Wavelets in the Astronomical Context

Shashikiran Ganesh

Physical Research Laboratory

Abstract

Wavelets are mathematical functions of zero mean. They have an enormous range of practical applications. In the present report I start[1] off with an introduction to wavelets. I discuss in brief the historical developments. In the following section I discuss data compression application of wavelets since that is one application which is useful in all fields. I analyse the usage of wavelets in astronomical research in a subsequent section. An application of wavelet transforms using the Multi-scale Vision Model for the detection of features in images of the Milky Way Galaxy is then shown. The basic principles of MVM analysis are presented. I then discuss results for the detection of positive features towards the starforming region Sagittarius B and cataloging of negative features (dark clouds) for another region. A brief survey of interesting WWW URLs, software links (including links to software used in the preparation of figures in this report), books and research papers follows.

[1]Following Lewis Carroll's rather overused dictum : "Begin at the beginning and go on till you come to the end; then stop."

1 What are wavelets

Wavelets are mathematical functions. In this age, almost everyone has used technology that has benefitted through the application of wavelets. Google, the all-knowing source of information, lists 2.7×10^5 results for the term 'wavelet' (and 1.51×10^5 for 'wavelets'). Most of the development in the usage of wavelets has happened in the last few decades - mirroring the development of computers and signal processing requirements.

The wavelet function can be used to approximate other functions. They can be used to represent or approximate other functions of more complicated form. They have wide ranging application in fields as diverse as astronomy, crime detection, data compression, genetics, geophysics, mathematics, music & movies, image and signal processing etc.

A study of the history of wavelets shows that the concepts developed in a parallel and independent manner in different fields such as geophysics and mathematics. To the extent that the wavelet transform can be used to approximate other complicated functions, analysis using wavelets is similar to that using Fourier transforms. However, while Fourier transforms represent any function as a series of sine/cosine terms (*the basis functions*) with varying frequencies, the basis functions for the wavelet transform can be defined in many different ways. The wavelet is a function with zero average and is defined over a limited domain (being zero everywhere else). Thus a wavelet basis is a localized function in contrast to the sines and cosines of the fourier series which are non-local. Wavelet analysis therefore allows to localise features (in time/frequency or space) and allows for very good approximation to sharp changes in signal levels (such as spikes and other 'choppy' signals). The basic idea underlying the use of wavelets is to analyze data according to scale or resolution. Analysis of a signal over a large scale provides information about the gross features, i.e. the low frequencies. On the otherhand, analysis with a small window (translated over the entire data set in steps) provides information about the high frequencies (i.e. the small scale variations or sharp discontinuities) and also localizes them. The resolution or localization then depends on the size of the window, being better for smaller windows.

To quote Daubechies & Gilbert : "Wavelets constitute a tool to decompose, analyze, and synthesize functions, with emphasis on time-frequency localization."

We represent a function f of t by

$$f(t) = \sum_{j,k\in\mathbb{Z}} C_{j,k}\ \psi_{j,k}(t). \tag{1}$$

Here, the functions $\psi_{j,k}$ are the wavelets obtained by scaling and translating a 'mother' wavelet ψ in the following way:

$$\psi_{j,k}(t) \ = \ 2^{j/2}\psi(2^j t - k) \text{ for } j,k \ \in \ \mathbb{Z}. \tag{2}$$

where wavelet ψ satisfies

$$\int_{-\infty}^{\infty} \psi(t)\ dt \ = \ 0 \tag{3}$$

In figure 1 we show examples from the Daubechies family of wavelets.

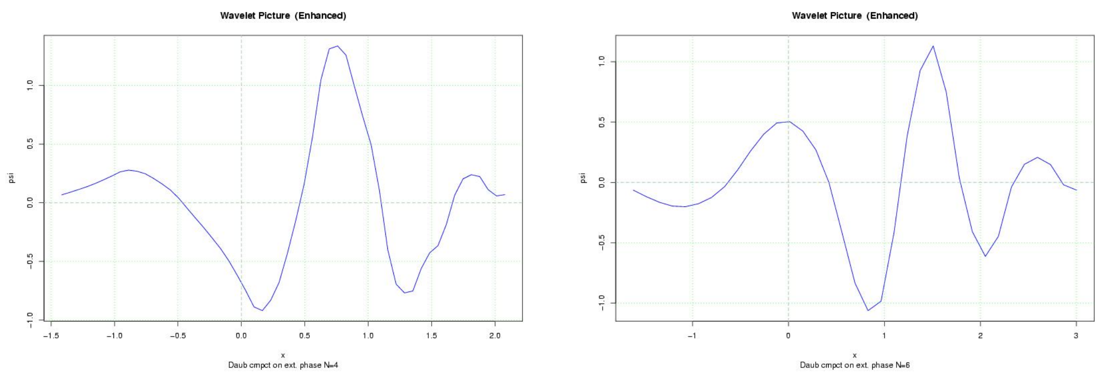

Figure 1: What do wavelets look like: Daubechies wavelet with order=4 (left) and order=6 (right)

Data analysis operations can be performed using the corresponding wavelet coefficients (from the linear combination of the wavelet functions that represent the original function). A large variety of wavelet basis functions are now known. The selection of the appropriate wavelet prototype function depends, largely, on the kind of data being analyzed and, to some extent, requires *a priori* information about the feature expected to be detected.

2 Historical developments

Historically, the roots of modern wavelet analysis goes back to the work of the French mathematician Fourier. He showed that functions can be represented as a sum of sine and cosine terms. This is known as the Fourier expansion. An extension of this is the Fourier transform which transforms a function from the time (or space) domain to the frequency domain.

Alfred Haar, in his thesis (1909), discussed orthogonal systems of functions. These form the simplest of the wavelet basis now known after him. Functions that were eventually to earn the sobriquet of 'wavelets' were studied by pure mathematicians, alongside tools such as the windowed fourier transform etc.

The importance and power of wavelets was realized (mid 1970s) in the work of Jean Morlet, a geophysicist. Morlet used the term wavelet to describe the functions he was using. His work was in the field of oil exploration. He mainly used windowed fourier analysis to study the echoes of the probing impulses which were used to probe the oil layer thickness underground. But the windowed fourier analysis technique is a time consuming process so Morlet evolved a new technique of varying not the frequency but the size of the window. Thus stretching the window stretched the function and compressing the window squeezed the function. Morlet's wavelet was basically a sine wave with a window defined by a gaussian. The similarity between the sinusoidal terms of classical Fourier analysis and the Morlet wavelets is quite apparent.

Building up on this foundation, Morlet collaborated with Alex Grossman and in the early 1980s came up with the ideas that a signal could be transformed into wavelet domain and then transformed back into the original without any loss of information. They discovered that the back transformation from the wavelet coefficients to the original signal could be done with a single integral. Another result was that small change in the wavelets caused only small changes in the reconstructed signal (as compared to the original signal). This is the basic idea behind data compression using wavelets.

The next major contribution to wavelet techniques was by Yves Meyer and Stephane Mallat.

They introduced (1986) the ideas of multiresolution analysis using wavelets. This was where the scaling function of wavelets was first discussed, allowing mathematicians to construct individual wavelet families. In what follows, we shall discuss multiresolution analysis using wavelets in a practical application to astronomy. Stephane Mallat describes the multiresolution methods (among other things) in his book *A Wavelet Tour of Signal Processing*.

Around 1988, Ingrid Daubechies used the idea of multiresolution analysis to create her own family of wavelets. The Daubechies family of wavelets satisfy a number of wavelet properties. Two examples of wavelets from this family are shown in figure 1. Her book of lectures *Ten lectures on wavelets* (1992) was awarded the Steele Prize for mathematical exposition by the American Mathematical Society (1994). The award citation reads:

The concept of wavelets has its origins in many fields, and part of the accomplishment of Daubechies is finding those places where the concept arose and showing how all the approaches relate to one another. The use of wavelets as an analytical tool is like Fourier analysis - simple and yet very powerful. In fact, wavelets are an extension of Fourier analysis to the case of localization in both frequency and space. ...

She has been awarded many other prizes and received many honours for her remarkable achievements. In 2000, the National Academy of Sciense made thier four yearly Award in Mathematics to her *for fundamental discoveries on wavelets and wavelet expansions and for her role in making wavelets methods a practical basic tool of applied mathematics.*

After Daubechies' unifying work, there have been many developments in various branches of wavelet analysis by various researchers and groups of researchers. In brief, one can mention, Donoho and collaborators in statistics, W. Dahmen in numerical analysis, Starck and collaborators in multiresolution analysis of astronomical data etc.

3 Data compression using wavelets

Data compression is one of the most important applications of the wavelet transform. It is usable across all disciplines and particularly applicable in fields with large datasets such as in astronomy, remote sensing, movies and other general databases. With this in mind, we briefly discuss here the general ideas behind compression schemes using wavelets.

The underlying idea of any compression scheme is to reduce or remove the correlation present in the data. When data points are correlated, one can predict the missing data on the basis of neighbouring data points. For example when data are spatially correlated, the value of a pixel in an image (for example) can be predicted based on it's neighbours' values. Similarly, correlation in the spectral (frequency component) or temporal component could be exploited for data compression. For example in video frames, the following video frame is very slightly different from the previous one.

Compression is implemented in two different ways : lossless and lossy. Lossless compression is used in the case where the reconstructed data are required to be exact copies of the original (for example in text data or computer binary executable programs). Lossy compression is used in the case of images where the reconstructed image is visually indistinguishable from the original. Much larger compression ratios are possible when one can accept some loss of information. The trick is to throw away information that does not contribute to the signifi-

Figure 2: Daubechies @ various levels of compression: original (left), reconstructed with 6% coefficients (center); reconstructed with 3% coefficients(right)

cant structure present in the data. Thus one way to achieve this method of compression is to use a wavelet transform to represent the data with a different mathematical basis where the correlations are apparent. In this new basis, the majority of the coefficients are small enough to be set to zero. Compression is thus achieved by taking the wavelet transform of the data, setting the coefficients below a threshold to zero, and losslessly encoding (compressing) the non-zero coefficients. The threshold is evaluated by a knowledge of the noise characteristics of the data.

An example of image compression and reconstruction using the Daubechies family of wavelets on a picture of Ingrid Daubechies is shown in figure 2.

4 Wavelets in Astronomy & Astrophysics

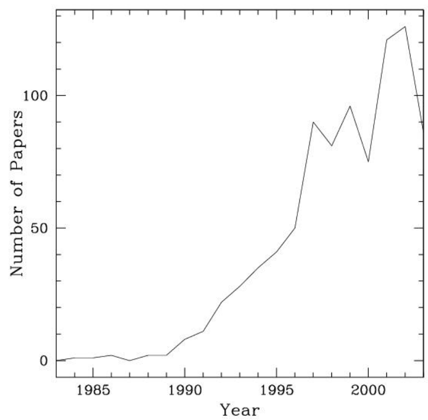

Figure 3: (see text)

A bibliographical search on the NASA Astronomy Data Service (`http://ads.iucaa.ernet.in/`) reveals 878 papers (in astronomical literature, including a few geophysics publications) with the term *wavelet* occuring in the title or abstract of the paper. Figure 3 shows the historical (as on 30^{th} October 2003) record of the usage of wavelets by the astronomical community. Presently over a hundred papers are being published per year. Data for 2003 (and possibly other recent years) are incomplete.

Reviewing the 878 titles we find the following topics of astronomical research where wavelets are being used (listed in no particular order)

- solar astronomy
- Cosmology - particularly Cosmic Microwave Background analysis
- Orbital mechanics of planetary influences on satellites of neighbouring planets
- Gamma Ray Bursts
- ROSAT source catalogging
- ISO data calibration and analysis
- Galaxy clustering
- Quasar detection in large scale surveys
- Data compression
- Pulsating stars
- Martian surface mapping
- Radio astronomy map denoising

- Morphological studies of spiral structure in external galaxies
- Molecular clouds
- Cometary jets (Hale-Bopp)
- Topology of the Universe
- Fractal structure in astronomical data
- Blazars and AGN
- ...

The above list shows that practically all aspects of modern astronomical research are being influenced by wavelet analysis techniques.

The most important contribution wavelets are making to astronomical data analysis is in the analysis of time series data. Using wavelets it is now possible not only to determine periodicities in the data but also determine the time at which these periodicities exist. In other words, it is now possible to study time dependence of the variability. Previously one used fourier analysis to determine the frequencies present in the data; but this did not allow for accurately pinpointing the time at which the frequencies were present (except in a poor resolution way using the windowed fourier transform). Most of the wavelet applications in the study of the sun, pulsating (variable) stars, variable sources such as blazars and other active galactic nuclei, is in the analysis of the light curves (study of how the intensity changes with time).

The next major application of wavelets is in the study of spectra. In this case the application is similar to that of light curve analysis but one is working with data that are spectral in nature. The application of wavelet analysis here is mostly to detect real spectral features and to suppress the background noise. The other application is in the usage of wavelet analysis at different scales in order to classify spectra automatically. This method is being used in the identification of quasars in the 2dF spectral survey. Of course the method can be used in characterizing non-quasar stellar sources as well, by using the appropriate coefficients.

Wavelets of the multiple resolution type are being used extensively in the calibration and analysis of modern astronomical instruments, as the procedures here are similar to the analysis of temporal or spectral domain data. As an example, multiple resolution analysis is being used to detect cosmic ray hits and to calibrate the highly complicated response of the ISOCAM mid-infrared detectors.

Images, from digital imaging devices (such as CCDs or array detectors) coupled to telescopes, form a large part of the astronomical dataset. Images at one wavelength are usually stored as two dimensional data files. Application of wavelets here is in noise removal and more importantly in the detection of features - either extended or point like. An important point where the wavelet based extraction scores over other traditional methods is in the fact that one does not need to know precisely the point spread function (PSF) of the system to apply wavelet analysis. When one has images of the same part of the sky at other wavelengths as well, then the data set could be treated as a multidimensional dataset. Wavelet analysis of such a data set is a very powerful technique to identify features in the images in different wavebands. Astronomical research requires that objects detected at different wavelengths be extracted and cataloged. The catalogs are then studied using different techniques to derive various physical parameters of the sources. Wavelet analysis plays an important role in building such catalogs - firstly in the extraction and then in the cross-identification of the same source at various wavelengths. In the following section I show a typical application of wavelet analysis in the detection of extended features in mid infrared images of selected areas of the Milky Way Galaxy.

5 Feature detection in astronomical images

Describing structures as a function of their characteristic scale is a precise property of the wavelet transform. This specificity has been used to define a *Multi-scale Vision Model (MVM)* to detect morphological features in images.

5.1 Mathematical underpinnings of Multi-scale Vision Model

The Multiscale transform of an image by the *á trous* algorithm produces a new set $\{w_j\}$ at each scale j. This set has the same number of pixels as the original image. In other words, it is a redundant transform. The original image I can be expressed as the sum of all the wavelet scales and a smoothed array C_J. This is mathematically represented as

$$I(k,l) \; = \; C_{J,k,l} \; + \; \sum_{j=1}^{J} \; w_{j,k,l}. \tag{4}$$

for each pixel location specified by (k,l). Thus each pixel of the input image is associated with a set of pixels in the wavelet multiscale transform domain. An image I of an astronomical subject can be decomposed into a series of subsets

$$I(k,l) \ = \ \sum_{i=1}^{n} O_i(k,l) + B(k,l) + N(k,l); \tag{5}$$

where n is the number of objects, O_i are the objects (stars, galaxies etc) in the data, B is the background image and N is the noise.

To setup such a decomposition, one has to detect, extract and measure the significant structures in an image. To achieve this, one computes the multi-resolution transform of the image and applies a segmentation scale by scale. The wavelet space of a 2D image is a three-dimensional one. The object is to be defined in this three-dimensional wavelet space.

Based on this principle, Bijaoui & Rué described an object as a hierarchical set of structures. They defined several parameters to set up their Multiscale Vision Model. Below, I summarize the method of identification of objects (for details see Starck & Murtagh, 2002) in a given image I:

1. compute the redundant wavelet transform;
2. determine the noise standard deviation;
3. deduce threshold at each scale based on the noise modeling;
4. threshold scale-by-scale and do a labeling of the objects;
5. determine the interscale relations;
6. identify all the wavelet coefficient maxima of the wavelet transform space;
7. extract the connected trees resulting from each wavelet transform space

5.2 Application to an image of the milky way

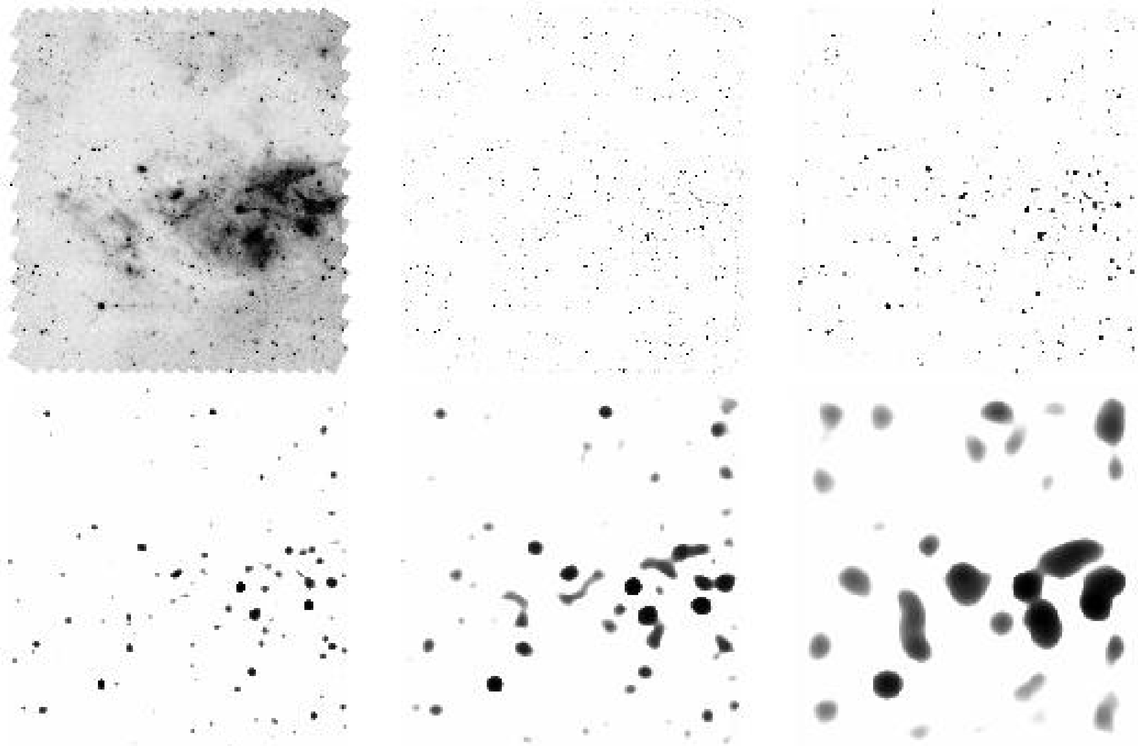

Figure 4: Multiscale wavelet transform of a mid infrared image of the Sgr B starforming region

In figure 4, I show an image of the Sagittarius B starforming region obtained by the Infrared Space Observatory at mid-infrared wavelengths. The corresponding redundant wavelet transform sets are also shown for different scales. The distinction between the point sources and the low frequency extended emission regions is quite apparent. Using this subset of the wavelet transform space of this image, I reconstruct back an image in which the features stand out quite sharply (see figure 5)[2] In both figures, the image is shown in reverse gray

[2]It is also available on the web at *http://www.prl.res.in/~shashi/tmw/sgrB_wavelet.png*

scale (bright features, such as stars, are represented by increasing levels of blackness) with the same colour tables for all frames.

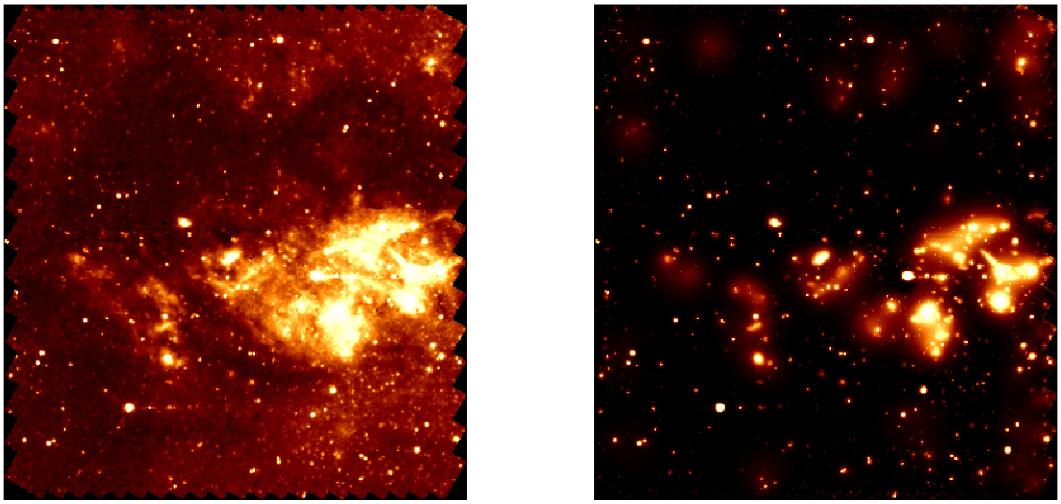

Figure 5: Reconstructed image of the Sgr B region using the multiscale wavelet transform shown in figure 4. An enlarged, false coloured version is attached at the end of this report.

In the rest of this section, I quote an extract from a paper (Hennebelle et al, 2001) in which we made use of MVM wavelet analysis to catalog the dark clouds seen in the ISOCAM images of the Milky Way galaxy. This catalog was used to place constraints on the interstellar extinction curve in the mid-infrared.

As illustrated in figure 6 the structures seen in absorption present various spatial scales, a complex morphology and various absorption contrasts. Galactic latitude gradient and diffuse emission structures produce significant variations in the background, at the scale of an individual image. Point sources or extended sources add up to this contrasted background and to the inherent complexity of these structures, preventing simple algorithms from properly extracting the dark structures. Although sophisticated multiscale analysis tools are appropriate for complex extraction, they are not designed for images with both positive and negative structures. The following strategy was thus adopted:

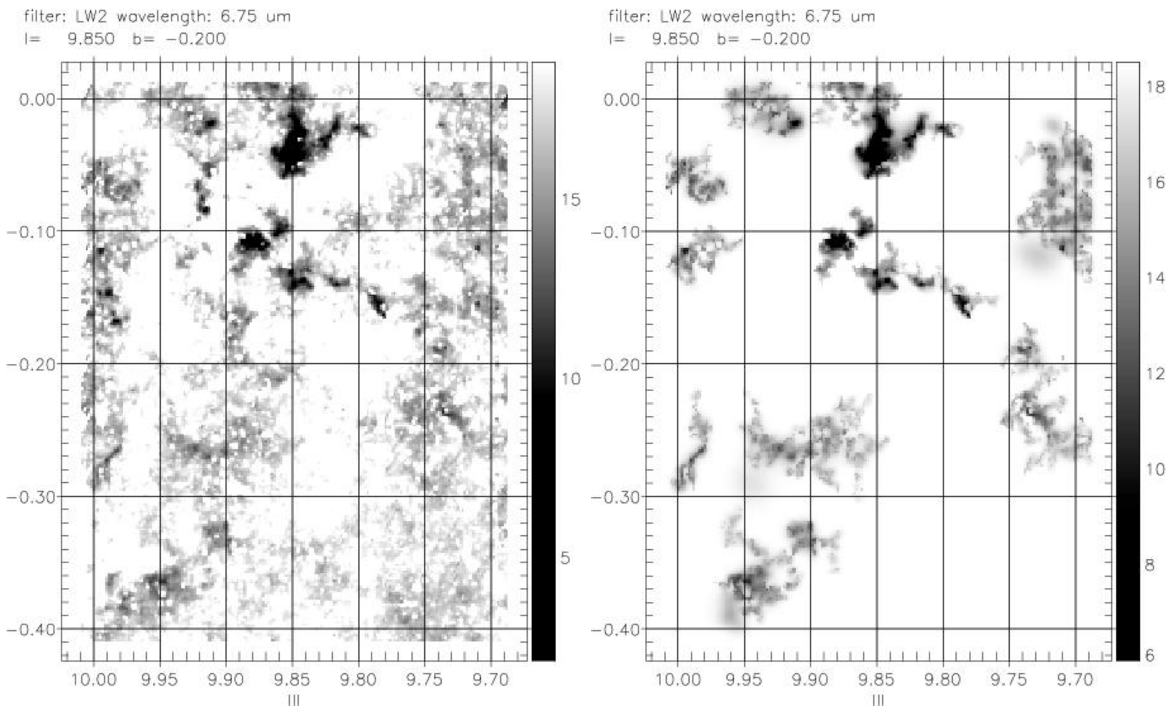

Figure 6: Image of the field $\ell = 9.85,\ b = -0.2$ with the bright (positive) features clipped to zero (left). Reconstructed image using MVM showing the dark clouds (right)

- clipping of all positive structures to a local background estimate obtained with a median filter;
- multi-scale (wavelet) analysis of the negative structures using the MVM package;
- object selection according to a minimal contrast criterion;
- cross identification of the 7 and 15μm list of objects.

The wavelet retained is a linear B-spline. Most of the reconstructed objects correspond to diffuse patches of moderate absorption or to emission minima and are not relevant for the present study. We thus only keep the objects for which the contrast (ratio of the intensity deficit to the background intensity), C, reaches at least 15% at one position of the object. The resulting lists at 7 and 15μm are then cross-identified on the basis of spatial concidence.

Diffuse and more compact structures, partly connected, are present in the reconstructed image. The reconstructed image obtained after performing all steps of MVM and applying the minimal contrast criterion, is displayed in right panel of figure 6. All significant structures, mostly in the upper part of the map, have been properly reconstructed. The isolated diffuse structures have been eliminated by the minimal contrast criterion whereas the diffuse parts spatially connected to the denser ones remain. Small structures at scale 1, have not been reconstructed either. One of the limitations of MVM is that, using square wavelets, narrow filaments are detected as lines of scale 1 elements, and thus fall off the catalogue.

It is satisfying to note that all large and most contrasted features, which represent the target of the present study, are similarly extracted in both images. Another limitation of our "local" detection is that dark objects larger than the size of the images (10-30') will escape detection. The risk also exists that a background area in the neighbourhood of a bright emissive region may be mistaken for a dark object. These false detections however will pop up at a later stage of the analysis (which is beyond the scope of this report)[3]

[3] This is not the end of the story of *wavelets in astronomy* but only the beginning...

6 Wavelets online

6.1 The online wavelet community

http://www.wavelet.org/ is the location of the wavelet digest - an online community of waveleticians. Includes useful links to introductory material, tutorials, demos, software. Archived mailing list of day to day postings by members. An equivalent site for francophones is at *http://www.ondelette.com/*; *Ondelette* is french for *wavelet*.

6.2 Personal webpages of waveleticians

- Paul Addison *http://sbe.napier.ac.uk/staff/paddison/*

  Author of book *The illustrated Wavelet Transform Handbook*. First three chapters are downloadable from here.

- Albert Bijaoui *http://www.obs-nice.fr/paper/bijaoui/EnglishMVM.html*

  Author of the MVM technique. Also co-author on several books and nearly 100 papers on astronomical analysis techniques including wavelets

- V Clemens *http://perso.wanadoo.fr/polyvalens/clemens/*

  webpage on *A Really Friendly Tutorial To Wavelets*

- Ingrid Daubechies *http://www.math.princeton.edu/~icd/*

  Includes links to her list of publications with some papers available online.

- Amara Graps *http://www.amara.com/*

  The one-stop collection of links to wavelet resources on the www!

- Palle Jorgensen *http://www.math.uiowa.edu/~jorgen/*

  Co-author of *Wavelets through a Looking Glass: The World of the Spectrum*

- Stephane Mallat *http://www.cmap.polytechnique.fr/~mallat/*

  Waveletician and author of *Wavelet tour of signal processing*

- Fionn Murtagh *http://www.cs.qub.ac.uk/~F.Murtagh/*

  Co-author of several books.

- Jean-Luc Starck *http://jstarck.free.fr/*

  Waveletician and co-author. Extensive work on ISO data calibration and analysis techniques using the multiple-resolution wavelet transforms.

6.3 Wavelet Software

- WaveLab *http://www-stat.stanford.edu/~wavelab*

  Built by David Donoho and collaborators, this is a package of matlab scripts for wavelet analysis. It is reported to run under the GNU equivalent of Matlab called *octave*. I have not verified this however.

- Wavelet Workbench *http://www.amara.com/wwbdev/wwbdev.html*

  This is an IDL port of the WaveLab software by Amara Graps. This software was used on a linux box to make figure 2. It has now evolved into the commercial package known as the IDL Wavelet Toolkit.

- IDL Wavelet Toolkit *http://www.rsinc.com*

  This is a commercial toolkit from IDL's developers. It requires a separate license apart from the IDL basic license. I tested this in evaluation mode.

- WaveThresh *http://www.stats.bris.ac.uk/~wavethresh/*

  This is a noncommercial software package for performing statistics based on wavelet techniques. It runs on the popular S-plus (commercial) statistics package. It works quite well with the GNU equivalent of S-plus (called 'R' - *http://www.r-project.org*). This package in combination with R on a linux PC was used to make figure 1.

- LastWave *http://www.cmap.polytechnique.fr/~bacry/LastWave/index.html*

  This is a signal processing oriented command language. I have not explored this package as yet. It is recommended on the webpages of Stephane Mallat's book.

- MVM

  This is a linux-only collection of binary programs that reads astronomical fits format data files and outputs wavelet transformed/reconstructed data sets (also in fits format). This was used in building the figures 4, 5 and 6.

- A Practical Guide to Wavelet Analysis *http://paos.colorado.edu/research/wavelets/* This site includes code in fortran, IDL and Matlab for wavelet analysis. It is also possible to make interactive wavelet plots using example data or data uploaded by user.

7 Further reading

In order to keep the report clear, I do not give explicit citations and cross-references in the above text. Below is a list of the references, books and web articles that serve as starting points in the exhaustive wavelet literature.

- *Lecture notes in Harmonic Analysis, Wavelets and Applications*, Daubechies, I., & Gilbert, A.
- *Where do Wavelets Come From? - A Personal Point of View*, Daubechies, I., 1996, IEEE article available on Daubechies' webpage
- *Ten lectures on wavelets*, Daubechies, I., SIAM, 1992
- *An introduction to wavelets*, Graps, A., IEEE 1995
- *Infrared dark clouds in the ISOGAL survey* Hennebelle, Perault, Teyssier, Ganesh, A&A, 2001, pp 598, vol 365
- *SIGGRAPH 1996 Notes on Wavelets*, SIGGRAPH http://www.multires.caltech.edu/teaching/courses/waveletcourse/
- *Astronomical Image and Data Analysis*, Starck, J.L., & Murtagh, F., Springer, 2002
- *Image Processing and Data Analysis*, Starck, J.L., Murtagh, F., Bijaoui, A., Cambridge, 1998
- *A Wavelet Tour of Signal Processing* Stephane Mallat, Academic Press 1999

This article is typeset in *PDF LaTeX*