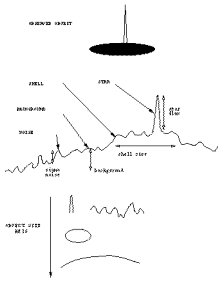
Wavelets and Multiresolution Transforms with Applications

Fionn Murtagh, M2

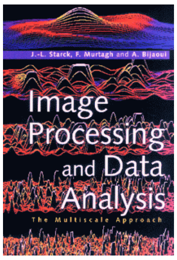
- Some theory and description of the algorithms
- Discussions and examples of: (i) visualization, (ii) noise filtering, (iii) deconvolution, (iv) compression

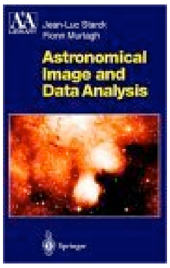
1

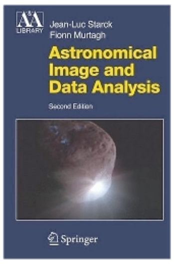
W h y   m u l t i r e s o l u t i o n ?



2







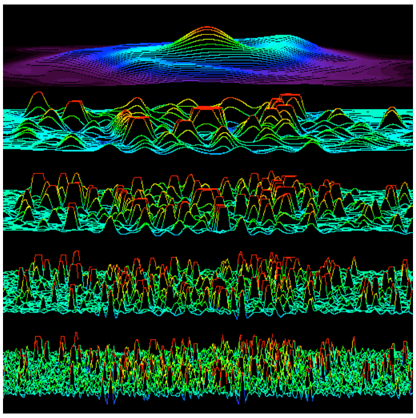
3

Some Examples of Wavelet Transforms for Visualization

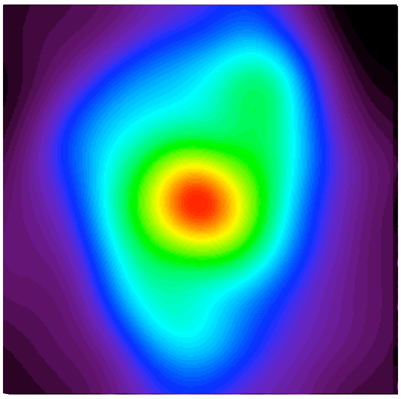
1) Visualization – comet core

2) Separating objects on different scales – galaxy vis-à-vis stars

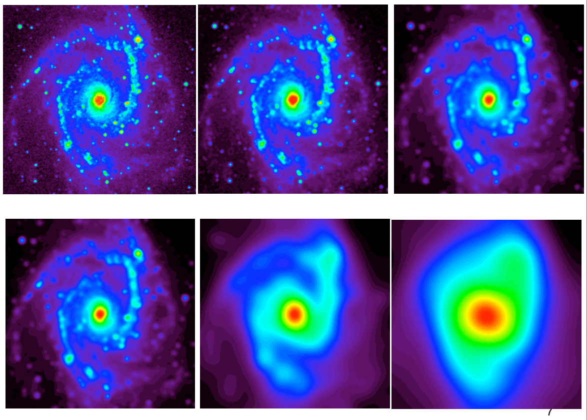
4



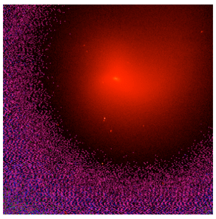
5

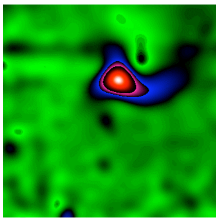


6



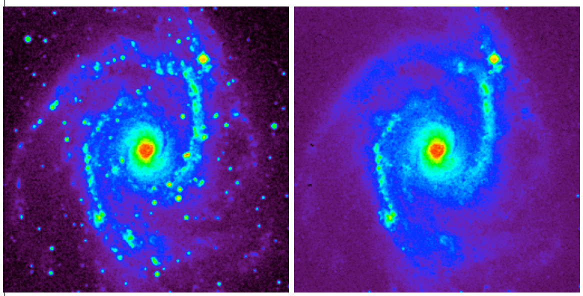
7





Left: Periodic comet P/Swift-Tuttle observed with the 1.2m telescope at Calar Alto Observatory in Spain in October and November 1992. (Data courtesy of Dr K Birkle, Max-Planck-Institut für Astronomie, Heidelberg). Irregularity of nucleus is indicative of the presence of jets in in the coma.
Right: Scale 5 of a B3 spline à trous wavelet transform.

8



Continuous wavelet transform

- A wavelet function is convolved with the signal
- Then it is dilated and again convolved with the signal
- So the signal is projected onto translated versions of the wavelet function, at each resolution level
- This scheme furnishes the continuous wavelet transform

The Continuous Wavelet Transform

$$W(a,b) = K \int_{-\infty}^{+\infty} \psi^*(\frac{x-b}{a}) f(x) dx$$

where:

- $W(a,b)$ is the wavelet coefficient of the function $f(x)$
- $\psi(x)$ is the analyzing wavelet
- a (> 0) is the scale parameter
- b is the position parameter

In Fourier space, we have: $\hat{W}(a,\nu) = \sqrt{a}\hat{f}(\nu)\hat{\psi}^*(a\nu)$
When the scale a varies, the filter $\hat{\psi}^*(a\nu)$ is only reduced or dilated while keeping the same pattern.

Properties

- CWT is a linear transformation:
  - if $f(x) = f_1(x) + f_2(x)$ then $W_f(a,b) = W_{f_1}(a,b) + W_{f_2}(a,b)$
  - if $f(x) = kf_1(x))$ then $W_f(a,b) = kW_{f_1}(a,b)$
- CWT is covariant under translation:

  if $f_0(x) = f(x - x_0)$ then $W_{f_0}(a,b) = W_f(a, b - x_0)$
- CWT is covariant under dilation:

  if $f_s(x) = f(sx)$ then $W_{f_s}(a,b) = \frac{1}{s^{\frac{1}{2}}} W_f(sa, sb)$

Whatever the scale and the position, the signal analysis is done using the same function.

The Inverse Transform

The inverse transform is:

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \frac{1}{\sqrt{a}} W(a,b) \psi\left(\frac{x-b}{a}\right) \frac{dadb}{a^2}$$

where

$$C_\psi = \int_{-\infty}^{+\infty} \mid \hat{\psi}(t) \mid^2 \frac{dt}{t} < +\infty$$

Reconstruction is only possible if C_ψ is defined (admissibility condition). This condition implies $\hat{\psi}(0) = 0$, i.e. the mean of the wavelet function is 0.

13

Morlet wavelet

- But usually a discrete wavelet transform is used, rather than the continuous wavelet transform
- For an n-length signal, through use of decimation, this means log2 n levels
- For a redundant wavelet transform, like the à trous wavelet transform - to be looked at later - a user-specified number of levels is used

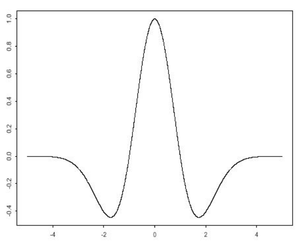
14

Mexican hat

The Mexican hat function is in one dimension:

$$g(x) = (1 - x^2) e^{-\frac{1}{2}x^2}$$

This is the second derivative of a Gaussian.



15

Mallat's multiresolution analysis

- Developed for image processing
- Uses wavelet and scaling functions, playing a complementary role
- Decimation is used: because the data is smoothed at successive stages, only one sample out of two is retained

16

Multiresolution Analysis

Multiresolution analysis (Mallat, 1989) results from the embedded subsets generated by the interpolations at different scales.

A function $f(x)$ is projected at each step j onto the subset V_j $(\ldots \subset V_{2^{j+1}} \subset V_{2^j} \subset V_{2^{j-1}} \subset V_{2^0} \ldots)$. This projection is defined by the scalar product $c_j(k)$ of $f(x)$ with the scaling function $\phi(x)$ which is dilated and translated:

$$c_j(x) \quad = \quad < f(x), \phi_j(x - 2^j k) >$$

$$\phi_j(x) = 2^j \phi(2^j x)$$

where $\phi(x)$ is the scaling function. ϕ is a low-pass filter.

Wavelets and Multiresolution Analysis

The difference between c_{j-1} and c_j is contained in the detail signal belonging to the space O_j orthogonal to V_j.

$$O_j \oplus V_j = V_{j-1}$$

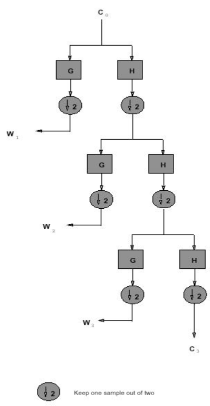
The set $\sqrt{2^{-j}}\psi_j(x - 2^{-j}k)_{k\in\mathcal{Z}}$ forms a basis of O_j.

$$\psi_j(x) = 2^j \psi(2^j x)$$

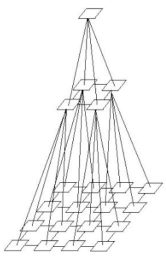
where $\psi(x)$ is the wavelet function.

The wavelet coefficients are obtained by:

$$w_j(x) \quad = \quad < f(x), \psi_j(x - 2^j k) >$$



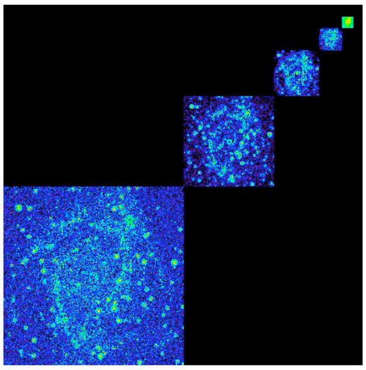
Pyramidal Algorithm



Several algorithms exist:

- The Laplacian Pyramid
- Pyramid with one wavelet
- Pyramidal wavelet transform using the FFT

NGC 2997 pyramidal coefficients



21

Example: the Haar Wavelet Transform

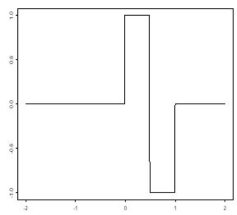
$$\psi(x) = 1 \quad \text{if } 0 \leq x < \tfrac{1}{2}$$
$$\psi(x) = -1 \quad \text{if } \tfrac{1}{2} \leq x < 1$$
$$\psi(x) = 0 \quad \text{otherwise}$$

And $\phi(x) = 1$ for $0 \leq x < 1$

Latter is low-pass (averaging) filter

ψ is band-pass (differencing) filter

Both are translated and dilated for the Haar WT



22

Input signal on top. Haar WT algorithm follows. Bottom line: Haar WT.

64	48	16	32	56	56	48	24
56	24	56	36	**8**	**-8**	**0**	**12**
40	46	**16**	**10**	**8**	**-8**	**0**	**12**
43	**-3**	**16**	**10**	**8**	**-8**	**0**	**12**

56 = average of 65 and 48; and +8 or -8 is needed to recreate inputs.
24 = average of 16 and 32; and -8 or -(-8) is needed to recreate inputs.
At scale 2, detail coefficients 8, -8, 0, 12 are just repeated.
At scale 3, detail coefficients 16, 10, 8, -8, 0, 12 are just repeated.

23

Sweldens' lifting scheme - a general wavelet transform algorithm

"... second generation wavelets, i.e. wavelets adapted to situations that do not allow translation and dilation like non-Euclidean spaces. ... A construction using lifting ... is entirely spatial and therefore ideally suited for building second generation wavelets when Fourier techniques are no longer available."

("Factoring wavelet transforms into lifting steps", I Daubechies and W Sweldens, http://cm.bell-labs.com/who/wim/papers/factor/factor.pdf
J. Fourier Anal. Appl., Vol. 4, Nr. 3, pp. 247-269, 1998)

24

Sweldens' lifting scheme - a general wavelet transform algorithm

"... second generation wavelets, i.e. wavelets adapted to situations that do not allow translation and dilation like non-Euclidean spaces. ... A construction using lifting ... is entirely spatial and therefore ideally suited for building second generation wavelets when Fourier techniques are no longer available."

("Factoring wavelet transforms into lifting steps", I Daubechies and W Sweldens, http://cm.bell-labs.com/who/wim/papers/factor/factor.pdf J. Fourier Anal. Appl., Vol. 4, Nr. 3, pp. 247-269, 1998)

Other important benefits:
- wavelet transform coefficients are all integers - which is very important for compression
- lifting scheme algorithm can be found for all wavelets

25

Haar algorithm - traditional

At any given step, we have two successive signal values, that we will term a, b

Smooth: s = (a + b)/2

Detail:

+d = (a + b)/2 - a = (b - a)/2

- d = (a + b)/2 - b = (a - b)/2

Reconstruction, then:

a = s - d

b = s + d

26

Haar algorithm - lifting

At any given step, we have two successive signal values, that we will term a, b

Smooth: s = a + b

Detail:

+d = b - a

- d = a - b

Reconstruction, then:

a = s/2 - d/2

b = s/2 + d/2

Big advantage: all wavelet coefficients are integer.

27

Lifting - general principle

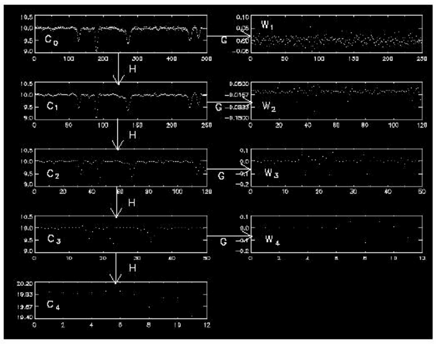
- Split a signal into its polyphase components: even indexed samples, and odd indexed samples. Typically they are correlated.
- Hence given one, e.g. odd, build a good predictor, P, for the other.
- odd set - P(even set) = detail
- In the Haar case, P(even set) is just the even set!
- Rewriting the above expression for 'detail': $d = x_o - P(x_e)$
- Given the even and d, we can recover the odd
- At each step we have a predictor operator, and an update operator
- Lifting has been used to develop the wavelet transform on the sphere

28

A few remarks…

- On 2D images, the algorithm may be carried out in a genuinely 2D way, but this may have a computational overhead. Separability (in x and y) is often used instead. The same principle applies for 3D image volumes.
- Because the signal is increasingly smooth, we can keep one sample out of two. This is called decimation. It is crucial in compression applications. Exact reconstruction is guaranteed. But for signal detection type applications, it can lead to aliasing if we interpret/analyze the data in the wavelet transform space.
- For general wavelet functions, we need to reverse the effects of transforming, when carrying out the inverse transform. The Haar WT is orthogonal and leads to a particularly simple algorithm. A more general algorithm is the biorthogonal WT.

29



Reconstruction with bi-orthogonal wavelets

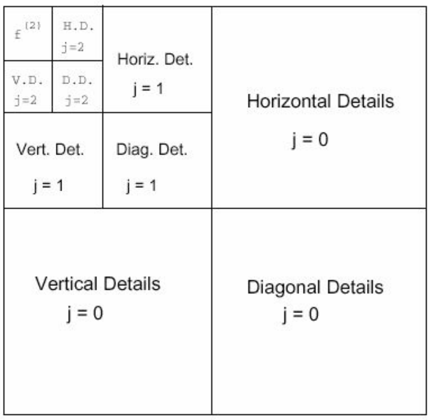
Two other filters $\tilde{h}$ and $\tilde{g}$ are used, defined to be conjugate to h and g. The reconstruction of the signal is performed with:

$$c_j(k) = 2\sum_l [c_{j+1}(l)\tilde{h}(k-2l) + w_{j+1}(l)\tilde{g}(k-2l)]$$

In order to get an exact reconstruction, two conditions are required for the conjugate filters:

- *Dealiasing condition*: $\hat{h}(\nu + \frac{1}{2})\hat{\tilde{h}}(\nu) + \hat{g}(\nu + \frac{1}{2})\hat{\tilde{g}}(\nu) = 0$
- *Exact restoration:* $\hat{h}(\nu)\hat{\tilde{h}}(\nu) + \hat{g}(\nu)\hat{\tilde{g}}(\nu) = 1$

31



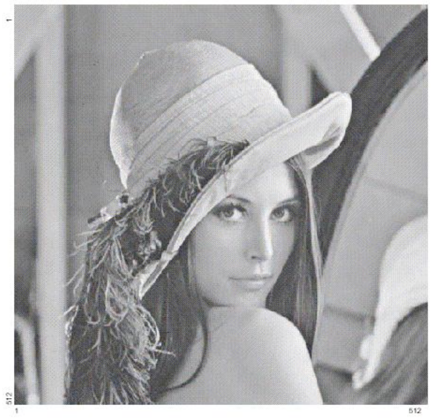
32

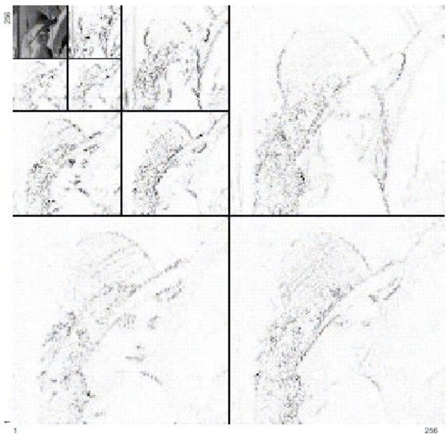
In two dimensions, we separate the variables x, y:

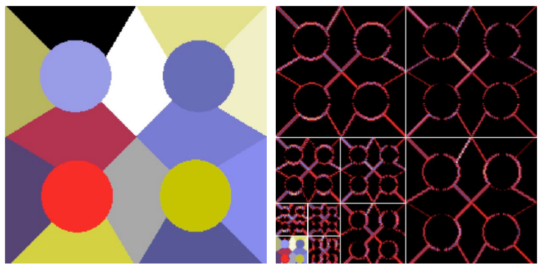
- vertical wavelet: $\psi^1(x,y) = \phi(x)\psi(y)$
- horizontal wavelet: $\psi^2(x,y) = \psi(x)\phi(y)$
- diagonal wavelet: $\psi^3(x,y) = \psi(x)\psi(y)$

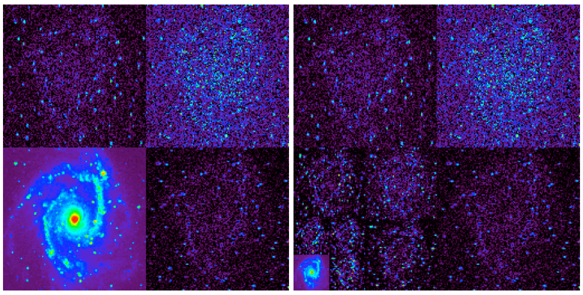
The detail signal is contained in three sub-images:

$$
\begin{aligned}
w_j^1(k_x, k_y) &= \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} g(l_x - 2k_x) h(l_y - 2k_y) c_{j+1}(l_x, l_y) \\
w_j^2(k_x, k_y) &= \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} h(l_x - 2k_x) g(l_y - 2k_y) c_{j+1}(l_x, l_y) \\
w_j^3(k_x, k_y) &= \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} g(l_x - 2k_x) g(l_y - 2k_y) c_{j+1}(l_x, l_y)
\end{aligned}
$$









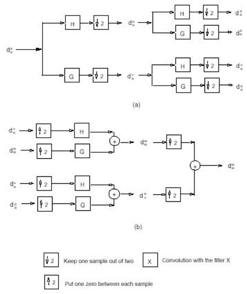
Redundant scheme, 3 wavelet scales

Mallat scheme, biorthogonal wavelets

37

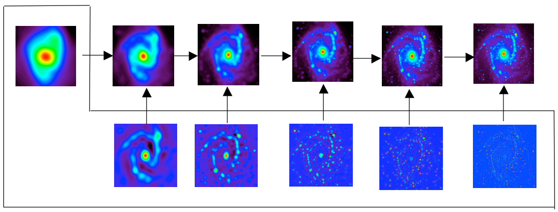
Wavelet Packet

Wavelet packets were introduced by Coifman, Meyer and Wickerhauser (1992). Instead of dividing only the approximation space, as in the standard orthogonal wavelet transform, detail spaces are also divided.



Wavelet packet Scheme: (a) forward direction, (b) reconstruction.

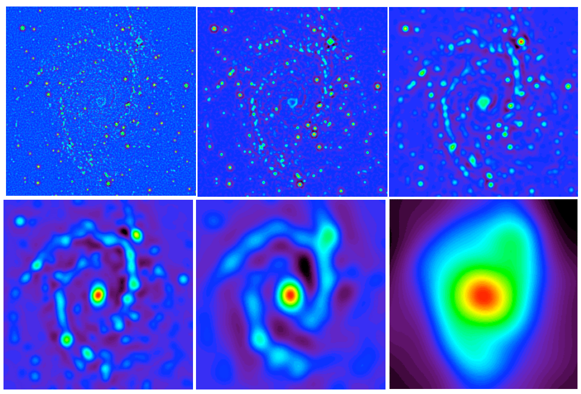
38



Redundant wavelet transform scheme.

We will in particular look at the
B3 spline à trous wavelet transform.

39



A trous wavelet transform

- A redundant transform: this means that there are as many values at each resolution level as there are values to begin with
- Often used in astronomy
- Wavelet function, and associated smoothing function, are "isotropic" or point symmetric

41

A trous wavelet transform

- Take the case of a 1D signal
- Use the B3 spline wavelet transform
- Step 1: convolve signal, s0, with filter (1/16, 1/4, 3/8, 1/4, 1/16). This gives first smoothed signal, s1
- First set of wavelet coefficients are: w1 = s0 - s1
- Smooth s1 again using above filter; skip one value between every two. This gives second smoothed signal, s2
- Second set of wavelet coefficients are: w2 = s2 - s1

42

A trous wavelet transform

Continue. Finally:

w1 = s1 - s0

w2 = s2 - s1

w3 = s3 - s2

w4 = s4 - s3

---- ----------

$\sum_j$ wj = - s0 + s4

Or: s0 = $\sum_j$ wj + s4

This is a decomposition of the original signal in terms of its wavelet coefficients, and the final smooth

43

The à trous Algorithm

This is a "stationary" or redundant transform, i.e. decimation is not carried out. The distance between samples increasing by a factor 2 from scale $(j-1)$ $(j>0)$ to the next, $c_j(k)$, is given by:

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

and the discrete wavelet coefficients by:

$$w_j(k) = \sum_l g(l) c_{j-1}(k + 2^{j-1} l)$$

Generally, the wavelet resulting from the difference between two successive approximations is applied:

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

The associated wavelet is $\psi(x)$.

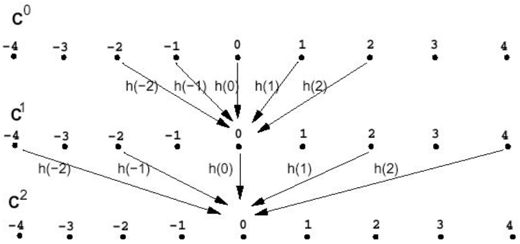
$$\frac{1}{2}\psi(\frac{x}{2}) = \phi(x) - \frac{1}{2}\phi(\frac{x}{2})$$

The reconstruction algorithm is immediate:

$$c_0(k) = c_{n_p}(k) + \sum_{j=1}^{n_p} w_j(k)$$

44

Passage from c_0 to c_1, and from c_1 to c_2

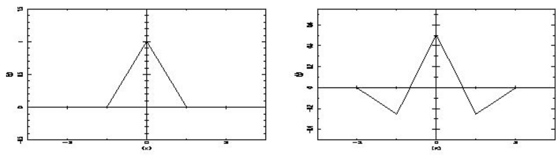


45

The triangle function

Choosing the triangle function as the scaling function ϕ leads to piecewise linear interpolation:

$$\phi(x) = 1- \mid x \mid \quad \text{if } x \in [-1, 1]$$
$$\phi(x) = 0 \quad \text{if } x \notin [-1, 1]$$



46

We have

$$\frac{1}{2}\phi(\frac{x}{2}) = \frac{1}{4}\phi(x+1) + \frac{1}{2}\phi(x) + \frac{1}{4}\phi(x-1)$$

c_{j+1} is obtained from c_j by:

$$c_{j+1}(k) = \frac{1}{4}c_j(k-2^j) + \frac{1}{2}c_j(k) + \frac{1}{4}c_j(k+2^j)$$

The wavelet coefficients at scale j are:

$$w_{j+1}(k) = -\frac{1}{4}c_j(k-2^j) + \frac{1}{2}c_j(k) - \frac{1}{4}c_j(k+2^j)$$

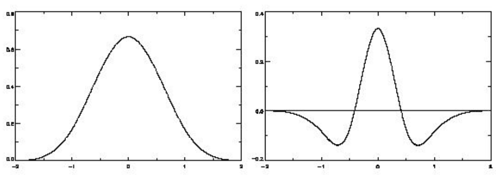
47

The cubic spline function

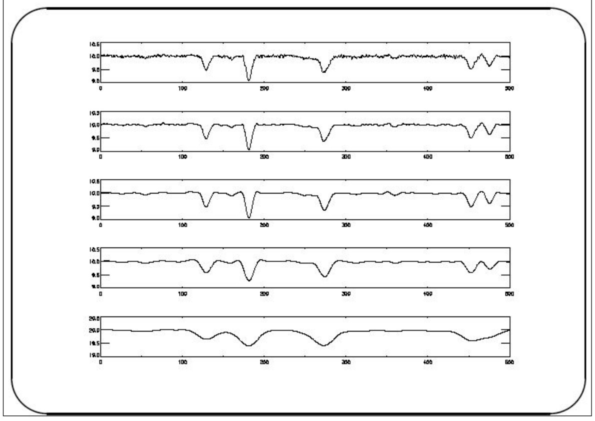
$$B_3(x) = \frac{1}{12}(\mid x-2 \mid^3 -4 \mid x-1 \mid^3 +6 \mid x \mid^3 -4 \mid x+1 \mid^3 + \mid x+2 \mid^3)$$

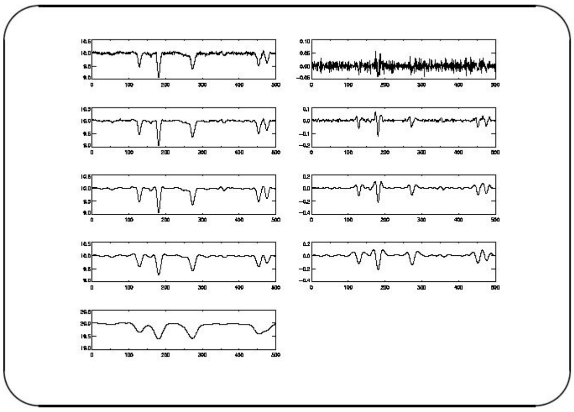
and

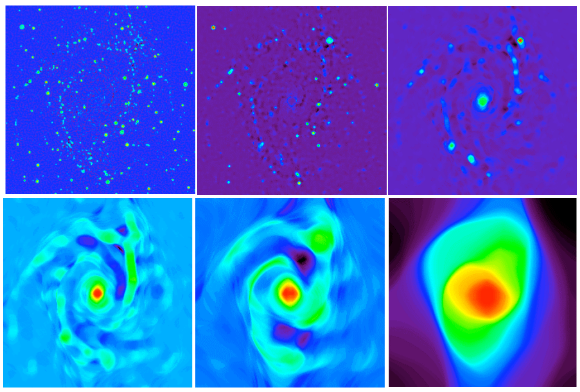
$$h(-2) = \frac{1}{16}; h(-1) = \frac{1}{4}; h(0) = \frac{3}{8}; h(1) = \frac{1}{4}; h(2) = \frac{1}{16}$$



48







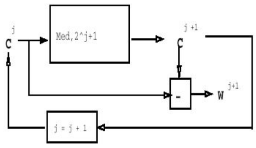
Multiresolution median transform

- The principle used in the wavelet transform can be generalized
- For example, the median transform has good properties - it is robust relative to extreme values
- So instead of, say, the B3 spline smoothing function, we use a median filter
- For the second resolution level dilate the median filter
- Actually: it would be better, computationally, to decimate the data and keep the median filter fixed!

52

Multiresolution Median Transform

1. Let $c_j = I$ with $j = 1$
2. Determine $c_{j+1} = med(f, 2s + 1)$.
3. The multiresolution coefficients w_{j+1} are defined as: $w_{j+1} = c_j - c_{j+1}$.
4. Let $j \longleftarrow j + 1$; $s \longleftarrow 2s$. Return to step 2 if $j < S$.



Reconstruction by: $I = c_p + \sum_j w_j$

53

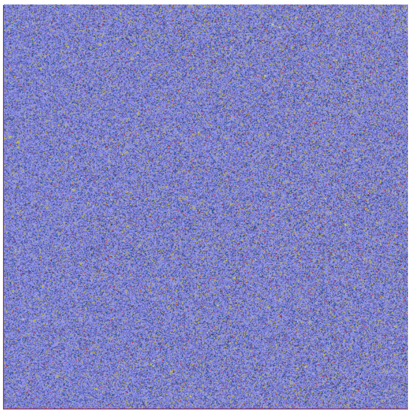
Pyramidal Median Transform

1. Let $c_j = I$ with $j = 1$.
2. Determine $c^*_{j+1} = med(c_j, 2s + 1)$ with $s = 1$.
3. The pyramidal multiresolution coefficients w_{j+1} are defined as:

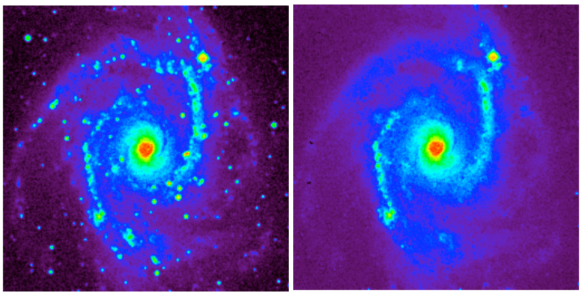
$$w_{j+1} = c_j - c^*_{j+1}$$

4. Let $c_{j+1} = dec(c^*_{j+1})$ (where the decimation operation, dec, entails 1 pixel replacing each 2×2 subimage).
5. Let $j \longleftarrow j + 1$. Return to step 2 so long as $j < S$.

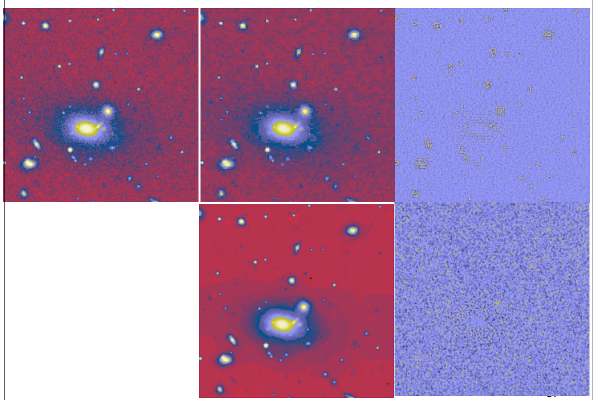
54

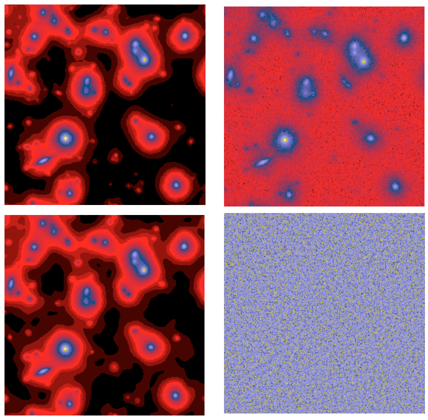


55

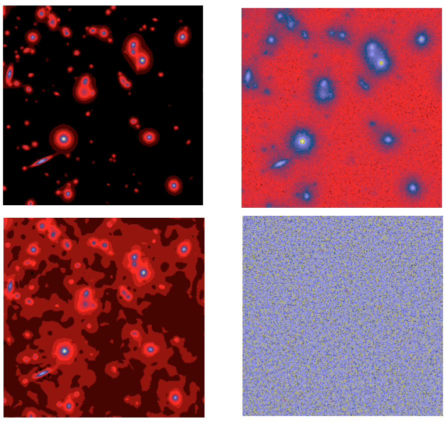


56





58



59

Noise in images and signals

- Photon counting detectors may be modelled using the Poisson distribution
- Note: the Poisson distribution becomes the Gaussian distribution when the number of “counts per bin” is sufficiently high, e.g. 30
- Also the Poisson is unreliable when there are very few counts per bin (e.g. often the case for X-ray data)

60

Gaussian noise

- Often the Gaussian distribution is a good one to use, if there is insufficient evidence in favour of any alternative
- CCDs, charged coupled devices, are widely used as digital detectors in astronomy
- CCDs have various noise components: a major one is “read-out noise”
- This is due to random, stray electrons being read

61

Gaussian noise - II

- Principle of operation of a CCD detector is as follows
- Photons arrive and eject electrons
- These electrons are read off, from the detector, in horizontal and vertical directions
- “Data numbers” are defined, proportional to the numbers of electrons read
- These data numbers give the pixel values

62

Noise model

Usual image (or signal) model is an additive one:

observed image = ideal image + noise

But there are many more complicated noise models

We will look at both (1) noise filtering, and also at (2) deconvolution. For the latter:

observed image = blur * ideal image + noise

63

Noise of wavelet coefficients

- It is often the case that we can say the following:
- The wavelet coefficients are additive transforms of the original data values
- Gaussian data remains Gaussian under an additive transformation
- Obviously the wavelet coefficients are of 0 mean; and their standard deviation changes of course, compared to std. dev. of original data

64

Noise filtering in wavelet space

- So this gives us a very simple principle for noise filtering in wavelet space
- The wavelet coefficients are "sparse" - bigger and smaller relative to the initial data values - so can (very often) be filtered better than any alternative of filtering the original data
- Gaussian properties "propagated" through to the wavelet resolution levels are nice and easy to filter - using e.g. a 3 sigma detection (i.e. = "real signal") threshold

65

Gaussian Noise

$$p(w_j(x,y)) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-w_j(x,y)^2/2\sigma_j^2}$$

Rejection of hypothesis $\mathcal{H}_0$ depends (for a positive coefficient value) on:

$$P = Prob(w_j(x,y) > W) = \frac{1}{\sqrt{2\pi}\sigma_j} \int_{w_j(x,y)}^{+\infty} e^{-W^2/2\sigma_j^2} dW$$

and if the coefficient value is negative, it depends on

$$P = Prob(w_j(x,y) < W) = \frac{1}{\sqrt{2\pi}\sigma_j} \int_{-\infty}^{w_j(x,y)} e^{-W^2/2\sigma_j^2} dW$$

Given stationary Gaussian noise, it suffices to compare $w_j(x,y)$ to $k\sigma_j$.

if $\mid w_j \mid \geq k\sigma_j$ then w_j is significant

if $\mid w_j \mid < k\sigma_j$ then w_j is not significant

66

Noise Modeling

We can introduce a statistical significance test for wavelet coefficients. Let $\mathcal{H}_0$ be the hypothesis that the image is locally constant at scale j. Rejection of hypothesis $\mathcal{H}_0$ depends (for a positive coefficient value) on:

$$P = Prob(W_N > w_j(x,y))$$

and if the coefficient value is negative

$$P = Prob(W_N < w_j(x,y))$$

Given a threshold, ϵ, if $P > \epsilon$ 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 only to the noise alone, and so the null hypothesis is rejected. In this case, a significant coefficient has been detected.

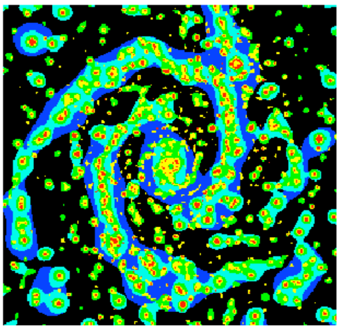
67

Multiresolution Support

- A "mask" image, maybe at each resolution scale, to "protect" the astronomical objects
- So this allows us to filter the noise in unimportant or less important parts of the image

68

NGC2997 MULTIRESOLUTION SUPPORT



69

Multiresolution Support

The *multiresolution support* M of an image is computed as follows:

- 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.

$M^{(I)}(j, x, y) = 1$ (or $= \ true$) $\Longrightarrow I$ contains information at scale j and at the position (x, y).
Remark:
Incorporating a priori knowledge: e.g. no interesting object smaller or larger than a given size in our image. Suppress, in the support, anything which is due to that kind of object. Can use mathematical morphology to do this.

70

Noise Modeling in Wavelet Space

Our noise modeling in wavelet space is based on the assumption that the noise in the data follows a statistical distribution, which can be:

- a Gaussian distribution
- a Poisson distribution
- a Poisson + Gaussian distribution (CCD noise)
- Poisson noise with few events (galaxy counting, X-ray images, ...)
- Speckle noise
- Root Mean Square map: we have a noise standard deviation of each data value.

71

Noise modeling with less constraint on the noise behavior

If the noise doesn't follows any of these distributions, we can derive a noise model from any of the following assumptions:

- it is additive, and non-stationary.
- it is multiplicative and stationary.
- it is multiplicative, but non-stationary.
- it is undefined but stationary.

If none of these assumptions can be applied, the only way to derive a noise model in wavelet space is to consider that the local variation of the wavelet coefficient follows a Gaussian distribution.

72

Poisson Noise with Few Photons

A wavelet coefficient at a given position and at a given scale j is

$$w_j(x, y) = \sum_{k \in K} n_k \psi\left(\frac{x_k - x}{2^j}\frac{y_k - y}{2^j}\right)$$

where K is the support of the wavelet function ψ and n_k is the number of events which contribute to the calculation of $w_j(x, y)$.

If a wavelet coefficient $w_j(x, y)$ is due to the noise, it can be considered as a realization of the sum $\sum_{k \in K} n_k$ of independent random variables with the same statistical distribution as is given by the wavelet function. Then we compare the wavelet coefficient of the data to the values which can be taken by the sum of n independent variables.

The distribution of one event in wavelet space is then directly given by the histogram H_1 of the wavelet ψ. Assuming independent events, the distribution of a coefficient W_n related to n events is given by n autoconvolutions of H_1

$$H_n = H_1 \otimes H_1 \otimes ... \otimes H_1$$

73

Other Types of Noise

1. Additive non-stationary noise: $R_\sigma(x, y)$ is calculated from the local variation in the image.
2. Multiplicative noise: the image is transformed by taking its logarithm. In the resulting image, the noise is additive, and a hypothesis of Gaussian noise can be used.
3. Multiplicative non-stationary noise: we take the logarithm of the image, and the resulting image is treated as for additive non-uniform noise above.
4. Undefined uniform noise: A k-sigma clipping is applied at each scale in order to find σ_j.
5. Undefined noise: The standard deviation is estimated for each wavelet coefficient, by considering a box around it, and the calculation of σ is done in the same way as for non-uniform additive noise.

74

Finding sigma at successive resolution levels

- One way is to simulate and have predefined values
- For filtering, we can even iteratively refine our estimate of the noise in the image
- I.e. assume value of noise -- sigma or standard deviation; carry out filtering in wavelet transform space; reconstruct the filtered data
- Then: subtract the filtered data from the original data; analyze the noise: is it close to the initially assumed noise parameter, sigma?

75

Next: variance stabilization

- What this means is performing some appropriate transformation on the original signal, such that Poisson, or a mixture of Poisson and Gaussian, distributed data becomes “tantamount” to Gaussian
- This makes the variance constant -- hence the name “variance stabilization”

76

Poisson Noise

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

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

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.

Poisson Noise + Gaussian

The generalization of the variance stabilizing is:

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

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

Filtering

- Hard thresholding
- Soft thresholding
- Donoho universal approach
- Multiresolution Wiener filtering
- Hierarchical Wiener filtering
- Hierarchical hard thresholding
- Iterative filtering from the multiresolution support
- Multiscale entropy filtering

Hard and soft thresholding

- Hard thresholding:

$$\begin{aligned} \tilde{w}_j &= w_j \text{ if } \mid w_j \mid \geq T_j \\ &= 0 \text{ otherwise} \end{aligned}$$

  with $T_j = k\sigma_j$. For an energy-normalized wavelet transform algorithm, we have $\sigma_j = \sigma$ for all j.

- Soft thresholding:

$$\begin{aligned} \tilde{w}_j &= sgn(w_j)(\mid w_j \mid - T_j) \text{ if } \mid w_j \mid \geq T_j \\ &= 0 \text{ otherwise} \end{aligned}$$

- Donoho approach:
  $T_j = \sqrt{2\log(n)}\sigma_j$ (where n is the number of pixels) instead of the standard $k\sigma$ value. This leads to a new soft and hard thresholding approach.

Filtering from the multiresolution support

The filtering can be seen as an inverse problem. Indeed, we want to reconstruct an image from the detected wavelet coefficient. The problem of reconstruction consists of searching for a signal $\tilde{I}$ such that its wavelet coefficients are the same as those of the detected structure. Denoting $\mathcal{T}$ the wavelet transform operator, and P the projection operator in the subspace of the detected coefficients (i.e. setting to zero all coefficients at scales and positions where nothing where detected), the solution is found by minimization of

$$J(\tilde{I}) = \parallel W - (P \circ \mathcal{T})\mathcal{S} \parallel$$

where W represents the detected wavelet coefficients of the image I.

Algorithm

The residual at iteration n is:

$$R^{(n)}(x,y) = I(x,y) - \tilde{I}^{(n)}(x,y)$$

By using the *à trous* wavelet transform algorithm:

$$R^{(n)}(x,y) = c_p(x,y) + \sum_{j=1}^{p} w_j(x,y)$$

Significant residual pixels are determined:

$$\bar{R}^{(n)}(x,y) = c_p(x,y) + \sum_{j=1}^{p} M(j,x,y) w_j(x,y)$$

and

$$\tilde{I}^{(n+1)} = \tilde{I}^{(n)} + \bar{R}^{(n)}$$

Multiscale Entropy Filtering method

The multiscale entropy method consists of measuring the information h relative to wavelet coefficients, and of separating this into two parts h_s, and h_n. The expression h_s is called the signal information and represents the part of h which is certainly not contaminated by the noise. The expression h_n is called the "noise information" and represents the part of h which may be contaminated by the noise. We have $h = h_s + h_n$. Following this notation, the corrected coefficient $\tilde{w}$ should minimize:

$$J(\tilde{w}_j) = h_s(w_j - \tilde{w}_j) + \alpha h_n(\tilde{w}_j)$$

i.e. there is a minimum of information in the residual $(w - \tilde{w})$ which can be due to the significant signal, and a minimum of information which could be due to the noise in the solution $\tilde{w}_j$.

82

Deconvolution

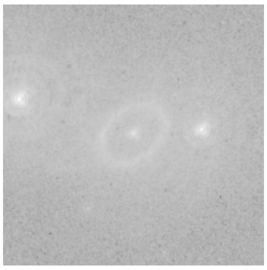
- … or deblurring
- Take observed signal,
- and approximate blurring function,
- and invert the blurring
- in order to obtain the original or ideal signal
- Usual principle:
- Redistribute image flux, iteratively, taking blurring function into account

83

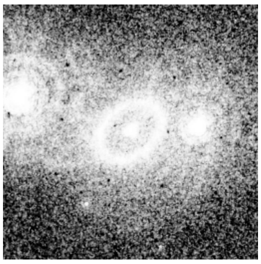
Hubble Space Telescope
Image restoration for tackling the spherical aberration problem between 1990 and 1993

- Concept of space telescope advanced by Lyman Spitzer Jr. (Princeton) in 1940s.
- First official mention by NASA in 1962, four years after NASA founded. Science team established by NASA in 1973. From 1975 ESA involved. Congress approved in 1978. Marshall (Huntsville) – construction, Goddard (Greenbelt) – science and ground control.
- Lockheed – basic spacecraft, Perkin-Elmer – primary mirror, ESA (BAe) – solar panels.
- 1986/1/26 – Challenger disaster.
- HST deployed in low Earth orbit – 600 km – by Discovery (STS-31) on 1990/4/25.
- First refurbishment mission Endeavour STS-61 in December 1993.

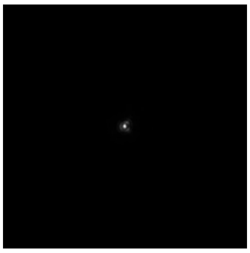
84



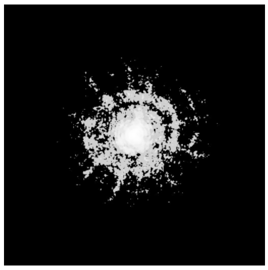
Supernova SN87A
HST image,
spherically aberrated,
original, histogram equalized



85



Point spread function
at original dynamic range
and histogram equalized



86

Deconvolution

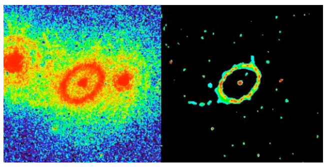
Observed data: I(x,y)
Real image: O(x,y)
Assume: linear, shift-invariant imaging system

Then:
$I(x,y) = (O * P)(x,y) + N(x,y)$
where P is point spread function, N is noise,
and * is convolution

Solve for O using coupled relations:
$O^{n+1} \leftarrow O^n \; [(I/I^n) * P^*]$
$I^n \leftarrow P * O^n$
where P* is conjugate of PSF

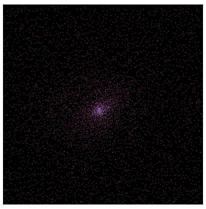
- Richardson, 1972, optical engineering literature
- Lucy, 1974, astronomy literature
- Shepp and Vardi, 1982, medical imaging literature
- Maximum likelihood solution, formally EM algorithm

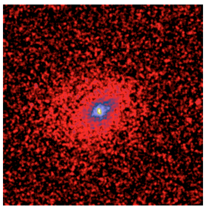
87

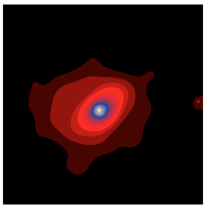


SN87a before (left) and after (right) deconvolution.
Both images histogram equalized and false colour coded.

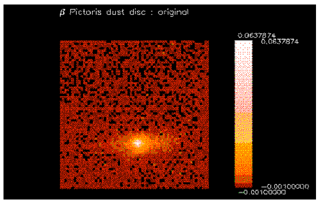
88

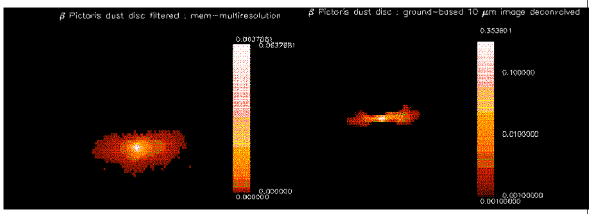






89





Regularization

- In image restoration, this means that (1) we have a criterion, say least squares goodness of fit, that we want to optimize, and (2) knowing that our data is noisy, we want to additionally have a smooth solution
- Overall we optimize:

(out - in)^2 + alpha f(out)

where here f is a smoothing function,

and alpha is a scalar “trade-off” value

91

Deconvolution

Formation of the image is expressed by the convolution integral

$$\begin{aligned} I(x,y) &= \int_{x_1=-\infty}^{+\infty}\int_{y_1=-\infty}^{+\infty} P(x-x_1, y-y_1)O(x_1,y_1)dx_1dy_1 + N(x,y) \\ &= (O * P)(x,y) + N(x,y) \end{aligned}$$

where I is the data, P the point spread function (PSF), and O is the object.
In Fourier space we have:

$$\hat{I}(u,v) = \hat{O}(u,v)\hat{P}(u,v) + \hat{N}(u,v)$$

We want to determine O knowing I and P.
The main difficulties are the existence of:

- a cut-off frequency of the point spread function,
- the noise.

This is in fact an **ill-posed problem**, and there is no unique solution.

92

Other topics related to deconvolution

- blind deconvolution – the PSF P is unknown.
- myopic deconvolution – the PSF P is partially known.
- superresolution – object spatial frequency information outside the spatial bandwith of the image formation system is recovered.
- image reconstruction – an image is formed from a series of projections (Computed Tomography, Positron Emission Tomography, ...), and the computation of the object from the projection image data embodies a computer synthesized PSF.

93

Regularization

To resolve these problems, some constraints must be added to the solution. The addition of such constraints is called regularization.

$$J(O) = \parallel I - P * O \parallel + \alpha C(O)$$

Several regularization methods exist:

- Tikhonov: Tikhonov's regularization consists of minimizing the term:

$$\parallel I(x,y) - (P * O)(x,y) \parallel + \lambda \parallel H * O \parallel$$

where H corresponds to a high-pass filter

- MEM: $C(O) = Entropy(O)$
- CLEAN: $Data = Stars +$ const. Background
- Markov Field: $O(x,y)$ is a function of its neighborhood.

Deconvolution Methods

In astronomy:

- without regularization: Lucy $O^{(n+1)} = O^{(n)}[\frac{I}{I^{(n)}} * P^*] = O^{(n)}[\frac{I^{(n)} + R^{(n)}}{I^{(n)}} * P^*]$ where $R^{(n)} = I - I^{(n)}$ and $I^{(n)} = P * O^{(n)}$
- regularized method: MEM and CLEAN.

In the signal processing domain:

- without regularization
  - Van Cittert: $O^{(n+1)} = O^{(n)} + \beta(I - P * O^{(n)})$
  - Fixed step gradient: $O^{(n+1)} = O^{(n)} + \beta P^* * (I - P * O^{(n)})$
- regularized method: Tikhonov and Markov field models.

95

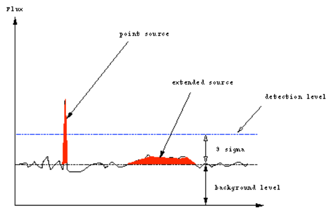
Novelty – regularization through wavelet-based denoising

- The problem with the traditional way of doing this is that the algorithm thinks that each and every noise spike is deserving of getting flux back, which has been blurred in its vicinity.
- "Regularization" is the term used for an imposed constraint to stop the algorithm doing this.
- Wavelet-based noise filtering: take **$R^n = I - I^n$**, i.e. **$R^n = I - (P * O^n)$, the observed image, I, minus the image we would observe if we were to use the current estimate of the ideal image,** which defines the residual at each iteration. Take **R^n** into wavelet space (known for its energy compacting properties). Denoise in wavelet space and inverse transform into direct space. Continue iterations. This allows a noise-"cleaned" version of **I^n**, while retaining and protecting signal.
- This is a computationally straightforward intervention at each iteration, based on the noise model used for the image. (E.g. widely-used CCD detectors have Gaussian read-out noise, together with Poisson photon noise.) Effective and efficient.

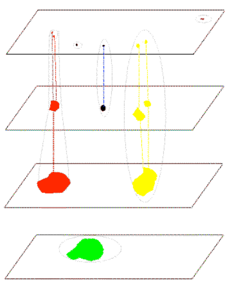
96

Object detection
Compression

97



98



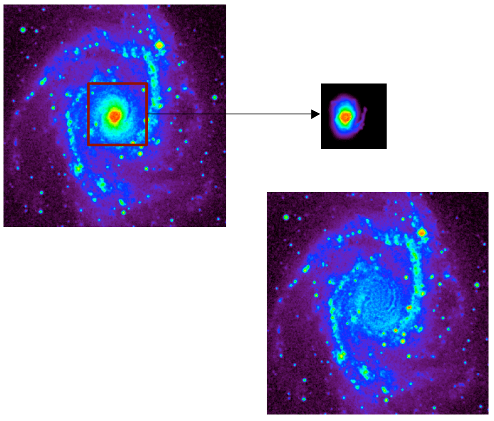
Structure: set of significant connected wavelet coefficients at a given scale.

Object: set of structures.

Object scale: the scale of an object is given by the scale of the maximum of its wavelet coefficients.

Interscale-relation: relation to group several structures into a single object.

99



100

A Short Review of Compression

- Use a pyramidal transform (so: a wavelet transform with decimation)
- Or a non-wavelet, multiresolution, transform, e.g. pyramidal median transform
- Incorporate noise modeling for e.g. CCD detectors
- Quantize, entropy encode
- Essential principle here: noise is inherently non-compressible, so separate out noise from signal

101

Image Compression

Why data compression?

- Transfer of data between satellites and ground-based stations.
- Images archiving.
- Fast access to large pictorial databases.
- Web-based data transmission.

102

Following the kind of images and the application needs, different strategies can be used:

1. Lossy compression: in this case the compression ratio is relatively low (< 5).
2. Compression without visual loss. This means that one cannot see the difference between the original image and the decompressed one. Generally, compression ratios between 10 and 20 can be obtained.
3. Good quality compression: the decompressed image does not contain any artifact, but some information is lost. Compression ratios up to 40 can be obtained in this case.
4. Fixed compression ratio: for some technical reasons, one may decide to compress all images with a compression ratio higher than a given value, whatever the effect on the decompressed image quality.
5. Signal/noise separation: if noise is present in the data, noise modeling can allow very high compression ratios just by including filtering in wavelet space during the compression.

Following the image type, and the selected strategy, the optimal compression method may differ. The main interest in using a multiresolution framework is to get, naturally, the possibility for progressive information transfer.

103

Redundancy

Compression methods try to use the redundancy contained in the raw data in order to reduce the number of bits. The main efficient methods belong to the transform coding family, where the image is first transformed into another set of data where the information is more compact (i.e. the entropy of the new set is lower than the original image entropy).

104

Compression steps

The typical steps are:

1. transform the image (for example using a discrete cosine transform, or a wavelet transform),
2. quantize the obtained coefficients, and
3. code the values by a Huffman or an arithmentic coder.

The first and third points are reversible, while the second is not. The distortion will depend on the way the coefficients are quantized. We may want to minimize the distortion with the minimum of bits, and a trade-off will then be necessary in order to also have "acceptable" quality. "Acceptable" is evidently subjective, and will depend on the application. Sometimes, any loss is unacceptable, and the price to pay is a very low compression ratio (often between one and two).

According to Shannon's theorem, the number of bits we need to code an image I without distorsion is given by its entropy H. If the image (with N pixels) is coded with L intensity levels, each level having a probability p_i to appear, the entropy H is

$$H(I) = \sum_{i=1}^{L} -p_i \log_2 p_i$$

The probabilities p_i can be easily derived from the image histogram. The compression ratio is given by:

$$\mathcal{C}(I) = \frac{\text{number of bits per pixel in the raw data}}{H(I)}$$

A Huffman, or an arithmetic, coder is generally used to transform the set of interger values into the new set of values, in a reversible way.

For lossy compression methods, the distortion is measured by

$$R = \| I - \tilde{I} \| = \sum_{k=1}^{N} (I_k - \tilde{I}_k)^2$$

where $\tilde{I}$ is the decompressed image.
The SNR is:

$$SNR_{dB} = 10 \log_{10} \frac{\sum_{k=1}^{N} I_k^2}{R}$$

and the Peak SNR (PSNR) is

$$PSNR_{dB} = 10 \log_{10} \frac{255^2}{\frac{R}{N}}$$

Quality criteria for lossy methods in Astronomy

1. Visual aspect
2. Signal to noise ratio
3. Photometry
4. Astrometry
5. Detection of real and faint objects
6. Object morphology

108

Quality criteria for lossy methods in Astronomy

1. Visual aspect
2. Signal to noise ratio
3. Photometry
4. Astrometry
5. Detection of real and faint objects
6. Object morphology

Methods in Astronomy

Standard methods in astronomy

- JPEG: is the standard video compression software for single frame images. It decorrelates pixel coefficients within 8 by 8 pixel blocks using the Discrete Cosine Transform and uniform quantization.
- FITSPRESS (Press, 1992): is based on a wavelet transform, using Daubechies-4 filters. It was developed at the Center for Astrophysics, Harvard.
- HCOMPRESS (White et al, 1992): was developed at Space Telescope Science Institute (STScI), Baltimore, and is commonly used to distribute archive images from the Digital Sky Survey. It is based on the Haar wavelet transform.

Other methods in astronomy

- Mathematical Morphology + Quadtree (Huang and Bijaoui, 1991)
- Pyramidal Median Transform (Starck et al, 1996)
- Iterated Hcompress method (Bijaoui et al, 1996)

Wavelet Transform and Compression

- Choice of the filters
  Antonini 7/9 filters are the most often used, with an L^2 normalization.
- Quantization
- Coding

Quantization + Coding Algorithm

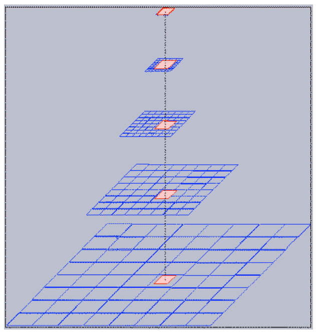
- Uniform quantization + Huffman encoding
- Vector quantization
- Uniform quantization + Quadtree + Huffman encoding
- Embedded Zerotree Wavelet

Large Images

With new technology developments, images furnished by detectors are larger and larger. For example, current astronomical projects plan to deal with images of sizes larger than 8000 by 8000 pixels (VLT: $8k \times 8k$, MEGACAM $16k \times 16k$, ...). Analysis of such images is obviously not easy, but the main problem is clearly archiving and network access to the data by users.

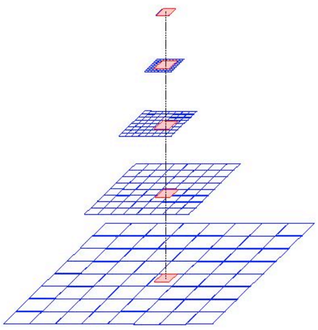
In order to visualize an image in a reasonable amount of time, transmission is based on two concepts:

- data compression
- progressive decompression



- We use a wavelet pyramidal representation to support lossless compression
- Supporting very fast decompression by resolution level and by block.

114



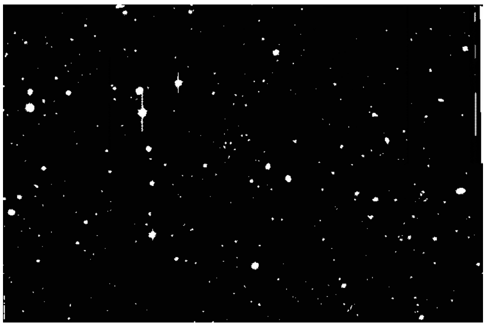
Large image, compressed by block, and represented at five resolution levels. At each resolution level, the visualization window is superimposed at a given position.

115

Large Image Visualization Environment: LIVE

A third concept is necessary, which is the region of interest. Indeed, images become so large it is impossible to display them in a normal window (typically of size 512x512), and we need to have to ability to focus on a given area of the image at a given resolution. To move from one area to another, or to increase the resolution of a part of the area is a user task, and is a new active element of the decompression. The goal of LIVE is to furnish this third concept, by adding the following functionality:

- Full image display at a very low resolution.
- Image navigation: the user can go up (the quality of an area of the image is improved) or down (return to the previous image). Then the new image represents only a quarter of the previous one.



117

Indicative results of lossy and lossless large astronomy image compression

- 12,451 x 8268 MegaCam image (Canada-France-Hawaii Telescope, Hawaii). 412 MB greyscale image file.
- Visually-lossless compression took 10 minutes on workstation. Compressed to 4.1 MB i.e. < 1% of original size
- Decompression to resolution scale 5 (as shown) took 0.43 sec.
- Rigorously lossless compression took just over 3 minutes and resulted in a compressed file of size 97.8 MB, i.e. 23.75% of original size.

118