Astronomy & Astrophysics manuscript no. November 5, 2002
(accepted)

Astronomical Image Representation by the Curvelet Transform

J.L. Starck[1], D.L. Donoho[2] and E.J. Candès[3]

[1] DAPNIA/SEDI-SAP, Service d'Astrophysique, CEA-Saclay, F-91191 Gif-sur-Yvette Cedex, France.
[2] Department of Statistics, Stanford University, Sequoia Hall, Stanford, CA 94305 USA.
[3]Department of Applied Mathematics, Mail Code 217-50, California Institute of Technology, Pasadena, CA 91125, USA.

November 5, 2002

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 show that the curvelet transform allows us also to well enhance elongated features contained in the data. Finally, we describe the Morphological Component Analysis, which consists in separating features in an image which do not present the same morphological characteristics. A range of examples illustrates the results.

Key words. methods: Data Analysis – techniques: Image Processing

1. Introduction

The wavelet transform has been extensively used in astronomical data analysis during the last ten years. A quick search with ADS shows that around 600 papers contain the keyword "Wavelet" in their abstract, and all astrophysical domains were concerned, from the sun study to the CMB analysis.

This large success of the wavelet transform (WT) is due to the fact that astronomical data presents generally complex hierarchical structures, often described as fractals. Using multiscale approaches such the wavelet transform (WT), an image can be decomposed into components at different scales, and the WT is therefore well-adapted to astronomical data study.

A series of recent papers (Candès and Donoho, 1999d; Candès and Donoho, 1999c), 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 (Candès and Donoho, 1999c) and ridgelets (Candès, 1999) 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 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

Send offprint requests to: jstarck@cea.fr

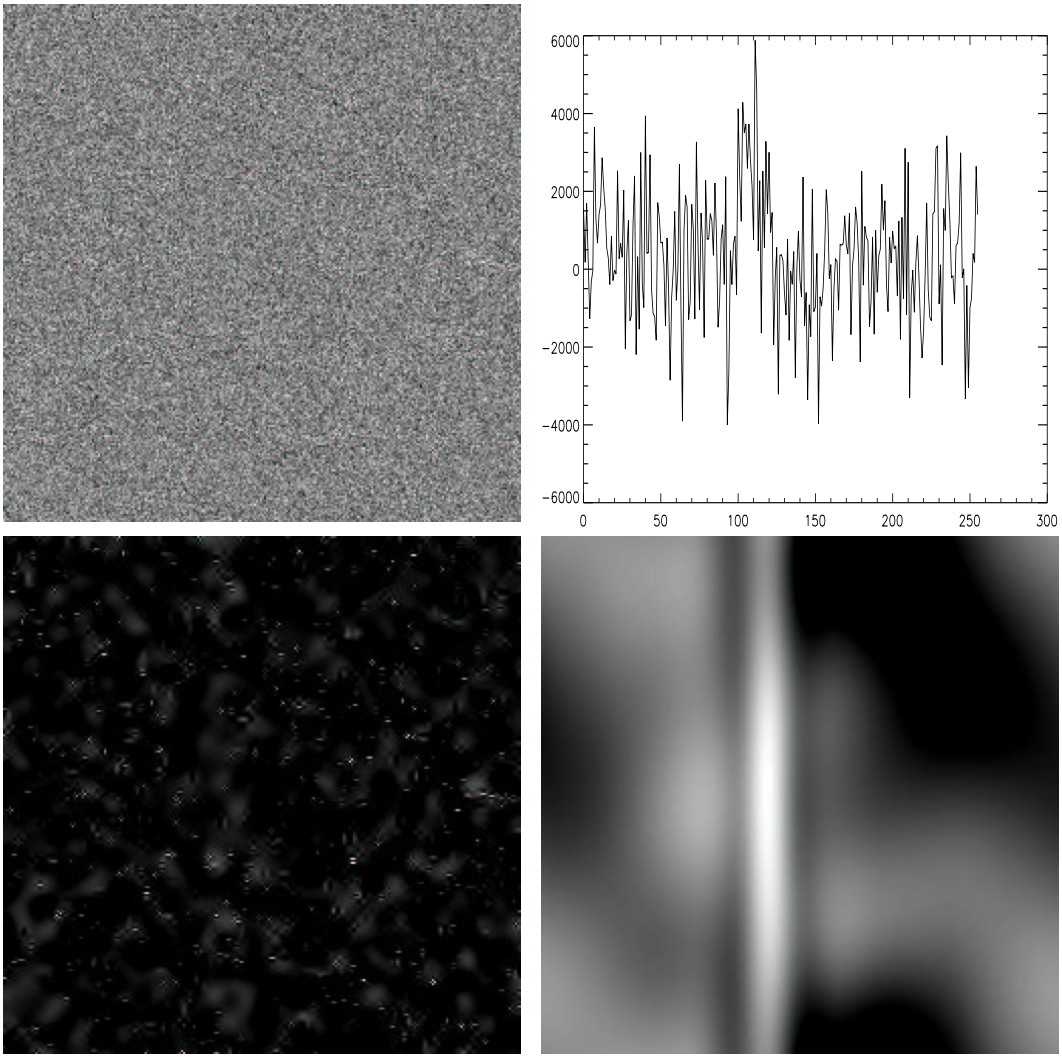

Fig. 1. Top left, original image containing a vertical band embedded in white noise with relatively large amplitude. Top right, signal obtained by integrating the image intensity over columns. Bottom left, reconstructed image for the undecimated wavelet coefficient, bottom right, reconstructed image from the ridgelet coefficients.

columns; that 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 (Starck et al., 2002), Section 2 rapidly outlines a possible implementation strategy. Sections 3 and 4 present respectively how to use the curvelet transform for image denoising and and image enhancement.

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.

2. The Curvelet Transform

2.1. The Ridgelet Transform

The two-dimensional continuous ridgelet transform in $\mathbf{R}^2$ can be defined as follows (Candès, 1999). 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* $\psi_{a,b,\theta} : \mathbf{R}^2 \to \mathbf{R}$ 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 2.1 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 constructed 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.

2.2. An Approximate Digital Ridgelet Transform

2.2.1. Radon tranform

A fast implementation of the RT can be performed in the Fourier domain. First the 2D FFT is computed. Then it is interpolated along a number of straight lines equal to the selected number of projections, each line passing through the origin of the 2D frequency space, with a slope equal to the projection angle, and a number of interpolation points equal to the number of rays per projection. The one dimensional inverse Fourier transform of each interpolated array is then evaluated. The FFT based RT is however not straitforward because we need to interpolate the Fourier domain. Furthermore, if we want to have an exact inverse transform, we have to make sure that the lines pass through all frequencies.

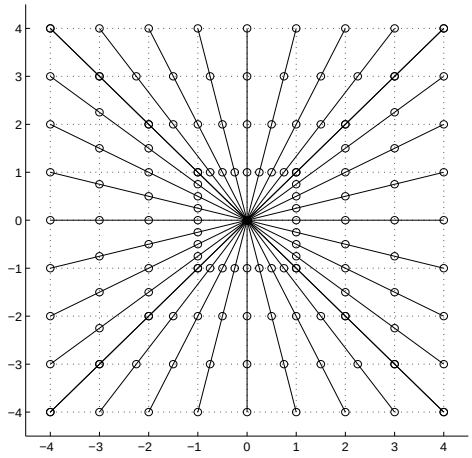

Fig. 3. Illustration of the digital polar grid in the frequency domain for an n by n image ($n = 8$). The figure displays the set of radial lines joining pairs of symmetric points from the boundary of the square. The rectopolar grid is the set of points – marked with circles – at the intersection between those radial lines and those which are parallel to the axes.

For our implementation, we use a pseudo-polar grid. The geometry of the rectopolar grid is illustrated on Figure 2.2.1. We select $2n$ radial lines in the frequency plane obtained by connecting the origin to the vertices (k_1, k_2) lying on the boundary of the array (k_1, k_2) , i.e. such that k_1 or $k_2 \in \{-n/2, n/2\}$. The polar grid $\xi_{\ell,m}$ (ℓ serves to index a given radial line while the position of the point on that line is indexed by m) that we shall use is the intersection between the set of radial lines and that of cartesian lines parallel to the axes. To be more specific, the sample points along a radial line $\mathcal{L}$ whose angle with the vertical axis is less or equal to $\pi/4$ are obtained by intersecting $\mathcal{L}$ with the set of horizontal lines $\{x_2 = k_2, \, k_2 = -n/2, -n/2+1, \ldots, n/2\}$. Similarly, the intersection with the vertical lines $\{x_1 = k_1, \, k_1 = -n/2, -n/2+1, \ldots, n/2\}$ defines our sample points whenever the angle between $\mathcal{L}$ and the horizontal axis is less or equal to $\pi/4$. The cardi-

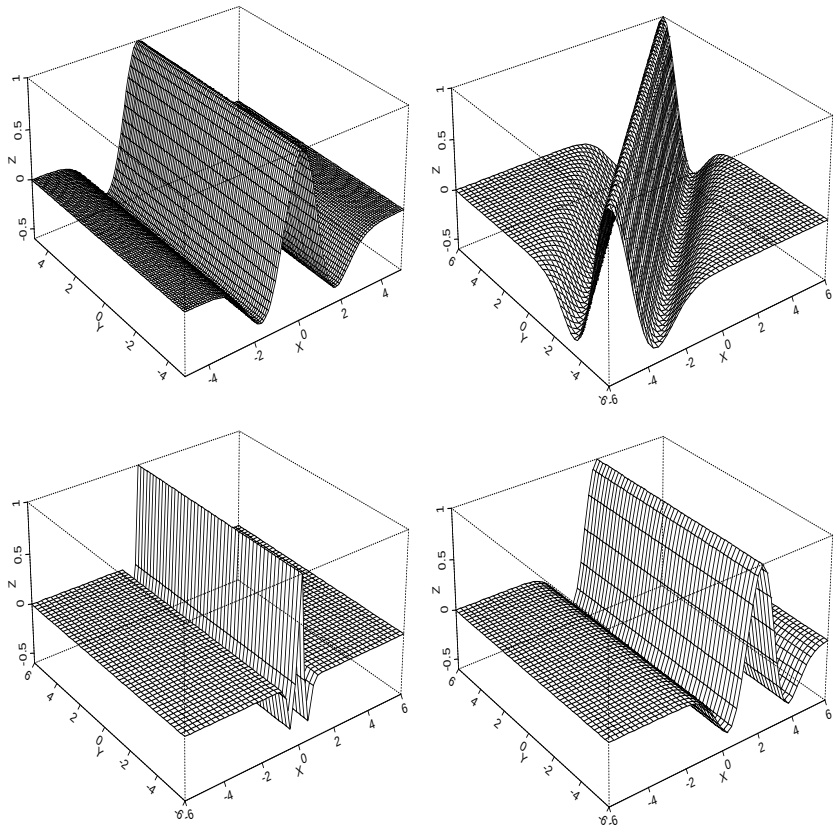

Fig. 2. A Few Ridgelets

nality of the rectopolar grid is equal to $2n^2$ as there are $2n$ radial lines and n sampled values on each of these lines. As a result, data structures associated with this grid will have a rectangular format. We observe that this choice corresponds to irregularly spaced values of the angular variable θ. More details can be found in (Starck et al., 2002).

2.2.2. 1D Wavelet Transform

To complete the ridgelet transform, we must take a one-dimensional wavelet transform along the radial variable in Radon space. We now discuss the choice of digital one-dimensional wavelet transform.

Experience has shown that compactly-supported wavelets can lead to many visual artifacts when used in conjunction with nonlinear processing - such as hard-thresholding of individual wavelet coefficients - particularly for decimated wavelet schemes used at critical sampling. Also, because of the lack of localization of such compactly-supported wavelets in the frequency domain, fluctuations in coarse-scale wavelet coefficients can introduce fine-scale fluctuations; this is undesirable in our setting. Here we take a frequency-domain approach, where the discrete Fourier transform is reconstructed from the inverse Radon transform. These considerations lead us to use band-limited wavelet – whose support is compact in the Fourier domain rather than the time-domain. Other implementations have made a choice of compact support in the frequency domain as well (Donoho, 1998; Donoho, 1997). However, we have chosen a specific overcomplete system, based on work of Starck et al. (1994; 1998), who constructed such a wavelet transform and applied it to interferometric image reconstruction. The wavelet transform algorithm is based on a scaling function ϕ such that $\hat{\phi}$ vanishes outside of the interval $[-\nu_c, \nu_c]$. We defined the scaling function $\hat{\phi}$ as a renormalized B_3-spline

$$\hat{\phi}(\nu) = \frac{3}{2} B_3(4\nu),$$

and $\hat{\psi}$ as the difference between two consecutive resolutions

$$\hat{\psi}(2\nu) = \hat{\phi}(\nu) - \hat{\phi}(2\nu).$$

Because $\hat{\psi}$ is compactly supported, the sampling theorem shows than one can easily build a pyramid of $n + n/2 + \ldots + 1 = 2n$ elements, see (Starck et al., 1998) for details. This transform enjoys the following features:

- The wavelet coefficients are directly calculated in the Fourier space. In the context of the ridgelet transform, this allows avoiding the computation of the one-dimensional inverse Fourier transform along each radial line.
- Each subband is sampled above the Nyquist rate, hence, avoiding aliasing –a phenomenon typically encountered by critically sampled orthogonal wavelet transforms (Simoncelli et al., 1992).

– The reconstruction is trivial. The wavelet coefficients simply need to be co-added to reconstruct the input signal at any given point. In our application, this implies that the ridgelet coefficients simply need to be co-added to reconstruct Fourier coefficients.

This wavelet transform introduces an extra redundancy factor, which might be viewed as an objection by advocates of orthogonality and critical sampling. However, we note that our goal in this implementation is not data compression/efficient coding - for which critical sampling might be relevant - but instead noise removal, for which it well-known that overcompleteness can provide substantial advantages (Coifman and Donoho, 1995).

Figure 4 shows the flowgraph of the ridgelet transform. The ridgelet transform of an image of size $n \times n$ is an image of size $2n \times 2n$, introducing a redundancy factor equal to 4.

We note that, because our transform is made of a chain of steps, each one of which is invertible, the whole transform is invertible, and so has the exact reconstruction property. For the same reason, the reconstruction is stable under perturbations of the coefficients.

Last but not least, our discrete transform is computationally attractive. Indeed, the algorithm we presented here has low complexity since it runs in $O(n^2 \log(n))$ flops for an $n \times n$ image.

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.

2.3. Local Ridgelet Transforms

The ridgelet transform is optimal to find only lines of the size of the image. To detect line segments, a partitioning must be introduced (Candès, 1998). 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. We present two competing strategies to perform the analysis and synthesis:

1. The block values are weighted (analysis) in such a way that the co-addition of all blocks reproduce exactly the original pixel value (synthesis).
2. The block values are those of the image pixel values (analysis) but are weighted when the image is reconstructed (synthesis).

Experiments have shown that the second approach leads to better results. We calculate a pixel value, $f(i,j)$ from its four corresponding block values of half-size $\ell = b/2$, namely, $B_1(i_1, j_1)$, $B_2(i_2, j_1)$, $B_3(i_1, j_2)$ and $B_4(i_2, j_2)$ with $i_1, j_1 > b/2$ and $i_2 = i_1 - \ell, j_2 = j_1 - \ell$, in the following way:

$$\begin{aligned} f_1 &= w(i_2/\ell)B_1(i_1, j_1) + w(1 - i_2/\ell)B_2(i_2, j_1) \\ f_2 &= w(i_2/\ell)B_3(i_1, j_2) + w(1 - i_2/\ell)B_4(i_2, j_2) \\ f(i,j) &= w(j_2/\ell)f_1 + w(1 - j_2/\ell)f_2 \end{aligned} \quad (5)$$

with $w(x) = \cos^2(\pi x/2)$. Of course, one might select any other smooth, nonincreasing function satisfying, $w(0) = 1$, $w(1) = 0$, $w'(0) = 0$ and obeying the symmetry property $w(x) + w(1 - x) = 1$.

It is worth mentioning that the spatial partitioning introduces a redundancy factor equal to 4.

2.4. Digital Curvelet Transform

2.4.1. Definition

The curvelet transform (Donoho and Duncan, 2000; Candès and Donoho, 1999a; Starck et al., 2002), open us the possibility to analyze an image with different block sizes, but with a single transform. The idea is to first decompose the image into a set of wavelet bands, and to analyze each band by a local ridgelet transform. The block size can be changed at each scale level. Roughly speaking, different levels of the multiscale ridgelet pyramid are used to represent different subbands of a filter bank output. At the same time, this subband decomposition imposes a relationship between the width and length of the important frame elements so that they are anisotropic and obey $width = length^2$.

The discrete curvelet transform of a continuum function $f(x_1, x_2)$ makes use of a dyadic sequence of scales, and a bank of filters with the property that the passband filter Δ_s is concentrated near the frequencies $[2^{2s}, 2^{2s+2}]$, e.g.

$$\Delta_s = \Psi_{2s} * f, \quad \widehat{\Psi_{2s}}(\xi) = \widehat{\Psi}(2^{-2s}\xi).$$

In wavelet theory, one uses a decomposition into dyadic subbands $[2^s, 2^{s+1}]$. In contrast, the subbands used in the discrete curvelet transform of continuum functions have the nonstandard form $[2^{2s}, 2^{2s+2}]$. This is nonstandard feature of the discrete curvelet transform well worth remembering.

The curvelet decomposition is the sequence of the following steps:

– *Subband Decomposition.* The object f is decomposed into subbands.
– *Smooth Partitioning.* Each subband is smoothly windowed into "squares" of an appropriate scale (of sidelength $\sim 2^{-s}$).
– *Ridgelet Analysis.* Each square is analyzed via the discrete ridgelet transform.

In this definition, the two dyadic subbands $[2^{2s}, 2^{2s+1}]$ and $[2^{2s+1}, 2^{2s+2}]$ are merged before applying the ridgelet transform.

2.4.2. Digital Realization

It seems that the "à trous" subband filtering algorithm is especially well-adapted to the needs of the digital curvelet

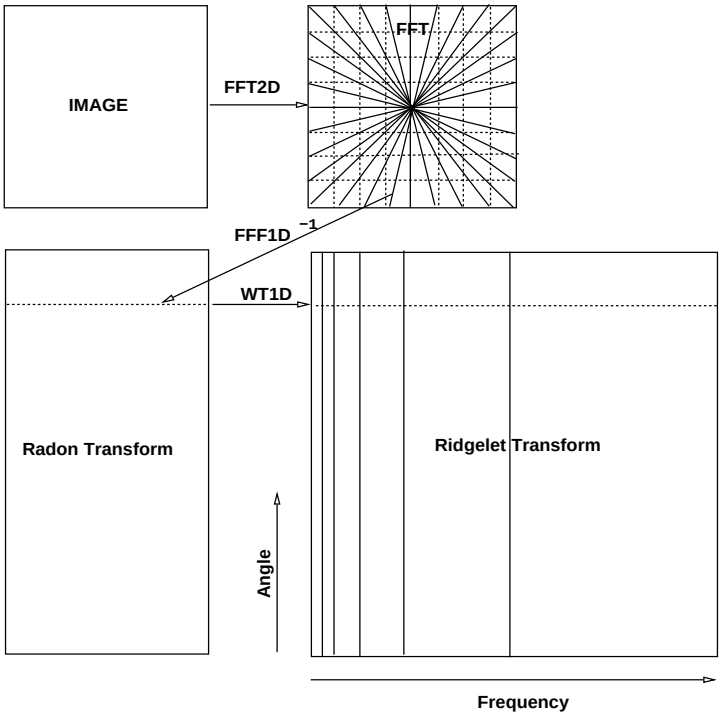

Fig. 4. Ridgelet transform flowgraph. Each of the $2n$ radial lines in the Fourier domain is processed separately. The 1-D inverse FFT is calculated along each radial line followed by a 1-D nonorthogonal wavelet transform. In practice, the one-dimensional wavelet coefficients are directly calculated in the Fourier space.

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 (Starck et al., 1998; Starck and Murtagh, 2002) 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).)

A sketch of the discrete curvelet transform algorithm is:

1. apply the à trous algorithm with J scales,
2. set $B_1 = B_{min}$,
3. for $j = 1, \ldots, J$ do,
   - partition the subband w_j with a block size B_j and apply the digital ridgelet transform to each block,
   - if j modulo $2 = 1$ then $B_{j+1} = 2B_j$,
   - else $B_{j+1} = B_j$.

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 $[2^j, 2^{j+1}]$. Note also that the coarse description of the image c_J is not processed. We used the default value $B_{min} = 16$ pixels in our implementation. Finally, Figure 5 gives an overview of the organization of the algorithm.

This implementation of the curvelet transform is also redundant. The redundancy factor is equal to $16J + 1$ whenever J scales are employed. Finally, the method enjoys exact reconstruction and stability, because these invertibility holds for each element of the processing chain.

3. Filtering

We now apply our digital transforms for removing noise from image data. The methodology is standard and is outlined mainly for the sake of clarity and self-containedness.

Suppose that one is given noisy data of the form

$$x_{i,j} = f(i,j) + \sigma z_{i,j},$$

where f is the image to be recovered and z is white noise, i.e. $z_{i,j} \overset{i.i.d.}{\sim} N(0,1)$. Unlike FFT's or FWT's, our discrete ridgelet (resp. curvelet) transform is not norm-preserving and, therefore, the variance of the noisy ridgelet (resp. curvelet) coefficients will depend on the ridgelet (resp. curvelet) index λ. For instance, letting F denote the discrete curvelet transform matrix, we have $Fz \overset{i.i.d.}{\sim} N(0, FF^T)$. Because the computation of FF^T is prohibitively expensive, we calculated an approximate value $\tilde{\sigma}_\lambda^2$ of the individual variances using Monte-Carlo simulations where the diagonal elements of FF^T are simply estimated by evaluating the curvelet transforms of a few standard white noise images.

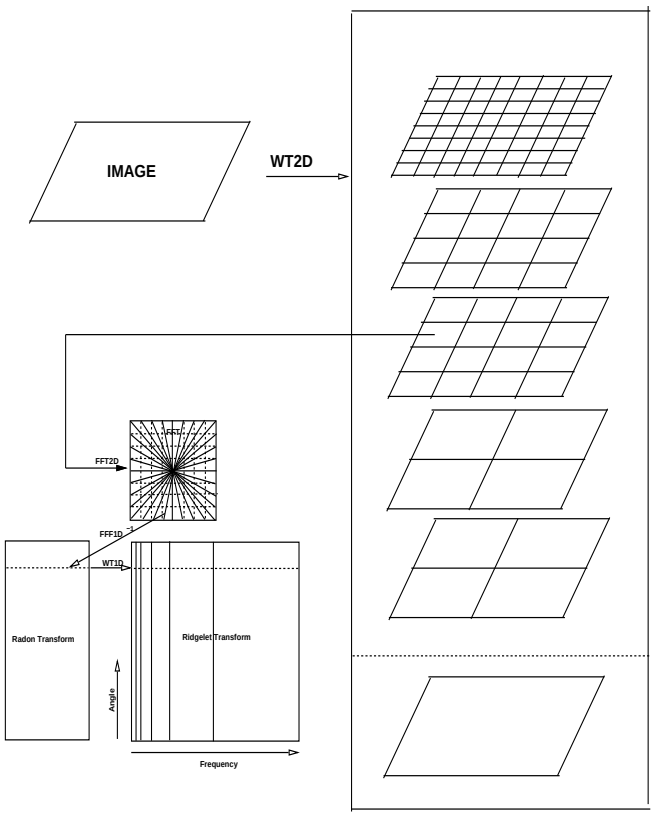

Fig. 5. Curvelet transform flowgraph. The figure illustrates the decomposition of the original image into subbands followed by the spatial partitioning of each subband. The ridgelet transform is then applied to each block.

Let y_λ be the noisy curvelet coefficients ($y = Fx$). We use the following hard-thresholding rule for estimating the unknown curvelet coefficients:

$$\hat{y}_\lambda = y_\lambda \quad \text{if} \quad |y_\lambda|/\sigma \geq k\tilde{\sigma}_\lambda \tag{6}$$

$$\hat{y}_\lambda = 0 \quad \text{if} \quad |y_\lambda|/\sigma < k\tilde{\sigma}_\lambda. \tag{7}$$

In our experiments, we actually chose a scale-dependent value for k; we have $k = 4$ for the first scale ($j = 1$) while $k = 3$ for the others ($j > 1$).

Poisson Noise

Assume now that we have Poisson data $x_{i,j}$ with unknown mean $f(i, j)$. The Anscombe transformation (Anscombe, 1948)

$$\tilde{x} = 2\sqrt{x + \frac{3}{8}} \tag{8}$$

stabilizes the variance and we have $\tilde{x} = 2\sqrt{f} + \epsilon$ where ϵ is a vector with independent and approximately standard normal components. In practice, this is a good approximation whenever the number of counts is large enough, greater than 30 per pixel, say.

For small number of counts, a possibility is to compute the Radon transform of the image, and then to apply the Anscombe transformation to the Radon data. The rationale being that, roughly speaking, the Radon transform corresponds to a summation of pixel values over lines and that the sum of independent Poisson random variables is a Poisson random variable with intensity equal to the sum of the individual intensities. Hence, the intensity of the sum may be quite large (hence validating the Gaussian approximation) even though the individual intensities may be small. This might be viewed as an interesting feature as unlike wavelet transforms, the ridgelet and curvelet transforms tend to average data over elongated and rather large neighborhoods.

Gaussian and Poisson Noise

The arrival of photons, and their expression by electron counts, on CCD detectors may be modeled by a Poisson distribution. In addition, there is additive Gaussian readout noise. The Anscombe transformation (eqn. 8) has been extended to take this combined noise into account. As an approximation, consider the signal's value, s_k, as a sum of a Gaussian variable, γ, of mean g and standard-deviation σ; and a Poisson variable, n, of mean m_0: we set $x = \gamma + \alpha n$ where α is the gain. The generalization of the variance stabilizing Anscombe formula is (Murtagh et al., 1995):

$$\tilde{x} == \frac{2}{\alpha}\sqrt{\alpha x + \frac{3}{8}\alpha^2 + \sigma^2 - \alpha g} \tag{9}$$

With appropriate values of α, σ and g, this reduces to Anscombe's transformation.

Then, for an image containing Poisson and Poisson+Gaussian noise, we apply first respectively the Anscombe and the Generalized Anscombe transform.

These variance stabilization transformations, it has been shown in Murtagh et al. (1995), are only valid for a sufficiently large number of counts (and of course, for a larger still number of counts, the Poisson distribution becomes Gaussian). The necessary average number of counts is about 20 if bias is to be avoided. Note that errors related to small values carry the risk of removing real objects, but not of amplifying noise. For Poisson parameter values under this threshold acceptable number of counts, the Anscombe transformation loses control over the bias. In this case, an alternative approach to variance stabilization is needed. It has been shown that the first step of the ridgelet transform consists in a Radon transform. As a Radon coefficient is an addition of pixel values along a line, the Radon transform of an image containing Poisson noise contains also Poisson noise. Then the Anscombe transform can be applied after the Radon transformation rather than on the original image. The advantage is that the number of counts per pixel will obviously be larger in the Radon domain than in the image domain, and the variance stabilization will be more robust.

Experiment

A Gaussian white noise with a standard deviation fixed to 20 was added to the `Saturn` image. We employed several methods to filter the noisy image:

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

Our experiments are reported on Figure 6. 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 is instructive. One notices that the decimated wavelet transform exhibits distortions of the boundaries and suffers substantial loss of important detail. The à trous wavelet transform gives better boundaries, but completely omits to reconstruct certain ridges. 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 7 shows an example of an X-ray image filtering by the ridgelet transform using such an approach. Figure 7 left and right shows respectively the XMM/Newton image of the Kepler SN1604 supernova and the ridgelet filtered image (using a five sigma hard thresholding).

Which transform should be chosen for a given data set ?

We have introduced in this paper two new transforms, the ridgelet transform and the curvelet transform. Several other transforms are often used in astronomy, such the Fourier transform, the isotropic à trous wavelet transform and the bi-orthogonal wavelet transform. The choise of the best transform may be delicate. Each transform has its own domain of optimality:

- The Fourier transform for stationary process.
- The à trous wavelet transform for isotropic features.
- The bi-orthogonal wavelet transform for features with a small anisotropy, typically with a width equals to half the length.
- The ridgelet wavelet transform for anisotropic features with a given length (i.e. block size).
- The curvelet transform for anisotropic features with different length and width equals to the square of the length.

Section 5 will show how several transforms can be used simultaneously, in order to benefit of the advantages of each of them.

4. Contrast Enhancement

Because some features are hardly detectable by eye in an image, we often transform it before display. Histogram equalization is one the most well-known methods for contrast enhancement. Such an approach is generally useful for images with a poor intensity distribution. Since edges play a fundamental role in image understanding, a way to enhance the contrast is to enhance the edges. For example, we can add to the original image its Laplacian ($I^{'} = I + \gamma \Delta I$, where γ is a parameter). Only features at the finest scale are enhanced (linearly). For a high γ value, only the high frequencies are visible.

Since the curvelet transform is well-adapted to represent images containing edges, it is a good candidate for edge enhancement. Curvelet coefficients can be modified in order to enhance edges in an image. The idea is to not modify curvelet coefficients which are either at the noise level, in order to not amplify the noise, or larger than a given threshold. Largest coefficients corresponds to strong edges which do not need to be amplified. Therefore, only curvelets coefficients with an absolute value in $[T_{min}, T_{max}]$ are modified, where T_{min} and T_{max} must be fixed. We define the following function y_c which modifies the values of the curvelet coefficients:

$$\begin{aligned} y_c(x) &= 1 \text{ if } x < T_{min} \\ y_c(x) &= \frac{x - T_{min}}{T_{min}} (\frac{T_{max}}{T_{min}})^p + \frac{2T_{min} - x}{T_{min}} \text{ if } x < 2T_{min} \\ y_c(x) &= (\frac{T_{max}}{x})^p \text{ if } 2T_{min} \leq x < T_{max} \\ y_c(x) &= 1 \text{ if } x \geq T_{max} \end{aligned} \quad (10)$$

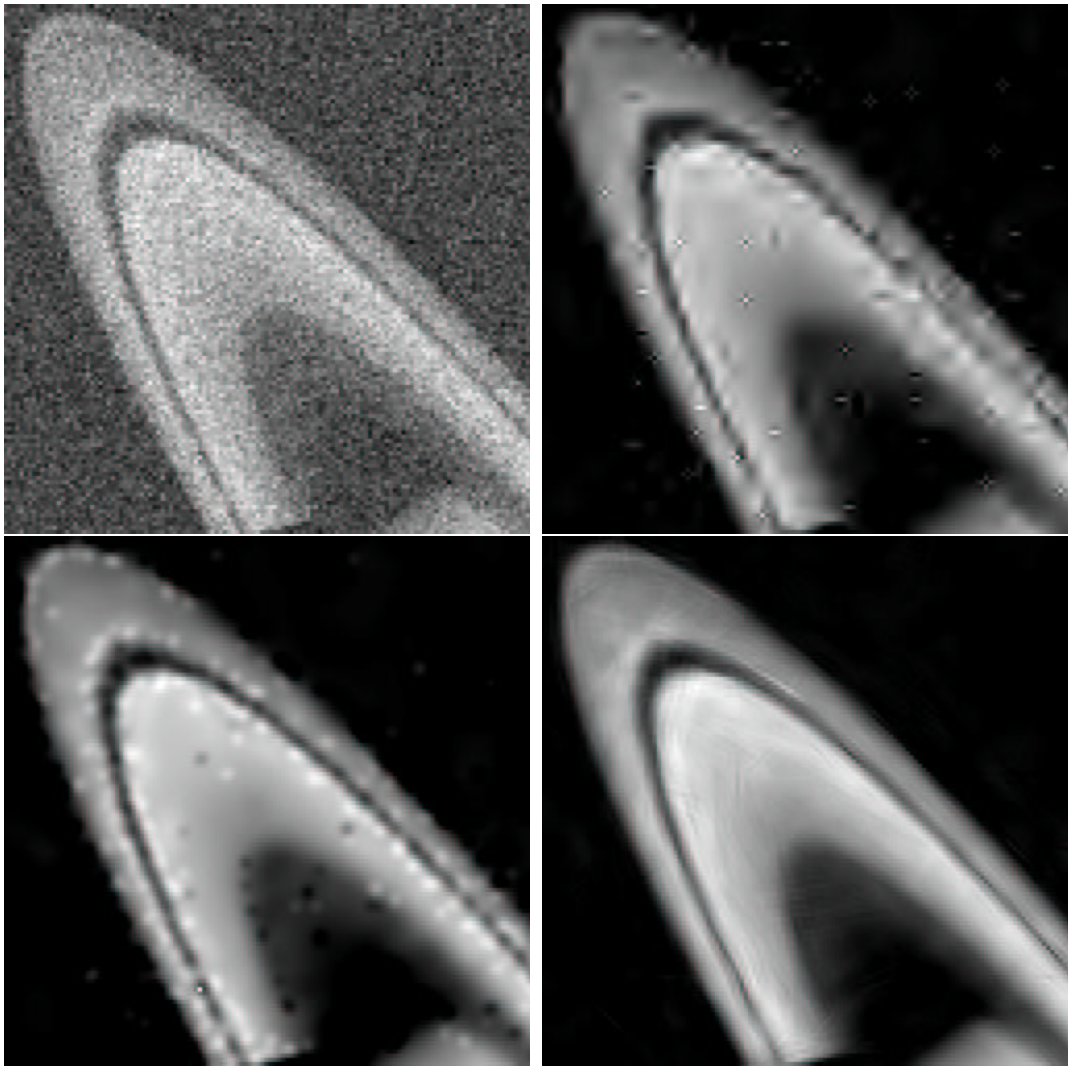

Fig. 6. Top left, part of Saturn image with a Gaussian noise. Top right, filtered image using the undecimated bi-orthogonal wavelet transform. Bottom left and right, filtered image by the à trous wavelet transform algorithm and the curvelet transform.

p determines the degree of non-linearity. T_{min} is derived from the noise level, $T_{min} = c\sigma$. A c value larger than 3 guaranties that the noise will not be amplified. The T_{max} parameter can be defined either from the noise standard deviation ($T_{max} = K_m\sigma$) or from the maximum curvelet coefficient M_c of the relative band ($T_{max} = lM_c$, with $l < 1$). The first choice allows the user to define the coefficients to amplify as a function of their signal to noise ratio, while the second one gives an easy and general way to fix the T_{max} parameter independently of the range of the pixel values. Figure 8 shows the curve representing the enhanced coefficients versus the original coefficients.

The curvelet enhancement method consists of the following steps:

1. Estimate the noise standard deviation σ in the input image I.
2. Calculate the curvelet transform of the input image. We get a set of bands w_j, each band w_j contains N_j coefficients and corresponds to a given resolution level.
3. Calculate the noise standard deviation σ_j for each band j of the curvelet transform (see (Starck et al., 2002) more details on this step).
4. For each band j do
   - Calculate the maximum M_j of the band.
   - Multiply each curvelet coefficient $w_{j,k}$ by $y_c(\mid w_{j,k} \mid)$.
5. Reconstruct the enhanced image from the modified curvelet coefficients.

Example: Saturn Image

Figures 9 shows respectively a part of the Saturn image, the histogram equalized image, the Laplacian enhanced image and the curvelet multiscale edge enhanced image (parameters were $p = 0.5$, $c = 3$, and $l = 0.5$). The curvelet multiscale edge enhanced image shows clearly better the rings and edges of Saturn.

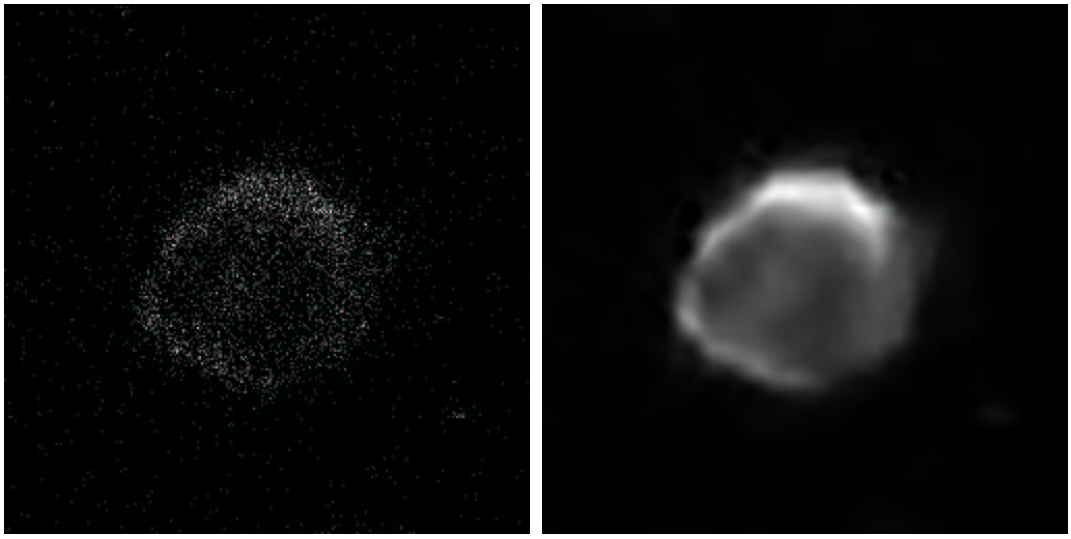

Fig. 7. Left, XMM/Newton image of the Kepler SN1604 supernova. Right, ridgelet filtered image.

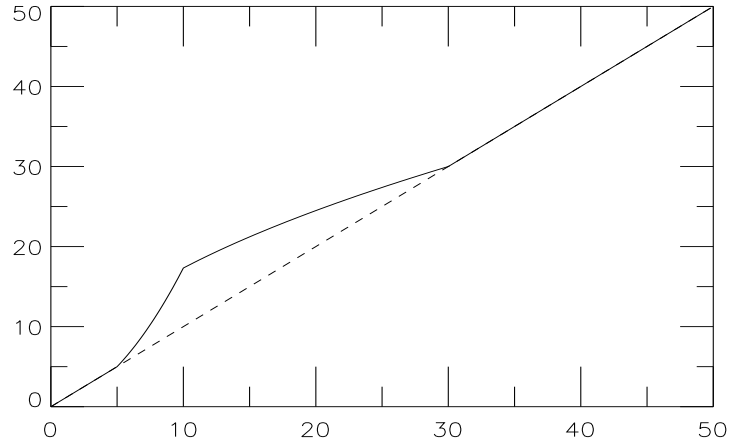

Fig. 8. Enhanced coefficients versus original coefficients. Parameters are T_{max} =30,c=5 and p=0.5.

5. Morphological Component Analysis

5.1. Introduction

The content of an image is often complex, and there is not a single transform which is optimal to represent all the contained features. For example, the Fourier transform better represents some textures, while the wavelet transform better represents singularities. Even if we limit our class of transforms to the wavelet one, decision have to be taken between an isotropic wavelet transform which produce good results for isotropic objects (such stars and galaxies in astronomical images, cells in biological images, etc), or an orthogonal wavelet transform, which is better for images with edges. This has motivated the development of different methods (Chen et al., 1998; Meyer et al., 1998; Huo, 1999), and the two most frequently discussed approaches are the Matching Pursuit (MP) (Mallat and Zhang, 1993) and the Basis pursuit (BP) (Chen et al., 1998). A dictionary $\mathcal{D}$ being defined as a collection of waveforms $(\varphi_\gamma)_{\gamma\in\Gamma}$, the general principe consists in representing a signal s as a "sparse" linear combination of a small number of basis such that:

$$s = \sum_\gamma a_\gamma \varphi_\gamma \tag{11}$$

or an approximate decomposition

$$s = \sum_{i=1}^{m} a_{\gamma_i} \varphi_{\gamma_i} + R^{(m)}. \tag{12}$$

Matching pursuit (Mallat and Zhang, 1993; Mallat, 1998) method (MP) uses a greedy algorithm which adaptively refines the signal approximation with an iterative procedure:

- Set $s^0 = 0$ and $R^0 = 0$.
- Find the element $\alpha_k \varphi_{\gamma_k}$ which best correlates with the residual.
- Update s and R:

$$\begin{aligned} s^{k+1} &= s^k + \alpha_k \varphi_{\gamma_k} \\ R^{k+1} &= s - s^k. \end{aligned} \tag{13}$$

Fig. 9. Top, Saturn image and its histogram equalization. Bottom, Saturn image enhancement the Laplacian method and by the curvelet transform.

In case of non orthogonal dictionaries, it has been shown (Chen et al., 1998) that MP may spend most of the time correcting mistakes made in the first few terms, and therefore is suboptimal in term of sparsity.

Basis pursuit method (Chen et al., 1998) (BP) is a global procedure which synthesizes an approximation $\tilde{s}$ to s by minimizing a functional of the type

$$\|s - \tilde{s}\|_{\ell_2}^2 + \lambda \cdot \|\alpha\|_{\ell_1}, \quad \tilde{s} = \Phi\alpha. \tag{14}$$

Between all possible solutions, the chosen one has the minimum l^1 norm. This choice of l_1 norm is very important. A l_2 norm, as used in the method of frames (Daubechies, 1988), does not preserve the sparsity (Chen et al., 1998).

In many cases, BP or MP synthesis algorithms are computationally very expensive. We present in the following an alternative approach, that we call *Combined Transforms Method* (CTM), which combines the different available transforms in order to benefit of the advantages of each of them.

5.2. The Combined Transformation

Depending on the content of the data, several transforms can be combined in order to get an optimal representation of all features contained in our data set. In addition to the ridgelet and the curvelet transform, we may want to use the à trous algorithm which is very well suited to astronomical data, or the undecimated wavelet transform which is commonly used in the signal processing domain.

Other transform such wavelet packets, the Fourier transform, the Pyramidal median transform (Starck et al., 1998), or other multiscale morphological transforms, could also be considered. However, we found that in practice, these four transforms (i.e. curvelet, ridgelet, à trous algorithm, and undecimated wavelet transform) furnishes a very large panel of waveforms which is generally large enough to well represents all features contained in the data.

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.

Therefore, we search a vector $\alpha = \alpha_1, \ldots, \alpha_K$ such that

$$s = \Phi\alpha \tag{15}$$

where $\Phi\alpha = \sum_{k=1}^{K} T_k^{-1}\alpha_k$. As our dictionary is overcomplete, there is an infinity of vectors verifing this condition, and we need to solve the following optimization problem:

$$min \parallel s - \phi\alpha \parallel^2 + \mathcal{C}(\alpha) \tag{16}$$

where $\mathcal{C}$ is a penalty term. We easily see that chosing $\mathcal{C}(\alpha) = \parallel \alpha \parallel_{l_1}$ leads to the BP method, where the dictionary $\mathcal{D}$ is only composed of the basis elements of the chosen transforms.

Two iterative methods, *soft-CTM* and *hard-CTM*, allowing us to realize such a combined transform, are described in this section.

5.3. Soft-CTM

Noting $T_1, ..., T_K$ the K transform operators, a solution α is obtained by minimizing a functional of the form:

$$J(\alpha) = \parallel s - \sum_{k=1}^{K} T_k^{-1}\alpha_k \parallel_2^2 + \lambda \sum_k \parallel \alpha_k \parallel_1 \tag{17}$$

where s is the original signal, and α_k are the coefficients obtained with the transform T_k.

An simple algorithm to achieve such an solution is:

1. Initialize $L_{\max}$, the number of iterations N_i, $\lambda = L_{\max}$, and $\delta_\lambda = \frac{L_{\max}}{N_i}$.
2. While $\lambda >= 0$ do
3. For k = 1, .., K do
   - Calculate the residual $R = s - \sum_k T_k^{-1}\alpha_k$.
   - Calculate the transform T_k of the residual: $r_k = T_k R$.
   - Add the residual to α_k: $\alpha_k = \alpha_k + r_k$.
   - Soft threshold the coefficient α_k with the λ threshold.
4. $\lambda = \lambda - \delta$, and goto 2.

Figure 10 illustrates the result in the case where the input image contains only lines and Gaussians. In this experiment, we have initialized $L_{\max}$ to 20, and δ to 2 (10 iterations). Two transform operators were used, the à trous wavelet transform and the ridgelet transform. The first is well adapted to the detection of Gaussian due to the isotropy of the wavelet function (Starck et al., 1998), while the second is optimal to represent lines (Candès and Donoho, 1999b). Figure 10 top, bottom left, and bottom right represents respectively the original image, the reconstructed image from the à trous wavelet coefficient, and the reconstructed image from the ridgelet coefficient. The addition of both reconstructed images reproduces the original one.

In some specific cases where the data are sparse in all bases, it has been shown (Huo, 1999; Donoho and Huo, 2001) that the solution is identical to the solution when using a $\parallel . \parallel_0$ penalty term. This is however generally not the case. The problem we met in image restoration applications, when minimizing equation 17, is that both the signal and noise are split into the bases. The way the noise is distributed in the coefficients α_k is not known, and leads to the problem that we do not know at which level we should threshold the coefficients. Using the threshold we would have used with a single transform makes a strong over-filtering of the data. Using the l^1 optimization for data restoration implies to first study how the noise is distributed in the coefficients. The hard-CTM method does not present this drawback.

5.4. Hard-CTM

The following algorithm consists in hard thresholding the residual successively on the different bases.

1. For noise filtering, estimate the noise standard deviation σ, and set $L_{\min} = k_\sigma$. Otherwise, set $\sigma = 1$ and $L_{\min} = 0$.
2. Initialize $L_{\max}$, the number of iterations N_i, $\lambda = L_{\max}$ and $\delta_\lambda = \frac{L_{\max} - L_{\min}}{N_i}$.
3. Set all coefficients α_k to 0.
4. While $\lambda >= L_{\min}$ do
5. for k = 1, .., K do
   - Calculate the residual $R = s - \sum_k T_k^{-1}\alpha_k$.
   - Calculate the transform T_k of the residual: $r_k = T_k R$.
   - For all coefficients $\alpha_{k,i}$ do
     - Update the coefficients: if $\alpha_{k,i} \neq 0$ or $\mid r_{k,i} \mid > \lambda\sigma$ then $\alpha_{k,i} = \alpha_{k,i} + r_{k,i}$.
6. $\lambda = \lambda - \delta_\lambda$, and goto 5.

For an exact representation of the data, k_σ must be set to 0. Choosing $k_\sigma > 0$ introduces a filtering. If a single transform is used, it corresponds to the standard k-sigma hard thresholding.

It seems that starting with a high enough $L_{\max}$ and a high number of iterations would lead to the l^0 optimization solution, but this remains to be proved.

5.5. Experiments

5.6. Experiment 1: Infrared Gemini Data

Fig. 11 upper left shows a compact blue galaxy located at 53 Mpc. The data have been obtained on ground with the GEMINI-OSCIR instrument at 10 μm. The pixel field of view is $0.089''$/pix, and the source was observed during 1500s. The data are contaminated by a noise and a stripping artifact due to the instrument electronic. The same kind of artifact pattern were observed with the ISOCAM instrument (Starck et al., 1999).

This image, noted D_{10}, has been decomposed using wavelets, ridgelets, and curvelets. Fig. 11 upper middle,

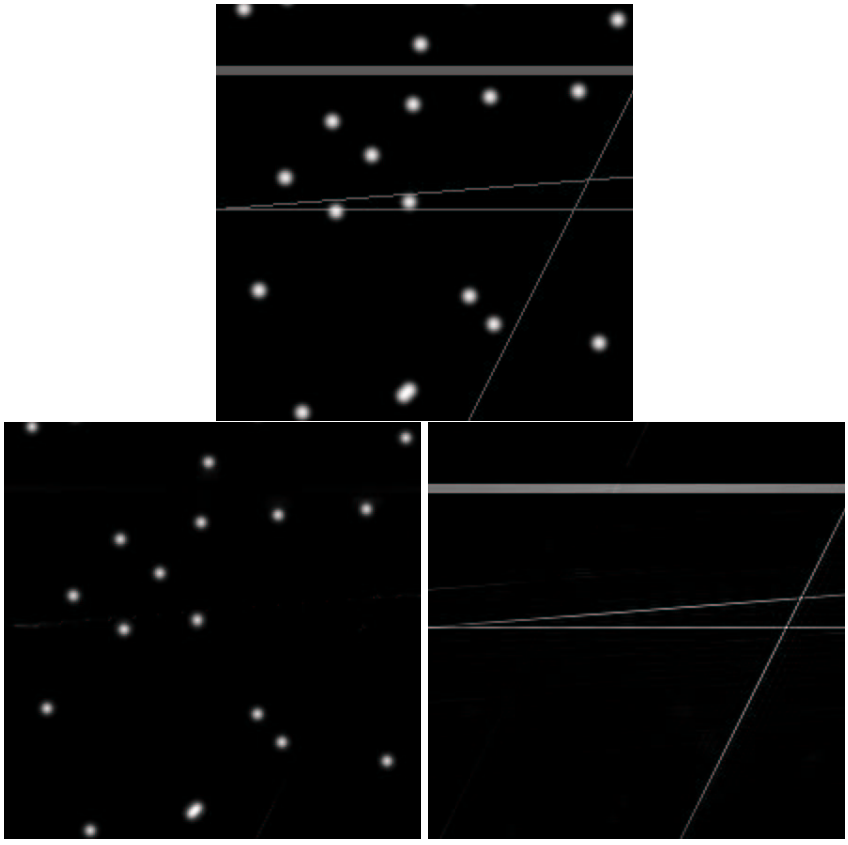

Fig. 10. Top, original image containing lines and gaussians. Botton left, reconstructed image for the à trous wavelet coefficient, bottom right, reconstructed image from the ridgelet coefficients.

upper right, and bottom left show the three images R_{10}, C_{10}, W_{10} reconstructed respectively from the ridgelets, the curvelets, and the wavelets. Image in Fig. 11 bottom middle shows the residual, i.e. $e_{10} = D_{10} - (R_{10} + C_{10} + W_{10})$. Another interesting image is the artifact free one, obtained by subtracting R_{10} and C_{10} from the input data (see Fig. 11 bottom right). The galaxy has well been detected in the wavelet space, while all stripping artifact have been capted by the ridgelets and curvelets.

Fig. 12 upper left shows the same galaxy, but at 20 μm. We have applied the same decomposition on D_{20}. Fig. 12 upper right shows the coadded image $R_{20} + C_{20}$, and we can see bottom left and right the wavelet reconstruction W_{20} and the residudal $e_{20} = D_{20} - (R_{20} + C_{20} + W_{20})$.

5.7. Experiment 2: A370

Figure 13 upper left shows the HST A370 image. It contains many anisotropic features such the gravitationnal arc, and the arclets. The image has been decomposed using three transforms: the ridgelet transform, the curvelet transform, and the à trous wavelet transform. Three images have then been reconstructed from the coefficients of the three basis. Figure 13 upper right shows the coaddition of the ridgelet and curvelet reconstructed images. The à trous reconstructed image is displayed in Figure 13 lower left, and the coaddition of the three images can be seen in Figure 13 lower right. The gravitational arc and the arclets are all represented in the ridgelet and the curvelet basis, while all isotropic features are better represented in the wavelet basis.

5.7.1. Elongated - point like object separation in astronomical images.

Figure 14 shows the result of a decomposition of a spiral galaxy (NGC2997). This image (figure 14 top left) contains many compact structures (stars and HII region), more or less isotropic, and large scale elongated features (NGC2997 spiral part). Compact objects are well represented by isotropic wavelets, and the elongated features are better represented by a ridgelet basis. In order to benefit of the optimal data representation of both transforms, the image has been decomposed on both the à trous wavelet transform and on the ridgelet transform by using the same method as described in section 5.4. When the functional is minimized, we get two images, and their coaddition is the filtered version of the original image. The reconstructions from the à trous coefficients, and from the ridgelet coefficients can be seen in figure 14 top right and bottom left. The addition of both images is presented in figure 14 bottom right.

We can see that this Morphological Component Analysis (MGA) allows us to separate automatically

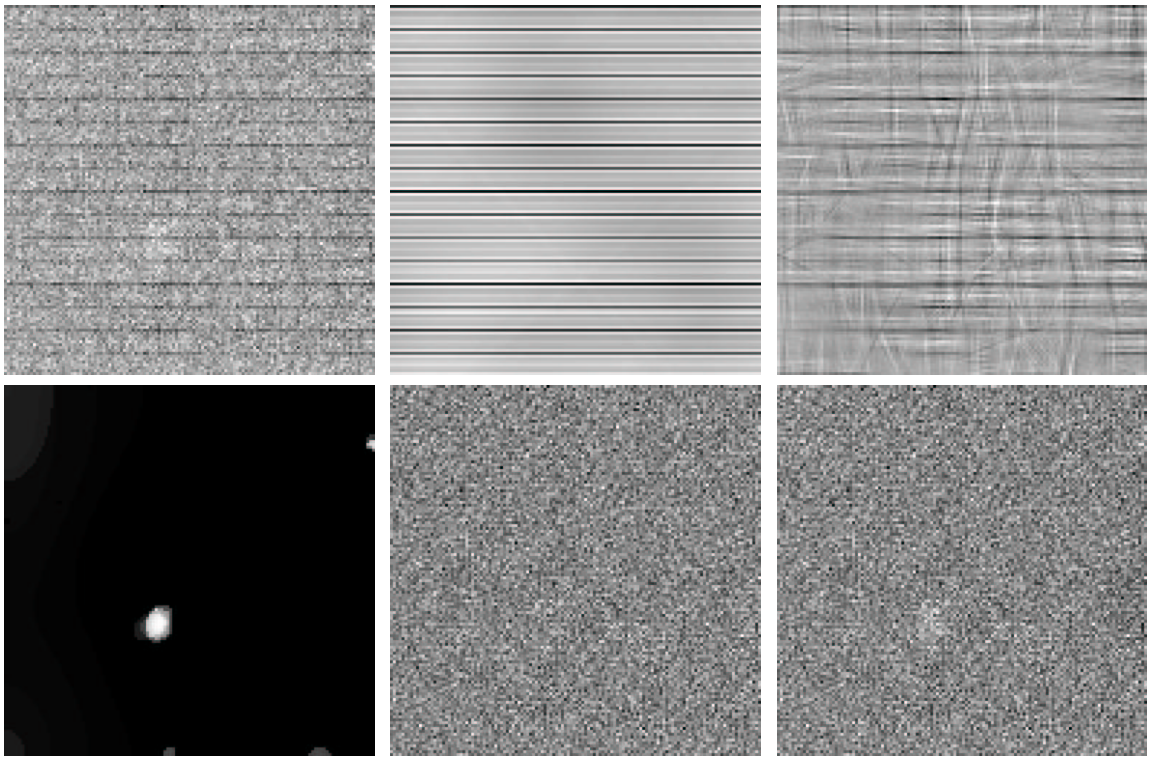

Fig. 11. Upper left, galaxy SBS 0335-052 (10 μm), upper middle, upper right, and bottom left, reconstruction respectively from the ridgelet, the curvelet and wavelet coefficients. Bottom middle, residual image. Bottom right, artifact free image.

features in an image which have different morphological aspects. It is very different from other techniques such as Principal Component Analysis or Independent Component Analysis (Cardoso, 1998) where the separation is performed via statistical properties.

Acknowledgments

We are grateful to the referee for helpful comments on an earlier version.

References

Anscombe, F.: 1948, *Biometrika* **15**, 246

Candès, E.: 1998, *Ph.D. thesis*, Stanford University

Candès, E. and Donoho, D.: 1999a, *Curvelets*, Technical report, Statistics, Stanford University

Candès, E. and Donoho, D.: 1999b, *Philosophical Transactions of the Royal Society of London A* **357**, 2495

Candès, E. J.: 1999, *Applied and Computational Harmonic Analysis* **6**, 197

Candès, E. J. and Donoho, D. L.: 1999c, in A. Cohen, C. Rabut, and L. Schumaker (eds.), *Curve and Surface Fitting: Saint-Malo 1999*, Vanderbilt University Press, Nashville, TN

Candès, E. J. and Donoho, D. L.: 1999d, *Philosophical Transactions of the Royal Society of London A* **357**, 2495

Cardoso, J.: 1998, *Proceedings of the IEEE* **86**, 2009

Chen, S., Donoho, D., and Saunder, M.: 1998, *SIAM Journal on Scientific Computing* **20**, 33

Coifman, R. and Donoho, D.: 1995, in A. Antoniadis and G. Oppenheim (eds.), *Wavelets and Statistics*, pp 125–150, Springer-Verlag

Daubechies, I.: 1988, *IEEE Transactions on Information Theory* **34**, 605

Donoho, D. and Duncan, M.: 2000, in H. Szu, M. Vetterli, W. Campbell, and J. Buss (eds.), *Proc. Aerosense 2000, Wavelet Applications VII*, Vol. 4056, pp 12–29, SPIE

Donoho, D. and Huo, X.: 2001, *IEEE Transactions on Information Theory* 47(7)

Donoho, D. L.: 1997, *Fast Ridgelet Transforms in Dimension 2*, Technical report, Stanford University, Department of Statistics, Stanford CA 94305–4065

Donoho, D. L.: 1998, *Digital Ridgelet Transform via RectoPolar Coordinate Transform*, Technical report, Stanford University

Huo, X.: 1999, *Ph.D. thesis*, Stanford Univesity

Mallat, S.: 1998, *A Wavelet Tour of Signal Processing*, Academic Press

Mallat, S. and Zhang, Z.: 1993, *IEEE Transactions on Signal Processing* **41**, 3397

Meyer, F., Averbuch, A., Stromberg, J.-O., and Coifman, R.: 1998, in *International Conference on Image Processing, ICIP'98, Chicago*

Murtagh, F., Starck, J.-L., and Bijaoui, A.: 1995, *Astronomy and Astrophysics, Supplement Series* **112**, 179

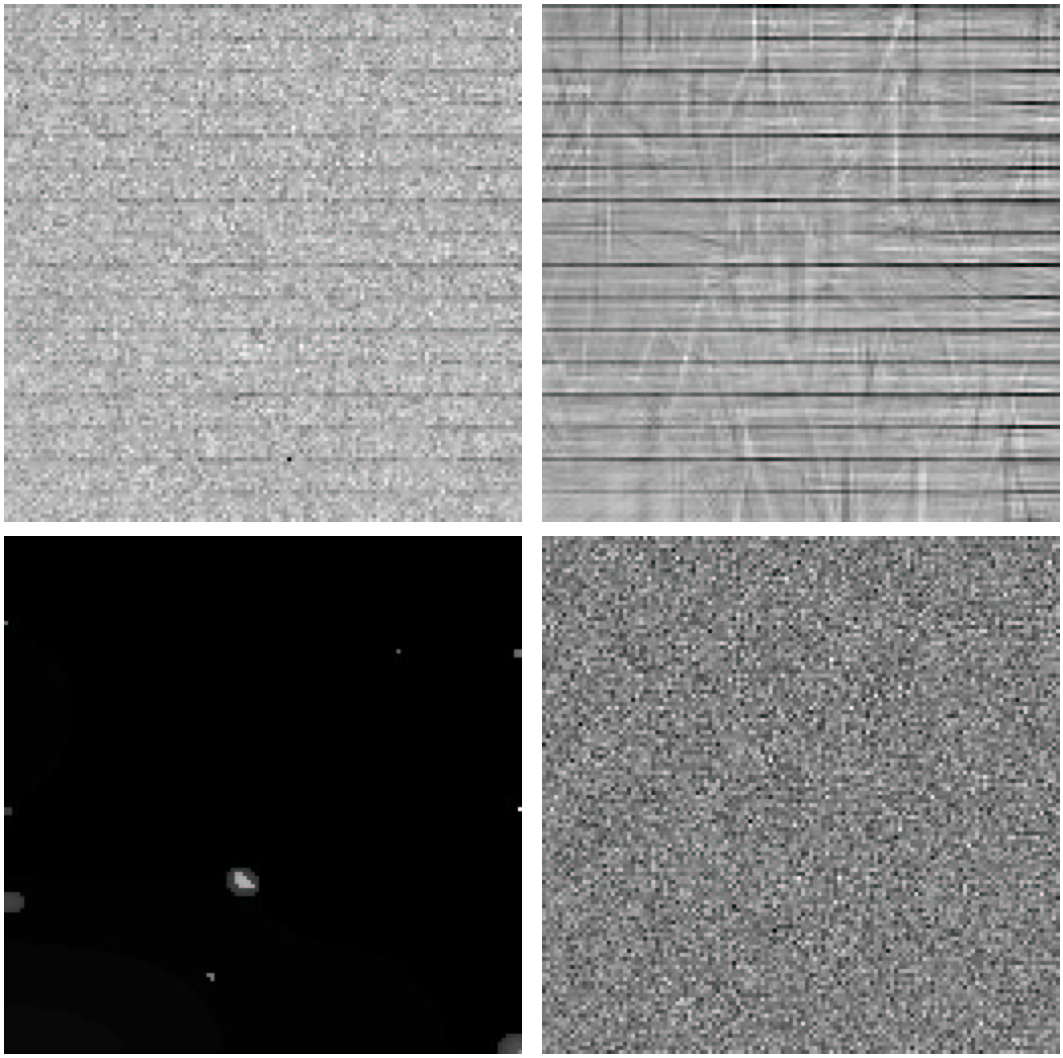

Fig. 12. Upper left, galaxy SBS 0335-052 (20 μm), upper right, addition of the reconstructed images from both the ridgelet and the curvelet coefficients, bottom left, reconstruction from the wavelet coefficients, and bottom right, residual image.

Simoncelli, E., Freeman, W., Adelson, E., and Heeger, D.: 1992, *IEEE Trans. Information Theory*

Starck, J.-L., Abergel, A., Aussel, H., Sauvage, M., Gastaud, R., Claret, A., Desert, X., Delattre, C., and Pantin, E.: 1999, *Astronomy and Astrophysics, Supplement Series* **134**, 135

Starck, J.-L., Bijaoui, A., Lopez, B., and Perrier, C.: 1994, *Astronomy and Astrophysics* **283**, 349

Starck, J.-L., Candès, E., and Donoho, D.: 2002, *IEEE Transactions on Image Processing* **11(6)**, 131

Starck, J.-L. and Murtagh, F.: 2002, *Astronomical Image and Data Analysis*, Springer-Verlag

Starck, J.-L., Murtagh, F., and Bijaoui, A.: 1998, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge University Press

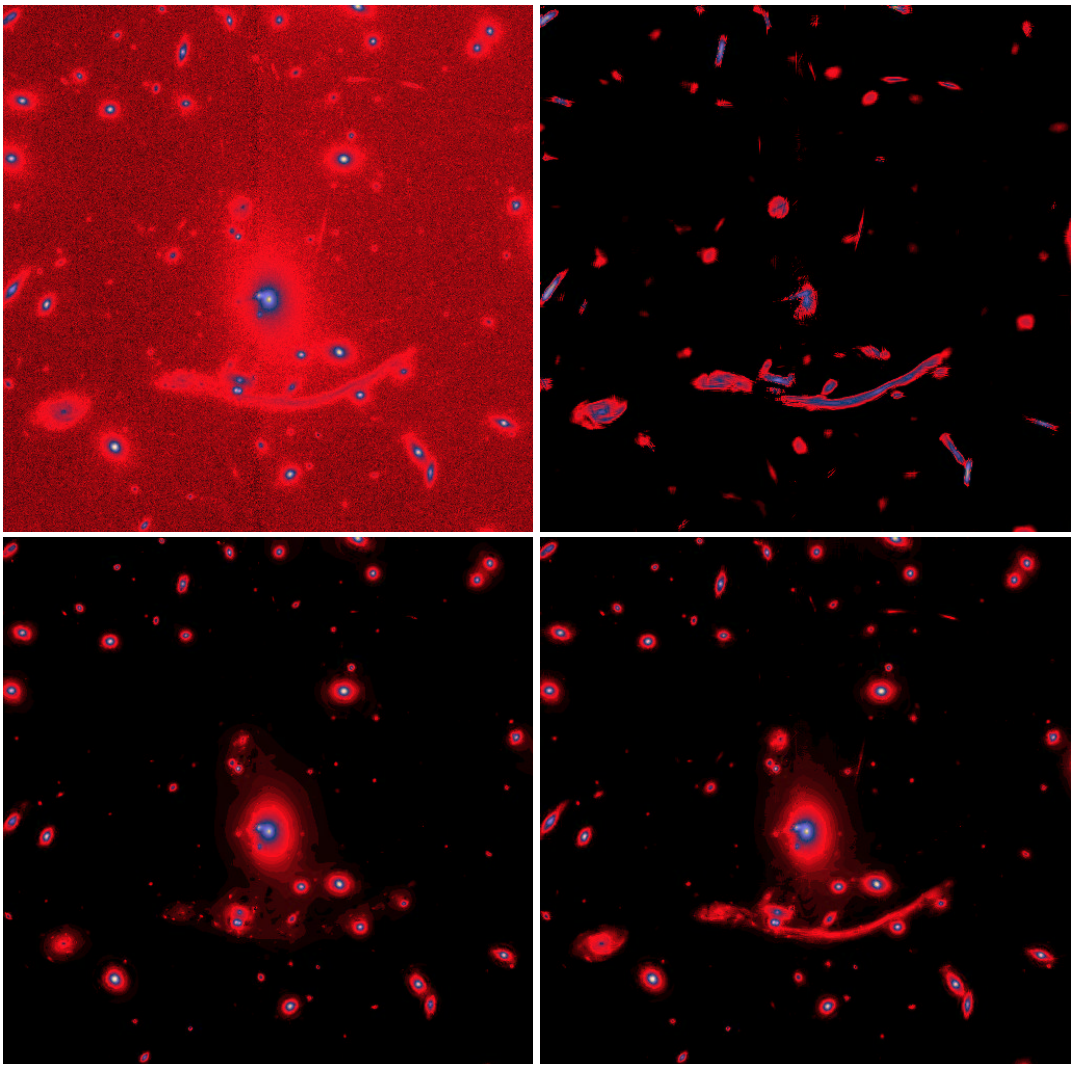

Fig. 13. Top left, HST image of A370, top right coadded image from the reconstructions from the ridgelet and the curvelet coefficients, bottom left reconstruction from the à trous wavelet coefficients, and bottom right addition of the three reconstructed images.

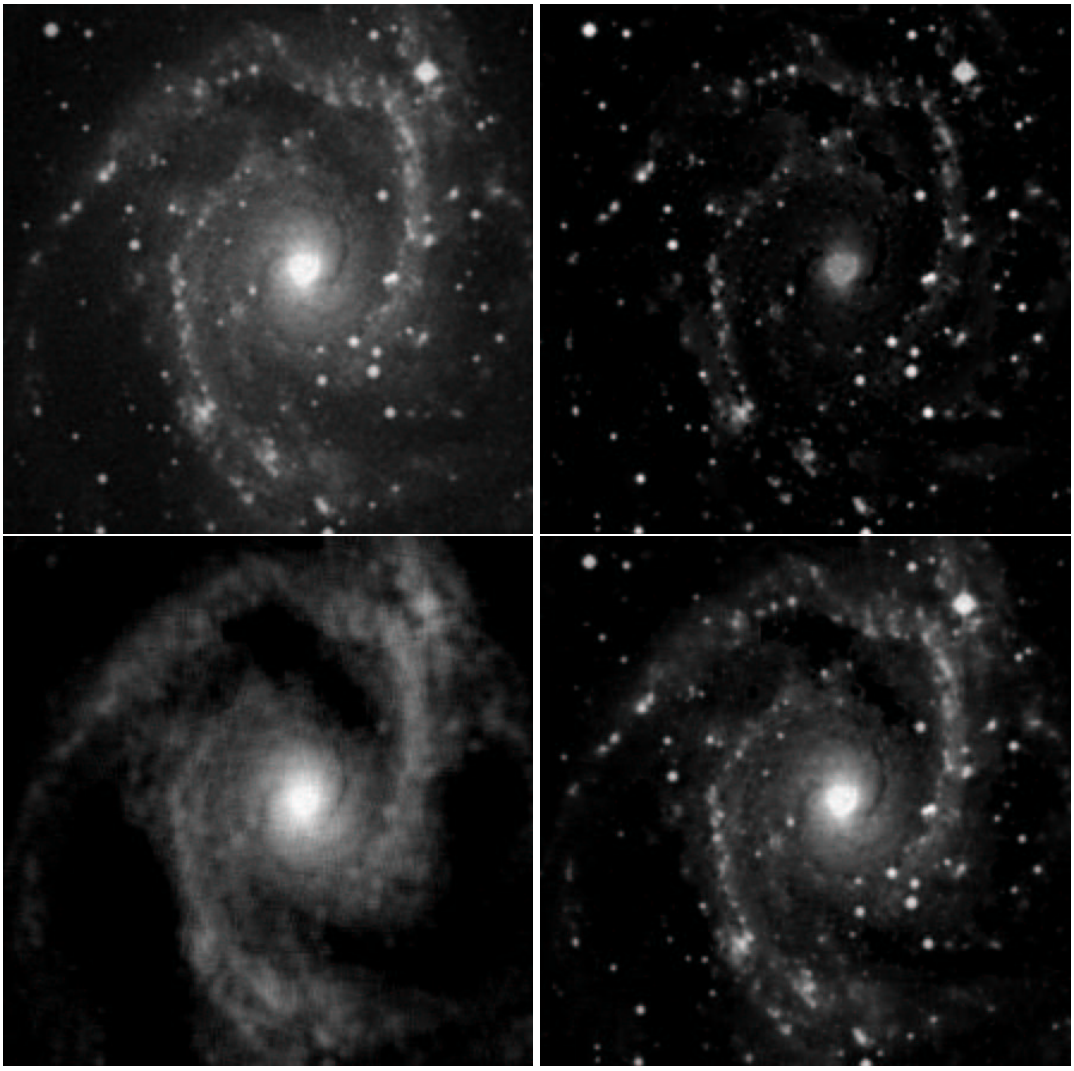

Fig. 14. Top left, galaxy NGC2997, top right reconstructed image from the à trous wavelet coefficients, bottom left, reconstruction from the ridgelet coefficients, and bottom right addition of both reconstructed images.