Image Processing through Multiscale Analysis and Measurement Noise Modeling

F. Murtagh (1) and J.-L. Starck (2)

(1) Faculty of Informatics, University of Ulster

BT48 7JL Londonderry, Northern Ireland

(2) DAPNIA/SEI-SAP, CEA-Saclay, F-91191 Gif-sur-Yvette Cedex, France

Abstract

We describe a range of powerful multiscale analysis methods. We also focus on the pivotal issue of measurement noise in the physical sciences. From multiscale analysis and noise modeling, we develop a comprehensive methodology for data analysis of 2D images, 1D signals (or spectra), and point pattern data. Noise modeling is based on the following: (i) multiscale transforms, including wavelet transforms; (ii) a data structure termed the multiresolution support; and (iii) multiple scale significance testing. The latter two aspects serve to characterize signal with respect to noise. The data analysis objectives we deal with include noise filtering and scale decomposition for visualization or feature detection.

1 Introduction

We will describe a recently-proposed methodology for data analysis in astronomy, medicine, Earth observation and other fields. We will survey results obtained. This methodology takes account of measurement noise in our input data. Noise models include Gaussian or Poisson distributions or mixtures of the two, and also include additive or multiplicative cases. Our methodology also assumes that the data embody characteristics which are on different resolution levels or scales, and therefore that analysis will be improved if these are taken into account.

The remainder of section 1 discusses various aspects of noise. Section 2 describes the wavelet transform mostly used by us, and compares it in some detail with alternative transform methods. This section then describes a data structure derived from the wavelet or some other multiscale transform, which is defined on the basis of noise modeling. In fact, as shown later in section 3, this data structure, the multiresolution support, can be iteratively used to get a better estimate of the noise, which in turn is used to then get a better estimate of the multiresolution support.

Section 3 discusses the treatment of noise in greater depth. This includes variance stabilization and other mechanisms for helping us more easily use the noise, once it has been defined.

Section 4 provides a number of examples of this innovative methodology.

1.1 Scientific Instruments and Detectors

Data in the physical sciences are characterized by the all-pervasive presence of noise, and often knowledge is available of the detector's and data's noise properties, at least approximately. We take as our point of departure image processing in astronomy. It is usual to distinguish between the *signal*, of substantive value to the analyst, and *noise* or clutter. The data signal can be a 2D image, a 1D time-series or spectrum, a 3D data cube, and variants of these.

Noise is a necessary evil in astronomical image processing. If we can reliably estimate noise, through knowledge of instrument properties or otherwise, subsequent analyses would be very much better behaved. In fact, major problems would disappear if this were the case – e.g. image restoration or sharpening based on solving inverse equations could become simpler.

One perspective on the theme of this paper is that we present a coherent and integrated algorithmic framework for a wide range of methods which may well have been developed elsewhere on pragmatic and heuristic grounds. We put such algorithms on a firm footing, through explicit noise modeling followed by computational strategies which benefit from knowledge of the data. The advantages are clear: they include objectivity of treatment; better quality data analysis due to far greater thoroughness; and possibilities for automation of otherwise manual or interactive procedures.

Noise is often taken as additive Poisson (related to arrival of photons) and/or Gaussian. Commonly used electronic CCD (charge-coupled device) detectors have a range of Poisson noise components, together with Gaussian readout noise (Snyder et al., 1993).

Digitized photographic images were found by Tekalp and Pavlović (1991) to be also additive Poisson and Gaussian (and subject to nonlinear distortions which we will not discuss here).

1.2 Automatic Estimation of Noise

The noise associated with a particular detector may be known in advance. In practice rule-of-thumb calculation of noise is often carried out. For instance, limited convex regions of what is considered as background are sampled, and the noise is determined in these regions. For common noise distributions, noise is classically specified by its variance.

There are different ways to more formally estimate the standard deviation of Gaussian noise in an image. Olsen (1993) carried out an evaluation of six methods and showed that the best was the average method, which is the simplest also. This method consists of filtering the data I with the average filter (filtering with a simple box function) and subtracting the filtered image from I. Then a measure of the noise at each pixel is computed. To keep image edges from contributing to the estimate, the noise measure is disregarded if the magnitude of the intensity gradient is larger than some threshold, T.

A new approach to automatic estimation of noise, which improves on the methods described by Olsen, is given in section 3.3 below. It uses a multiscale transform and the multiresolution support data structure.

1.3 Working on Signal versus Working on Noise

In various studies of image and spectral filtering, and image compression (see Starck et al., 1995; Starck et al., 1996), the basis of the methodology lay in determining noise, and separating signal from noise. Signal is what we term the scientifically interesting part of the data. Signal is often very compressible, whereas noise by definition is not compressible. Effective separation of signal and noise is evidently of great importance in the physical sciences.

As has been pointed out, our initial focus is on accurate determination of the noise. Other types of signal modeling, e.g. distribution mixture modeling or density estimation, are more easily carried out subsequently. We contend that the noise modeling is a desirable, and in many cases necessary, preliminary to such signal modeling.

2 The à trous and Other Wavelet Transforms

2.1 The à trous Wavelet Transform

. Furthermore, using a wavelet defined as the difference between the scaling functions of two successive scales ($\frac{1}{2}\psi(\frac{x}{2}) = \phi(x) - \phi(\frac{x}{2})$), the original image c_0 can be expressed as the sum of all the wavelet scales and the smoothed array c_p

$$c_0 = c_p + \sum_{j=1}^{p} w_j \tag{1}$$

and a pixel at position x, y can be expressed also as the sum of all the wavelet coefficients at this position, plus the smoothed array:

$$c_0(x,y) = c_p(x,y) + \sum_{j=1}^{p} w_j(x,y) \tag{2}$$

The wavelet transform of an image produces, at each scale j, a set zero-mean coefficient values $\{w_j\}$. Using an algorithm such as the à trous method (Holschneider et al., 1989; Shensa 1992) this set has the same number of pixels as the image and thus this wavelet transform is a redundant one. Furthermore, using a wavelet defined as the difference between the scaling functions of two successive scales ($\frac{1}{2}\psi(\frac{x}{2}) = \phi(x) - \phi(\frac{x}{2})$), the original image c_0, with a pixel at position x, y, can be expressed as the sum of all the wavelet scales and the smoothed array c_p

$$c_0(x,y) = c_p(x,y) + \sum_{j=1}^{p} w_j(x,y) \tag{3}$$

To simplify notation, let us take one index running over all pixels,

$$c_0(k) = c_p(k) + \sum_{j=1}^{p} w_j(k) \tag{4}$$

A summary of the à trous wavelet transform algorithm is as follows.

1. Initialize i to 0, starting with an image $c_i(k)$. Index k ranges over all pixels.

2. Increment i, and carry out a discrete convolution of the data $c_i(k)$ using a filter h (see Annex A), yielding $c_{i+1}(k)$. The convolution is an interlaced one, where the filter's pixel values have a gap (growing with level, i) between them of 2^{i-1} pixels, giving rise to the name à trous ("with holes"). "Mirroring" is used at the data extremes.

3. From this smoothing we obtain the discrete wavelet transform, $w_i(k) = c_{i-1}(k) - c_i(k)$.

4. If i is less than the number p of resolution levels wanted, return to step 2.

The set $\mathcal{W} = \{w_0, w_1, ..., w_p, c_p\}$, where c_p is a residual, represents the wavelet transform of the data.

The discrete filter h is derived from the scaling function $\phi(x)$ (see Annex A). In our calculations, $\phi(x)$ is a spline of degree 3, which leads (in one dimension) to the the filter h $(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$. A 2D implementation can be based on two 1D sets of (separable) convolutions (Starck and Murtagh, 1994).

The associated wavelet function is of mean zero, of compact support, with a central bump and two negative side-lobes. Of interest for us is that, like the scaling function, it is isotropic (point symmetric).

2.2 Multiscale Transforms Compared to Other Data Transforms

In this section we will discuss in general terms why the wavelet transform has very good noise filtering properties, and how it differs from other data preprocessing transforms in this respect. Among the latter, we can include principal components analysis (PCA) and correspondence analysis, which decompose the input data into a new orthogonal basis, with axes ordered by "variance (or inertia) explained". PCA used with images as observation vectors can be used, for example, for a best synthesis of multiple band images, or for producing eigen-faces in face recognition. Among other data preprocessing transforms, we also include the discrete cosine transform (DCT), which decomposes the data into an orthogonal basis of cosine functions; and the Fourier transform (FT) which uses a basis of sine and cosine functions, each at different frequencies.

PCA, DCT, and FT have the property of *energy packing* (Seales et al. 1996): most of the energy (second order moment) of the input vector is packed into the first few values of the output vector. Thus, one can roughly approximate, or even eliminate, all but the most important values and still preserve most of the input energy.

The wavelet transform (WT), whether orthonormal as in the case of the Haar or Daubechies transforms or non-orthogonal as in the case of the à trous method, is different. It can be viewed as an automatic method for laying bare superimposed scale-related components of the data. Our analysis of the data may be considerably improved by removing noise in *all* scale-related components. This perspective differs from the usual approach of PCA, DCT, and FT: in these methods we remove output scales (or "levels") entirely to filter the data.

We turn attention now to denoising through modification of scale information at all levels. This is the preferred method of denoising using the wavelet transform.

Donoho et al. (1995) have proposed a "universal threshold", $\sqrt{2 \log n}\sigma$, used in the additive Gaussian noise case where σ is the known or estimated standard deviation of the data, and n is the size of the input data set. Wavelet coefficients above this threshold are retained, and those below the threshold are set to zero. The authors also propose a soft threshold, referred to as wavelet shrinkage, which reduces wavelet values by a fraction of their initial values.

As an alternative to such hard and soft thresholding, Starck et al. (1995) assume known or estimated noise properties for the input data, and then derive or make use of wavelet coefficient probability distributions at each level, under a null hypothesis of stochastic input. Other noise modeling work in this direction can be found in Kolaczyk

(1997) and Powell et al. (1995), albeit with different wavelet transforms.

In the work described in this paper we employ thresholding in a data- and noise-driven manner.

2.3 Choice of Multiscale Transform

Background texts on the wavelet transform include Meyer (1993), Combes et al. (1989), Mallat (1989), Shensa (1992), Bijaoui et al. (1994), Wickerhauser (1994) and Strang and Nguyen (1996). Some important properties of the à trous wavelet transform are as follows.

As already noted, the à trous transform is isotropic. Unlike it, Mallat's widely-used multiresolution algorithm (Mallat, 1989) leads to a wavelet transform with three wavelet functions (at each scale there are three wavelet coefficient subimages) which does not simplify the analysis and the interpretation of the wavelet coefficients. Other anisotropic wavelets include the similarly widely-used Haar and Daubechies wavelet transforms. An isotropic wavelet seems more appropriate for images containing features or objects with no favored orientation.

An important property of the à trous wavelet transform over other wavelet transforms is shift invariance. Lack of independence to pixel shift is a problem in the case of any pyramidal wavelet transform (Haar, Daubechies, Mallat, etc.) due to the downsampling or decimating. The reason is simply that shift-variance is introduced because Nyquist sampling is violated in each of the (wavelet-decomposed) subbands – wavelets are not ideal filters. By not downsampling the problem is avoided. Various authors have proposed solutions to this problem. The à trous algorithm is in fact a fast imple-

mentation of a wavelet transform with no downsampling.

Two inconvenient aspects of many wavelet transforms are negative values and lack of robustness. By definition, the wavelet coefficient mean at each level is null. Every time we have a positive structure at a scale, we have negative values surrounding it. These negative values often create artifacts during the data reconstruction process, or complicate the analysis. For instance, if we threshold small values (noise, non-significant structures, etc.) in the wavelet transform, and if we reconstruct the image at full resolution, the structure's total intensity will be modified. Furthermore, if an object is associated with high intensity values, the negative values will be significant too and will lead to false structure detections. Point artifacts (e.g. cosmic ray hits in optical astronomy, glitches in the infrared ISO, Infrared Satellite Observatory, detectors) can "pollute" all scales of the wavelet transform. The wavelet transform is non-robust relative to such real or detector faults.

One way around both of these issues – negative wavelet coefficient values, and non-robustness relative to anomalous values – is to keep certain aspects of the multiscale decomposition algorithm provided by the à trous wavelet transform, but to base our algorithm on a function other than the wavelet function. The median smoothing transform provides us with one such possibility. A multiscale pyramidal median transform, for instance, was investigated in Starck et al. (1996). We conclude that the wavelet transform, à trous or otherwise, is not sacrosanct. Depending on the data, it may well be advisable and necessary to develop new multiresolution tools.

2.4 The Multiresolution Support

We will say that a multiresolution support (Starck et al., 1995) of an image describes in a logical or boolean way if an image I contains information at a given scale j and at a given position (x, y). If $M^{(I)}(j, x, y) = 1$ (or $= true$), then I contains information at scale j and at the position (x, y). M depends on several parameters:

- The input image.
- The algorithm used for the multiresolution decomposition.
- The noise.
- All constraints which we want the support additionally to satisfy.

Such a support results from the data, the treatment (noise estimation, etc.), and from knowledge on our part of the objects contained in the data (size of objects, linearity, etc.). In the most general case, a priori information is not available to us.

The multiresolution support of an image is computed in several steps:

- Step 1 is to compute the wavelet transform of the image.
- Booleanization of each scale leads to the multiresolution support.
- A priori knowledge can be introduced by modifying the support.

The last step depends on the knowledge we have of our images. For instance, if we know there is no interesting object smaller or larger than a given size in our image, we can suppress, in the support, anything which is due to that kind of object. This can often be done conveniently by the use of mathematical morphology (Breen et al.,

1998). In the most general setting, we naturally have no information to add to the multiresolution support.

The multiresolution support will be obtained by detecting at each scale the significant coefficients. The multiresolution support is defined by:

$$M(j,x,y) = \begin{cases} 1 & \text{if } w_j(x,y) \text{ is significant} \\ 0 & \text{if } w_j(x,y) \text{ is not significant} \end{cases} \tag{5}$$

In order to visualize the support, we can create an image S defined by:

$$S(x,y) = \sum_{j=1}^{p} 2^j M(j,x,y) \tag{6}$$

Given stationary Gaussian noise, to define if w_j is significant, it suffices to compare $w_j(x,y)$ to $t\sigma_j$, where σ_j is the noise standard deviation at scale j. Often t is chosen as 3. If $w_j(x,y)$ is small, it is not significant and could be due to noise. If $w_j(x,y)$ is large, it is significant:

$$\begin{aligned} &\text{if } \mid w_j \mid \geq t\sigma_j \quad \text{then } w_j \text{ is significant} \\ &\text{if } \mid w_j \mid < t\sigma_j \quad \text{then } w_j \text{ is not significant} \end{aligned} \tag{7}$$

So we need to estimate, in the case of Gaussian noise models, the noise standard deviation at each scale.

3 Noise Modeling

3.1 The Overall Procedure

Images and sets of point patterns generally contain noise. Hence the wavelet coefficients are noisy too. In most applications, it is necessary to know if a wavelet coefficient is due to signal (i.e. it is significant) or to noise.

The wavelet transform yields a set of resolution-related views of the input image. A wavelet image scale at level j has coefficients given by $w_j(x, y)$. If we obtain the distribution of the coefficient $w_j(x, y)$ for each resolution plane, based on the noise, we can introduce a statistical significance test for this coefficient. This procedure is the classical significance-testing one. Let $\mathcal{H}_0$ be the hypothesis that the image is locally constant at scale j. Rejection of hypothesis $\mathcal{H}_0$ depends (for interpretational reasons, restricted to positive coefficient values) on:

$$P = Prob(\mid w_j(x, y) \mid < \tau \mid \mathcal{H}_0) \tag{8}$$

The detection threshold, τ, is defined for each scale. Given an estimation threshold, ϵ, if $P = P(\tau) > \epsilon$ the null hypothesis is not excluded. Although non-null, the value of the coefficient could be due to noise. On the other hand, if $P < \epsilon$, the coefficient value cannot be due to the noise alone, and so the null hypothesis is rejected. In this case, a significant coefficient has been detected.

3.2 The Case of Gaussian Noise

We start with a consideration of the Gaussian case, and then incorporate the Poisson case. Such noise distributions are the most important for CCD, digitized photographic, and other common astronomical image detectors.

The following discussion will be based on the à trous wavelet transform. For a non-wavelet transform such as the multiscale median transform, we can use a very similar perspective, but it then becomes an approximate one.

The appropriate value of σ_j in the succession of wavelet planes is assessed from

the standard deviation of the noise σ_I in the original image and from study of the noise in the wavelet space. This study consists of simulating an image with zero signal, containing Gaussian noise with a standard deviation equal to 1, and taking the wavelet transform of this image. Then we compute the standard deviation σ_j^e at each scale. We get a curve σ_j^e as a function of j, giving the behavior of the noise in the wavelet space. (Note that if we had used an orthogonal wavelet transform, this curve would be linear.) Due to the properties of the wavelet transform, we have

$$\sigma_j = \sigma_I \sigma_j^e \tag{9}$$

This numerical simulation therefore gives us values for σ_j^e which we will make use of subsequently.

3.3 Automatic Estimation of Gaussian Noise

The Gaussian noise σ_I can be estimated automatically in an image I. This estimation is very important, because all the noise standard deviations σ_j in the scales j are derived from σ_I. Thus an error associated with σ_I will introduce an error on all σ_j. This measure of σ_I can be refined by the use of the multiresolution support. Indeed, if we consider the set of pixels $\mathcal{S}$ in the image which are due only to the noise, and if we take the standard deviation of them, we would obtain a good estimate of σ_I. This set is easily obtained from the multiresolution support. We say that a pixel (x, y) belongs to the noise if $M(j, x, y) = 0$ for all j (i.e. there is no significant coefficient at any scale). The new estimation of σ_I is then computed by the following iterative algorithm:

1. Estimate the standard deviation of the noise in I (usually by sampling the image in areas of relatively constant signal): we have $\sigma_I^{(0)}$.

2. Compute the wavelet transform (à trous algorithm) of the image I with p scales, providing the additive decomposition see in equation (1) above.

3. Set n to 0.

4. Compute the multiresolution support M which is derived from the wavelet coefficients and from $\sigma_I^{(n)}$ (equations (7) and (5)).

5. Select the pixels which belong to the set $\mathcal{S}$: if $M(j,x,y) = 0$ for all j in $1 \ldots p$, then the pixel $(x,y) \in \mathcal{S}$.

6. For all the selected pixels (x,y), compute the values $I(x,y) - c_p(x,y)$ and compute the standard deviation $\sigma_I^{(n+1)}$ of these values (we compute the difference between I and c_p in order not to include the background in the noise estimation).

7. $n = n + 1$

8. If $\frac{|\sigma_I^{(n)} - \sigma_I^{(n-1)}|}{\sigma_I^{(n)}} > \epsilon$ then go to 4.

This method converges in a few iterations, and allows the noise estimation to be improved.

The approach is in fact very physically-meaningful. It consists of detecting the set $\mathcal{N}$ of pixels which does not contain any significant signal (only the background + noise). A pixel (x,y) is dominated by the noise if all wavelet coefficients at this position are not significant. The background affects only the last scale of the wavelet transform.

We subtract this last scale from the original image, and we compute the standard deviation of the set $\mathcal{N}$ in this background-free image. Wavelet coefficients larger than $3\sigma_j$ are considered as significant, but a small fraction of them will be due to the noise. This introduces a small systematic bias in the final solution, which is easily corrected by dividing the standard deviation by a given constant value, found experimentally as equal to 0.974. Therefore we downgrade the empirical variance in this way. The method is robust and whatever the initial estimation of noise, it converges quickly to a good estimate.

More information on this framework for automated noise estimation can be found in Starck and Murtagh (1998a).

3.4 Cases of Poisson or Poisson and Gaussian Noise

If the noise in the data I is Poisson, the transformation

$$t(I(x,y)) = 2\sqrt{I(x,y) + \frac{3}{8}} \tag{10}$$

acts as if the data arose from a Gaussian white noise model (Anscombe, 1948), with $\sigma = 1$, under the assumption that the mean value of I is large. 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 read-out noise. The Anscombe transformation has been extended to take this combined noise into account. The generalization of the variance stabilizing Anscombe formula is derived as (Murtagh et al., 1995):

$$t(I) = \frac{2}{\alpha}\sqrt{\alpha I(x,y) + \frac{3}{8}\alpha^2 + \sigma^2 - \alpha g} \tag{11}$$

where α is the detector gain, σ and g the standard deviation and the mean of the read-out noise respectively.

These 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 10 if bias is to be avoided. In the next section, we will look at an approach which also handles the case of very low average numbers of counts.

3.5 Case of Poisson Noise with Few Counts

The low count situation is what we face when dealing with planar point patterns. This also is the case with certain types of astronomical X-ray images. We have often in such images a low number of counts, and extensive very sparsely populated regions.

A wavelet coefficient is by definition obtained by convolution product between the input data and the dilated wavelet function. Then a wavelet coefficient at a given position and at a given scale j is

$$w_j(x,y) = \sum_{k=1}^{n_e} \psi(\frac{x_k - x}{2^j}, \frac{y_k - y}{2^j}) \tag{12}$$

where n_e is the number of events which contribute to $w_j(x,y)$, i.e. the number of events included in the support of the dilated wavelet function $\psi(\frac{x}{2^j}, \frac{y}{2^j})$ centered at (x,y); and (x_k, y_k) are the event coordinates (which can be non integer values).

If a wavelet coefficient $w_j(x,y)$ is due to noise, it can be considered as a realization of the sum of n_e independent random variables with the same distribution as that of the wavelet function. This allows comparison of the wavelet coefficients of the data

with the values which can be taken by the sum of n_e independent variables.

The distribution of one event in wavelet space is then directly given by the histogram H_1 of the wavelet ψ. As we consider independent events, the distribution of a coefficient w_n (note the changed subscripting for w, for convenience) related to n events is given by n autoconvolutions of H_1:

$$H_n = H_1 \otimes H_1 \otimes ... \otimes H_1 \tag{13}$$

This principle of using autoconvolutions of the wavelet in order to model the null hypothesis marginal distribution was also used or described by Slezak et al. (1993), Bury (1995), and Starck and Pierre (1998c). For a large number of events, H_n converges to a Gaussian.

It may be helpful to comment somewhat further on the implementation of this procedure. Consider a wavelet coefficient, at a given pixel, at some resolution level. At any resolution level, the support of the wavelet function defines a region and the input data provides a number of counts in this region. The autoconvolutions then show the anticipated distribution of the wavelet coefficient. This is the distribution of the wavelet-transformed Poisson noise. This distribution function summarizes all possible eventualities, from the extreme of a uniform arrangement of counts, through to the extreme of all counts being stacked in one pixel. Tables representing the distribution functions – or, in practice, detection thresholds – are determined once, and stored for subsequent data analyses.

In order to facilitate the comparisons, the variable w_n of distribution H_n is reduced

by $c = [w_n - E(w_n)]/\sigma(w_n)$ and the distribution function is

$$F_n(c) = \int_{-\infty}^{c} H_n(u)du$$

From F_n, we derive c_{min} and c_{max} such that $F(c_{min}) = \epsilon$ and $F(c_{max}) = 1 - \epsilon$.

Then a reduced wavelet coefficient $w_j^r(x, y)$, calculated from $w_j(x, y)$, resulting from n events is significant (positive coefficients only being considered for interpretational reasons) if $F(w^r) > c_{max}$. Finally, $w_r(j, x, y)$ is obtained from $w_j(x, y)$ ($w_j(x, y)$ being obtained using the à trous algorithm) by

$$\begin{aligned} w_j^r(x,y) &= \frac{w_j(x,y)}{\sqrt{n}\sigma_{\psi_j}} \\ &= \frac{w_j(x,y)}{\sqrt{n}\sigma_{\psi}} 4^j \end{aligned}$$

where σ_ψ is the standard deviation of the wavelet function, and σ_{ψ_j} is the standard deviation of the dilated wavelet function ($\sigma_{\psi_j} = \sigma_\psi / 4^j$).

4 Examples

4.1 Determining Noise in 2D Images

A simulated image containing stars and galaxies is shown in Fig. 1 (top left). This image was constructed by sampling from a library of object shapes, reflecting realistic luminosity distributions, and placing these objects in the image at randomly chosen locations. Then background and noise were added, in fairly close analogy with common detector characteristics. The simulated noisy image, the filtered image and the residual

image are respectively shown in Fig. 1 top right, bottom left, and bottom right. We can see that there is no structure in the residual image. The filtering was carried out using the multiresolution support, and thereby determining of signal at each level, before reconstruction of the image.

It is possible for us to see in Fig. 1, and below in Fig. 4, the influence of the isotropic wavelet function on the structures reconstructed following filtering, or found at different resolution scales. In Fig. 1, this is certainly accentuated by the display color function used: the residual image confirms this. An isotropic wavelet function provides an excellent general analyzing function for the myriad range of objects or structures found in practice.

4.2 Filtering in 1D Images

Fig. 2 shows a noisy spectrum (top). For the astronomer, the spectral lines – here mainly absorption lines extending downwards – are of interest. The continuum may also be of interest, i.e. the overall spectral tendency or overall spectral shape. The spectral lines are unaffected in the filtered version (center). The bottom part of the figure shows the input and filtered result superimposed. Noise is suppressed in the unimportant parts of the spectrum, and the important parts are untouched.

A similar filtering was used, based on the Daubechies coefficient 8 (Daubechies, 1992), a compactly-supported orthonormal wavelet, followed by thresholding based on estimated variance of the coefficients as proposed by Donoho (1995). This did not take into account the image's noise properties as we have done, and results were very poor. A problem- (or image-) driven choice of wavelet and filtering strategy is indispensable.

4.3 Point Patterns

Point patterns can represent very diverse situations, e.g. mines in a coastal waters minefield, or earthquake locations (Dasgupta and Raftery, 1995; Murtagh and Starck, 1998). Given a planar point pattern, a 2-dimensional image is created by:

1. Considering a point at (x, y) as defining the value one at that point, yielding the tuple $(x, y, 1)$.

2. Projection onto a plane by (i) using a regular discrete grid (an image) and (ii) assigning the contribution of points to the image pixels by means of the interpolation function, ϕ, used by the chosen wavelet transform algorithm (in our case, the à trous algorithm with a B_3 spline).

3. The à trous algorithm is applied to the resulting image. Based on a noise model for the original image (i.e. tuples $(x, y, 1)$), significant structures are detected at each resolution level.

Fig. 3 shows a point pattern set. Fig. 4 shows the corresponding wavelet transform. Wavelet scales 1–6 are shown in sequence, left to right, starting at the upper right corner. The images shown in Fig. 4 may be summed pixel-wise to exactly reconstitute a *smoothed version* of Fig. 3, the smoothing being carried out, as mentioned above, by a B_3 spline. Two technical remarks regarding Fig. 4 are that (i) we rebinned each image to 128×128 from the input 256×256 to cut down on space, and (ii) this figure is shown histogram-equalized to more clearly indicate structure.

Fig. 3 shows two Gaussian clusters designed with centers $(64, 64)$ and $(190, 190)$; and with standard deviations in x and y directions respectively $(10, 20)$ and $(18, 10)$. In

the first (lower) of these clusters, there are 300 points, and there are 250 in the second. Background Poisson clutter was provided by 300 points. The 5th wavelet scale (see Fig. 4), following application of the significance threshold for positive values ($\epsilon = 0.001$), provides good reconstruction of the input clusters. The centroid values of the "island" objects were found to be respectively $(62, 63)$ and $(190, 190)$ which are very good fits to the design values. The standard deviations in x and y were found to be respectively $(9.3, 14.4)$ and $(14.9, 8.9)$, again reasonable fits to the input data.

A remark to make in regard to this result is that the wavelet method used prioritizes the finding of Gaussian-like shapes. The wavelet, as already mentioned, is a B_3 spline.

5 Conclusion

This article has discussed a general methodology for data analysis, focusing on image data and including also low-dimensional point pattern data. In various aspects, it is an alternative methodology to more classical procedures. As a methodology, it has been developed to tackle a wide range of real data analytic problems. We have shown in many case-studies and practical situations that results (filtering, deconvolution, data registration, feature detection, compression, and so on) are better than more classical alternatives. Many such data processing and data analysis methods are discussed and exemplified in Starck et al. (1998a).

For the types of data worked on, the transforms used were the most appropriate. It may be the case that other aspects can be better handled by other wavelet transforms. Edges or sharp jumps in features may be enhanced by use of an anisotropic

wavelet transform. In Starck and Murtagh (1998b,1998e) and Starck et al (1998d), we investigate and compare the combination of various wavelet, filtering strategies, and multiscale transforms to allow for performance enhancements over and above use of a single multiscale transform.

The work described in this paper has also resulted in a large software package, supporting a range of wavelet and other multiscale transforms, and also supporting a wide range of noise models. Visualization, noise filtering, deconvolution and compression are among the many objectives covered. Further information on this software package can be obtained from the authors or at

http://ourworld.compuserve.com/homepages/multires

References

Anscombe, F.J. (1948) The transformation of Poisson, binomial and negative-binomial data. Biometrika, 15, 246–254.

Bijaoui, A., Starck, J.-L. and Murtagh, F. (1994) Restauration des images multi-échelles par l'algorithme à trous, Traitement du Signal, 11, 229–243.

Breen, E.J., Jones, R. and Talbot, H. (1998) Mathematical morphology: a useful set of tools for image analysis, Statistics and Computing (this issue).

Bruce, A. and Gao, H.-Y. (1994) S+Wavelets User's Manual, Version 1.0, StatSci Division, MathSoft Inc., Seattle.

Bury, P. (1995) De la distribution de matière à grande échelle à partir des amas d'Abell

(On the large-scale distribution of matter using Abell clusters), PhD Thesis, University of Nice, Sophia Antipolis.

Combes, J.M., Grossmann, A. and Tchamitchian, Ph., Eds. (1989) Wavelets: Time-Frequency Methods and Phase Space, Springer-Verlag, Berlin.

Dasgupta, A. and Raftery, A.E. (1995) Detecting features in spatial point processes with clutter via model-based clustering, Technical Report 295, Department of Statistics, University of Washington. Available at http://www.stat.washington.edu/tech.reports/. Journal of the American Statistical Association, in press.

Daubechies, I. (1992) Ten Lectures on Wavelets, SIAM (Society for Industrial and Applied Mathematics), Philadelphia

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia? (with discussion). Journal of the Royal Statistical Society, Series B, 57, 301–370.

Holschneider, M., Kronland-Martinet, R., Morlet, J. and Tchamitchian, Ph. (1989) A real-time algorithm for signal analysis with the help of the wavelet transform, in J.M. Combes, A. Grossmann and Ph. Tchamitchian, Eds., Wavelets: Time-Frequency Methods and Phase Space, Springer-Verlag, Berlin, 286–297.

Kolaczyk, E.D. (1997) Estimation of intensities of burst-like processes using Haar wavelets. Submitted to Journal of the Royal Statistical Society Series B, Department of Statistics, University of Chicago, 27 pp..

Mallat, S. (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 674–693.

Meyer, Y. (1995) Wavelets: Algorithms and Applications, SIAM, Philadelphia.

Murtagh, F., Starck, J.-L. and Bijaoui, A. (1995) Image restoration with noise suppression using a multiresolution support, Astronomy and Astrophysics Supplement Series, 112, 179–189.

Murtagh, F. and Starck, J.-L. (1998) Pattern clustering based on noise modeling in wavelet space. Pattern Recognition, 31, 847–855.

Mccoy, E.J. and Walden, A.T. (1996) Wavelet analysis and synthesis of stationary long-memory processes, Journal of Computational and Graphical Statistics, 5, 26–56.

Meyer, Y. (1993) Wavelets: Algorithms and Applications, Phildelphia: SIAM (Society for Industrial and Applied Mathematics). Original French version (1990), Ondelettes, Hermann, Paris.

Olsen, S.I. (1993) Estimation of noise in images: an evaluation. CVGIP: Graphical Models and Image Processing, 55, 319–323.

Powell, K.J., Sapatinas, T., Bailey, T.C. and Krzanowski, W.J. (1995) Application of wavelets to pre-processing of underwater sounds. Statistics and Computing, 5, 265–273.

Seales, W.B., Yuan, C.J., Hu W. and Cutts, M.D. (1996) Content analysis of compressed video, University of Kentucky Computer Science Department Technical Report No. 2, 28 pp.

Shensa, M.J. (1992), The discrete wavelet transform: wedding the à trous and Mallat algorithms. IEEE Transactions on Signal Processing, 40, 2464–2482.

Slezak, E., de Lapparent, V., and Bijaoui, A. (1993) Objective detection of voids and high density structures in the first CfA redshift survey slice, Astrophysical Journal, 409, 517–529.

Snyder, D.L., Hammoud, A.M. and White, R.L. (1993) Image recovery from data acquired with a charge-coupled camera. Journal of the Optical Society of America, 10, 1014–1023.

Starck, J.-L. and Bijaoui, A. (1994), Filtering and deconvolution by the wavelet transform, Signal Processing, 35, 195–211.

Starck, J.-L. and Murtagh, F. (1994) Image restoration with noise suppression using the wavelet transform, Astronomy and Astrophysics, 288, 342–348.

Starck, J.-L., Murtagh, F. and Bijaoui, A. (1995) Multiresolution support applied to image filtering and deconvolution, Graphical Models and Image Processing, 57, 420–431.

Starck, J.-L., Murtagh, F., Pirenne, B. and Albrecht, M. (1996) Astronomical image compression based on noise suppression. Publications of the Astronomical Society of the Pacific, 108, 446–455.

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

Starck, J.-L. and Murtagh, F. (1998a) Automatic noise estimation from the multiresolution support. Publications of the Astronomical Society of the Pacific, 193–199.

Starck, J.-L. and Murtagh, F. (1998b) Image filtering from the combination of multivision models. Submitted.

Starck, J.-L. and Pierre, M., (1998c) Structure Detection in Low Intensity X-Ray Images, Astronomy and Astrophysics, Suppl. Ser., 128, pp 397–407.

Starck, J.-L., Murtagh, F., and Gastaud, R. (1998d) A New Entropy Measure Based on the Wavelet Transform and Noise Modeling, IEEE Transactions on Circuits and Systems II on Multirate Systems, Filter Banks, Wavelets, and Applications, 45, 8.

Starck, J.-L. and Murtagh, F. (1998e) Multiscale Entropy Filtering, submitted to Signal Processing, 1998.

Strang, G. and Nguyen, T. (1996) Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley.

Tekalp, A.M. and Pavlović, G. (1991) Restoration of scanned photographic images, in A.K. Katsaggelos, Ed., Digital Image Restoration, Springer-Verlag, New York, 209–239.

Wickerhauser, M.V. (1994) Adapted Wavelet Analysis from Theory to Practice, A.K. Peters, Wellesley.

Annex A: the à trous algorithm

We consider spectra, $\{c_0(k)\}$, defined as the scalar product at samples k of the function $f(x)$ with a scaling function $\phi(x)$ which corresponds to a low pass filter:

$$c_0(k) =< f(x), \phi(x-k) > \tag{14}$$

The scaling function is chosen to satisfy the dilation equation:

$$\frac{1}{2}\phi(\frac{x}{2}) = \sum_l h(l)\phi(x-l) \tag{15}$$

where h is a discrete low-pass filter associated with the scaling function ϕ. This means that a low-pass filtering of the signal is, by definition, closely linked to another resolution level of the signal. The distance between levels increases by a factor 2 from one scale to the next.

The smoothed data $c_j(k)$ at a given resolution j and at a position k is the scalar product

$$c_j(k) = \frac{1}{2^j} < f(x), \phi(\frac{x-k}{2^j}) > \tag{16}$$

This is consequently obtained by the convolution:

$$c_j(k) = \sum_l h(l) \;\; c_{j-1}(k + 2^{j-1}l) \tag{17}$$

The signal difference w_j between two consecutive resolutions is:

$$w_j(k) = c_{j-1}(k) - c_j(k) \tag{18}$$

or:

$$w_j(k) = \frac{1}{2^j} < f(x), \psi(\frac{x-k}{2^j}) > \tag{19}$$

Here, the wavelet function ψ is defined by:

$$\frac{1}{2}\psi(\frac{x}{2}) = \phi(x) - \frac{1}{2}\phi(\frac{x}{2}) \tag{20}$$

Equation 19 defines the discrete wavelet transform, for a resolution level j.

A series expansion of the original signal, c_0, in terms of the wavelet coefficients is now given as follows. The final smoothed array $c_p(x)$ is added to all the differences w_j:

$$c_0(k) = c_p + \sum_{j=1}^{p} w_j(k) \tag{21}$$

This equation provides a reconstruction formula for the original signal. At each scale j, we obtain a set $\{w_j\}$ which we call a wavelet scale. The wavelet scale has the same number of samples as the signal.

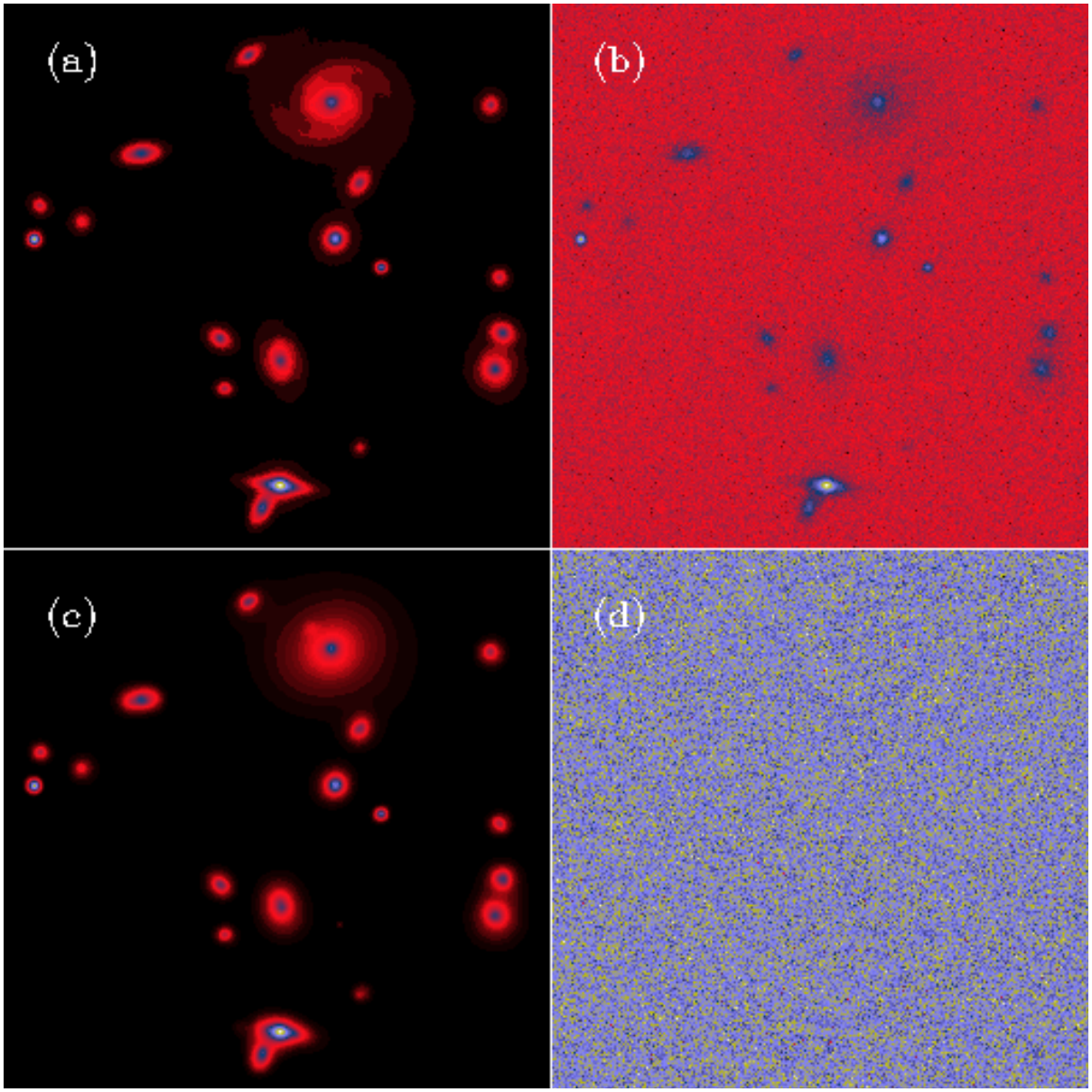

Figure 1: (a) Simulated image, (b) simulated image and Gaussian noise, (c) filtered image, and (d) residual image.

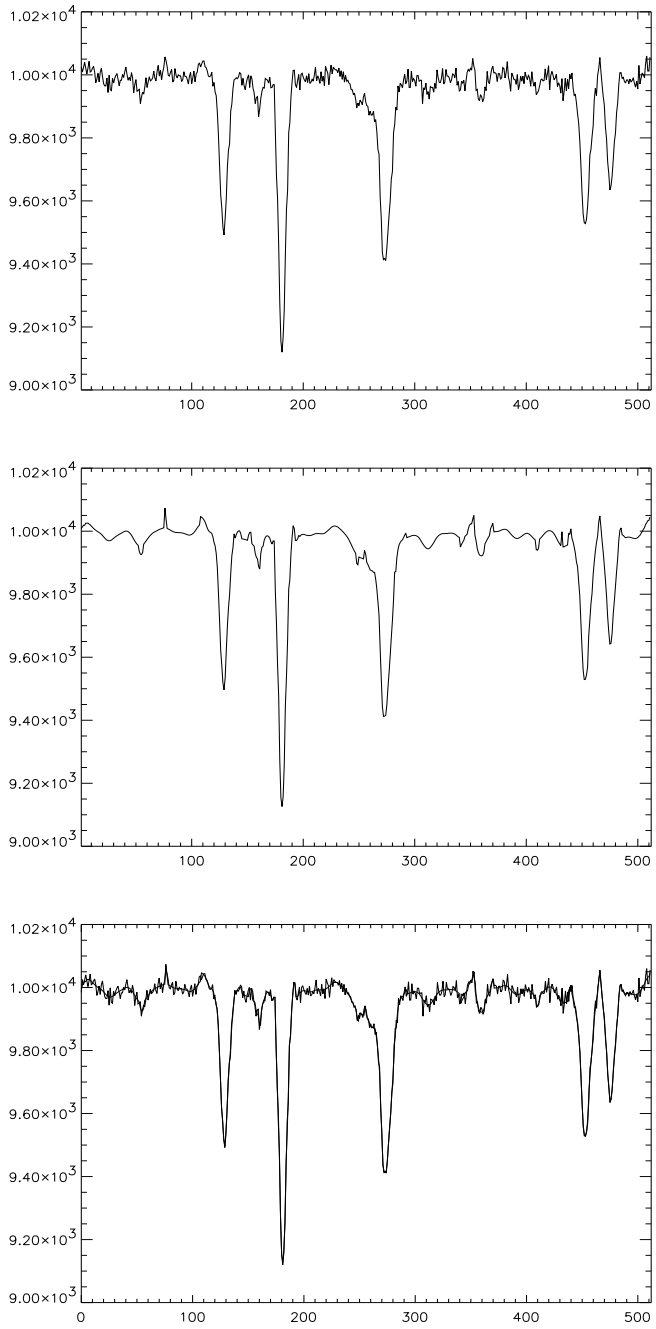

Figure 2: From top to bottom: input spectrum; denoised spectrum; and both overplotted.

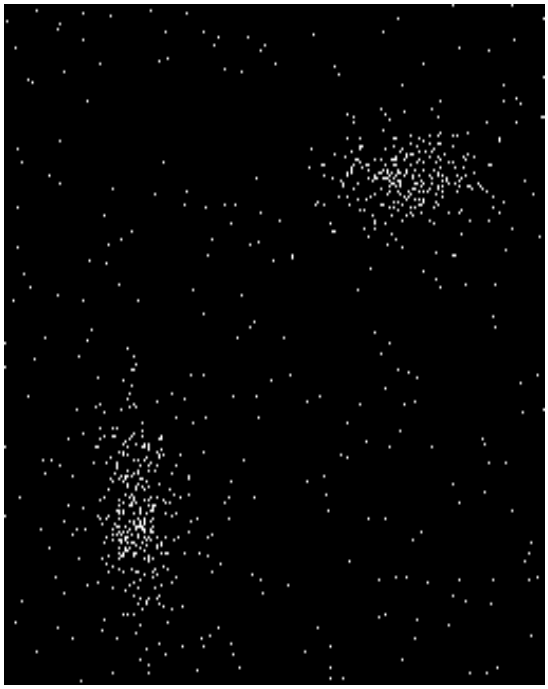

Figure 3: Simulated Gaussian clusters with 300 and 250 points; and background Poisson noise with 300 points.

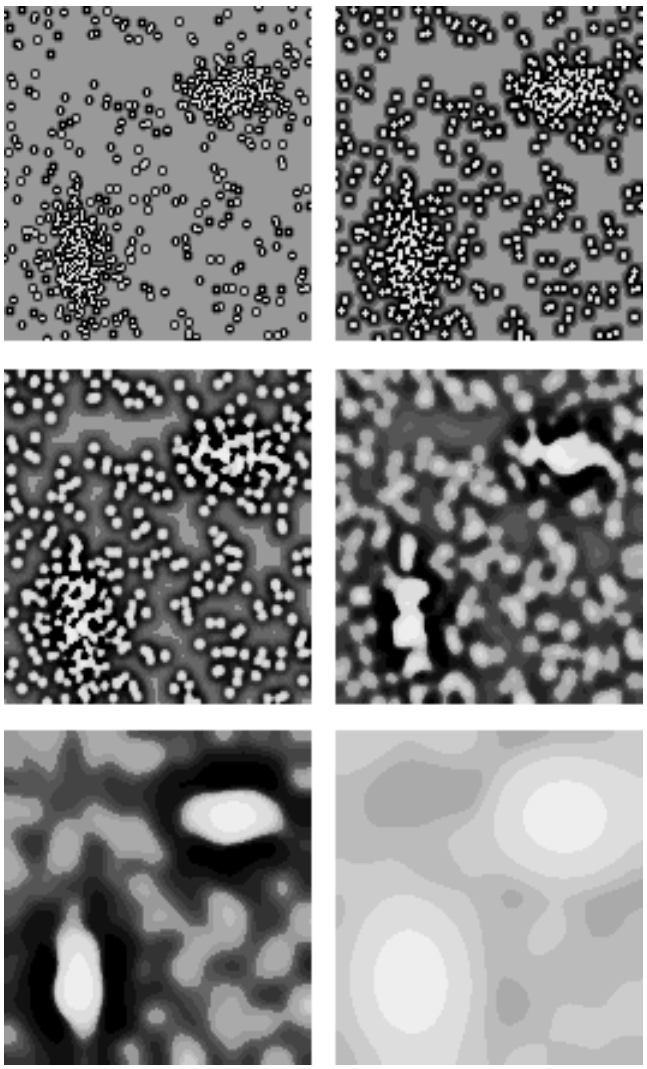

Figure 4: Wavelet transform (à trous method) of Fig. 3.