. The problem that we will want to solve in this combined approach is not to represent a signal with overcomplete dictionaries, but to make sure that our reconstruction will incorporate information judged as significant by any of our representations.

In general, suppose that we are given K linear transforms $T_1, \ldots, T_K$ and let α_k be the coefficient sequence of an object x after applying the transform T_k, i.e. $\alpha_k = T_k x$. We will suppose that for each transform T_k we have available a reconstruction rule that we will denote by T_k^{-1} although this is clearly an abuse of notations. Finally, T will denote the block diagonal matrix with the T_k's as building blocks and α the amalgamation of the α_k's.

A hard thresholding rule associated with the transform T_k synthesizes an estimate $\tilde{s}_k$ via the formula

$$\tilde{s}_k = T_k^{-1}\delta(\alpha_k) \tag{6}$$

where δ is a rule that sets to zero all the coordinates of α_k whose absolute value falls below given a sequence of thresholds (such coordinates are said to be nonsignificant).

In practice, a widely used approach is to compute the average of the $\tilde{s}_k$'s giving a reconstruction of the form

$$\tilde{s} = \sum_k \tilde{s}_k / K. \tag{7}$$

For instance, in the literature of image processing it is common to average reconstructions obtained after thresholding the wavelet coefficients of translated versions of the original dataset (cycle-spinning), i.e. the T_k's are obtained by composing translations and the wavelet transform. In our setup, we do not find this solution very appealing as this creates the opportunity to average high-quality and low-quality reconstructions.

Given data y of the form $y = s + \sigma z$, where s is the image we wish to recover and z is standard white noise, we propose solving the following optimization problem:

$$\min \|T\tilde{s}\|_{\ell_1}, \quad \text{subject to} \quad s \in C, \tag{8}$$

where C is the set of vectors $\tilde{s}$ which obey the linear constraints

$$\begin{cases} \tilde{s} \geq 0, \\ |T\tilde{s} - Ty| \leq e; \end{cases} \tag{9}$$

here, the second inequality constraint only concerns the set of significant coefficients, i.e. those indices μ such that $\alpha_\mu = (Ty)_\mu$ exceeds (in absolute value) a threshold t_μ. Given a vector of tolerance (e_μ), we seek a solution whose coefficients $(T\tilde{s})_\mu$ are within e_μ of the noisy empirical α_μ's. Think of α_μ as being given by

$$y = \langle y, \varphi_\mu \rangle,$$

so that α_μ is normally distributed with mean $\langle f, \varphi_\mu \rangle$ and variance $\sigma_\mu^2 = \sigma^2 \|\varphi_\mu\|_2^2$. In practice, the threshold values range typically between three and four times the noise level σ_μ and in our experiments we will put $e_\mu = \sigma_\mu/2$. In short, our constraints guarantees that the reconstruction will take into account any pattern which is detected as significant by a any of the K transforms.

We use an ℓ_1 penalty on the coefficient sequence because we are interested in *low complexity* reconstructions. There are other possible choices of complexity penalties; for instance, an alternative to (8) would be

$$\min \|\tilde{s}\|_{TV}, \quad \text{subject to} \quad s \in C.$$

where $\|\cdot\|_{TV}$ is the Total Variation norm, i.e. the discrete equivalent of the integral of the euclidian norm of the gradient.

We propose solving (8) using the method of hybrid steepest descent (HSD).[10] HSD consists in building the sequence

$$s^{n+1} = P(s^n) - \lambda_{n+1} \nabla_J (P(s^n)); \tag{10}$$

here, P is the ℓ_2 projection operator onto the feasible set C, ∇_J is the gradient of equation 8, and $(\lambda_n)_{n\geq 1}$ is a sequence obeying $(\lambda_n)_{n\geq 1} \in [0,1]$ and $\lim_{n\to+\infty} \lambda_n = 0$.

Unfortunately, the projection operator P is not easily determined and in practice we will use the following proxy; compute $T\tilde{s}$ and replace those coefficients which do not obey the constraints $|T\tilde{s} - Ty| \leq e$ (those which fall outside of the prescribed interval) by those of y; apply the inverse transform.

The combined filtering algorithm (CFA) is:

1. Initialize $L_{\max} = 1$, the number of iterations N_i, and $\delta_\lambda = \frac{L_{\max}}{N_i}$.
2. Estimate the noise standard deviation σ, and set $e_k = \frac{\sigma}{2}$.
3. for k = 1, .., K calculate the transform: $\alpha_k^{(s)} = T_k s$.
4. Set $\lambda = L_{\max}$, $n = 0$, and $\tilde{s}^n$ to 0.
5. While $\lambda >= 0$ do
   - $u = \tilde{s}^n$.
   - for k = 1, .., K do
     - Calculate the transform $\alpha_k = T_k u$.
     - For all coefficients $\alpha_{k,l}$ do
       - Calculate the residual $r_{k,l} = \alpha_{k,l}^{(s)} - \alpha_{k,l}$
       - if $\alpha_{k,l}^{(s)}$ is significant and $\mid r_{k,l} \mid > e_{k,l}$ then $\alpha_{k,l} = \alpha_{k,l}^{(s)}$
       - $\alpha_{k,l} = sgn(\alpha_{k,l})(\mid \alpha_{k,l} \mid -\lambda)_+$.
     - $u = T_k^{-1} \alpha_k$
   - Threshold negative values in u and $\tilde{s}^{n+1} = u$.
   - $n = n + 1$, $\lambda = \lambda - \delta_\lambda$, and goto 5.

Figure 6. Noisy image (top left), and filtered images by the undecimated wavelet transform (top right), the curvelet transform (bottom left) and the combined transforms (bottom right).

3.0.1. Filtering

The noisy `Lenna` image (the noise standard deviation being equal 20) has been filtered by the undecimated wavelet transform, the curvelet transform, and by our combined transform approach (curvelet and undecimated wavelet transforms). The results are reported in Table 1. Figure 6 displays the noisy image (top left), and the restored images after denoising by means of the undecimated wavelet transform (UWT) (top right), the curvelet transform (bottom left), and the combined transforms (bottom right). Details are displayed on Figure 7 left. Figure 7 right shows the full residual image, and can be compared to the residual images shown in Figure 5. The residual is much better when the combined filtering is applied, and no feature can anymore be detected by eye. This was not the case for both the wavelet and the curvelet filtering.

Figure 8 displays the PSNR of the solution versus the number of iterations. On this example, the algorithm is shown to converge rapidly. From a practical viewpoint only four or five iterations are truly needed. Note that the result obtained after a single iteration is already superior than those available using methods based on the thresholding of wavelet or curvelet coefficients alone.

Method	PSNR	Comments
Noisy image	22.13	
OWT7-9 + ksigma Hard thresh.	28.35	many artifact
UWT7-9 + ksigma Hard thresh.	31.94	very few artifact
Curvelet (B=16)	31.95	no artifact
Combined filtering	32.72	no artifact

Table 1. PSNR after filtering the simulated image (Lena + Gaussian noise (sigma=20)). In the combined filtering, a curvelet and an undecimated wavelet transform have been used.

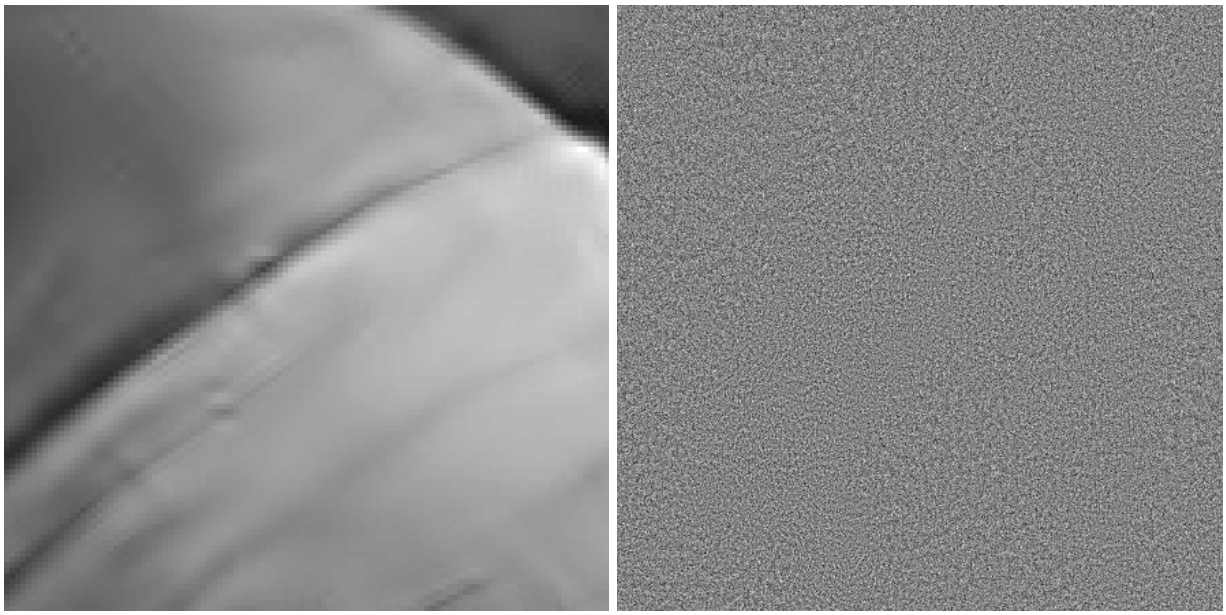

Figure 7. The left panel shows a detail of the filtered image by the combined method. The full residual image is displayed on the right.

4. CONCLUSION

We believe that the denoising experiments presented in this paper are of very high quality;

1. The combined filtering leads to a real improvement both in terms of PSNR and visual appearance.
2. The combined approach arguably challenges the eye to distinguish structure/features from residual images on real image data (at least for the range of noise levels that was considered here). Single transforms cannot manage such a feat.

We also note that the combined reconstruction may tend to be free of major artifacts which is very much unlike typical thresholding rules. Although the ease of implementation is clear we did not address the computational issues associated with our method. In a nutshell, the algorithms we described requires calculating each transform and its inverse only a limited number of times.

In our examples, we constructed a combined transform from linear transforms (wavelets, ridgelets and curvelets) but our paradigm extends to any kind of nonlinear transform such as the Pyramidal Median Transform[5] or morphological multiscale transforms.[11]

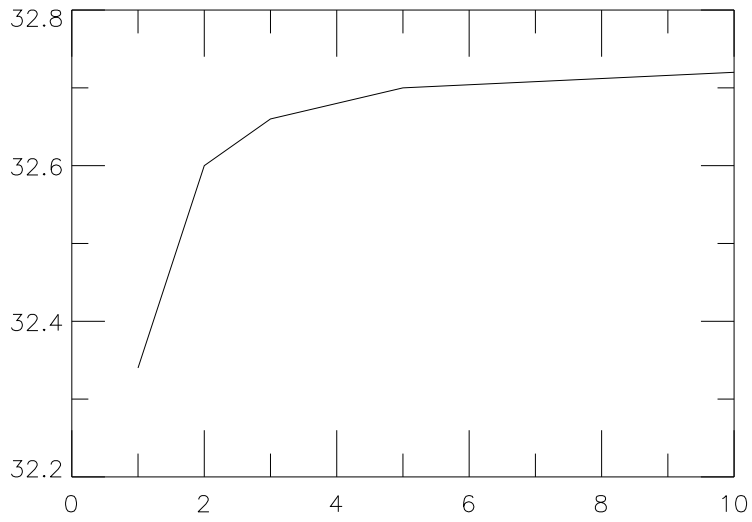

Figure 8. PSNR versus the number of iterations.

REFERENCES

1. E. J. Candès and D. L. Donoho, "Ridgelets: the Key to Higher-dimensional Intermittency?," *Phil. Trans. R. Soc. Lond. A.* **357**, pp. 2495–2509, 1999.
2. E. J. Candès and D. L. Donoho, "Curvelets – a surprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting: Saint-Malo 1999*, A. Cohen, C. Rabut, and L. Schumaker, eds., Vanderbilt University Press, (Nashville, TN), 1999.
3. E. J. Candès, "Harmonic analysis of neural netwoks," *Applied and Computational Harmonic Analysis* **6**, pp. 197–218, 1999.
4. J. Starck, E. Candès, and D. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing* , 2001. submitted.
5. J. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge University Press, Cambridge (GB), 1998.
6. S. Chen, D. Donoho, and M. Saunder, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing* **20**(1), pp. 33–61, 1998.
7. F. Meyer, A. Averbuch, J.-O. Stromberg, and R. Coifman, "Multi-layered image representation: Application to image compression," in *International Conference on Image Processing, ICIP'98, Chicago*, 1998.
8. X. Huo, *Sparse Image Representation via Combined Transforms.* PhD thesis, Stanford Univesity, August 1999.
9. S. Mallat and Z. Zhang, "Atomic decomposition by basis pursuit," *IEEE Transactions on Signal Processing* **41**(12), pp. 3397–3415, 1993.
10. I. Yamada, "Inherently parallel algorithm for feasibility and optimization," Elsevier, 2001.
11. H. Heijmans and J. Goutsias, "Multiresolution signal decomposition schemes. part 2: Morphological wavelets," *IEEE Transactions on Image Processing* **9**(11), pp. 1897–1913, 2000.