Source Separation based on Morphological Diversity

J.-L. Starck

Dapnia/SEDI-SAP,

Service d'Astrophysique

CEA-Saclay, France.

jstarck@cea.fr

http://jstarck.free.fr

Collaborators:

P. Abrial and J. Bobin, *CEA-Saclay, France*

D.L. Donoho, *Department of Statistics, Stanford*

M. Elad, *The Technion, Israel Institute of Technology*

J. Fadili, *Caen University , France*

Y. Moudden, *CEA-Saclay, France*

1. **Introduction**
2. **The MCA algorithm**
3. **MCA texture extraction**
4. **MCA Inpainting**
5. **Multichannel MCA**

What is a good representation for data?

- Computational harmonic analysis seeks representations of a signal as linear combinations of basis, frame, dictionary, element :

$$f = \sum_k a_k \, \mathbf{b}_k$$

coefficients basis, frame

- Fast calculation of the coefficients a_k
- Analyze the signal through the statistical properties of the coefficients
- Approximation theory uses the sparsity of the coefficients.

Seeking sparse and generic representations

- Sparsity

Non-linear approximation curve (reconstruction error versus nbr of coeff)

- Why do we need sparsity?
  - data compression
  - Feature extraction, detection
  - Image restoration

Truncated Fourier series give very good approximations to smooth functions, but

–Provides poor representation of non stationary signals or image.

–Provides poor representations of discontinuous objects (Gibbs effect)

JPEG / JPEG2000

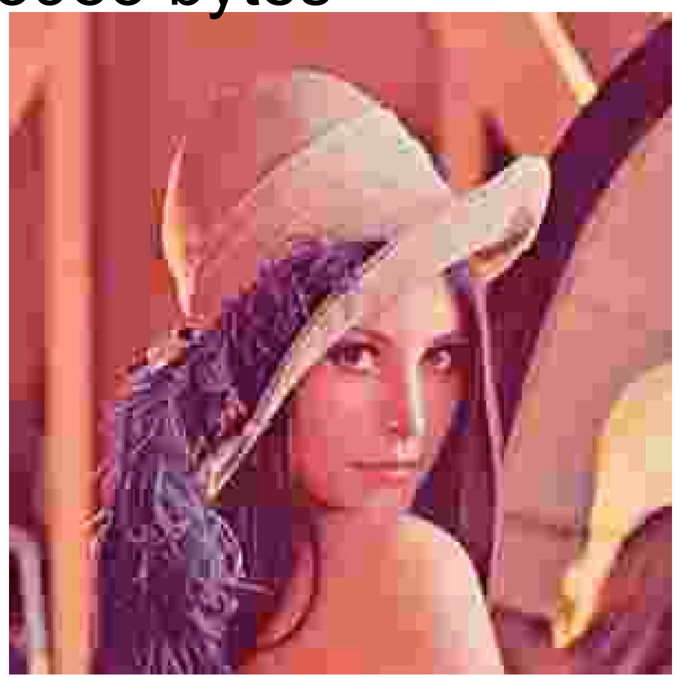
Original
BMP

300x300x24

270056
bytes

JPEG 1:68

3983 bytes

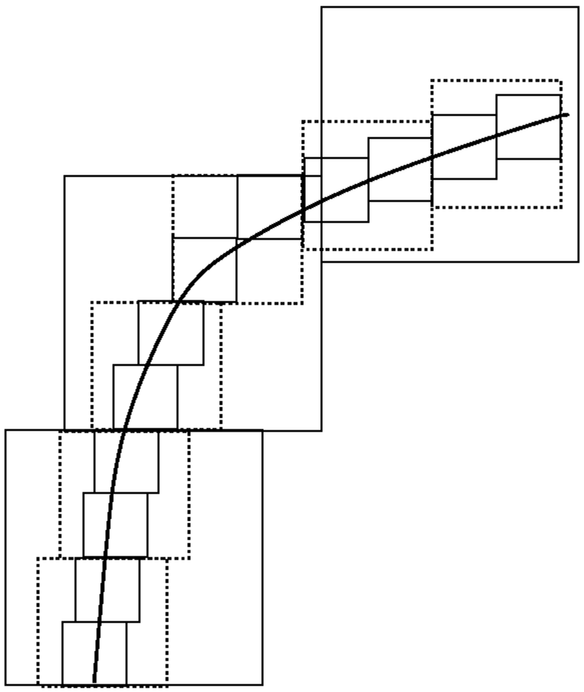


JPEG2000 1:70

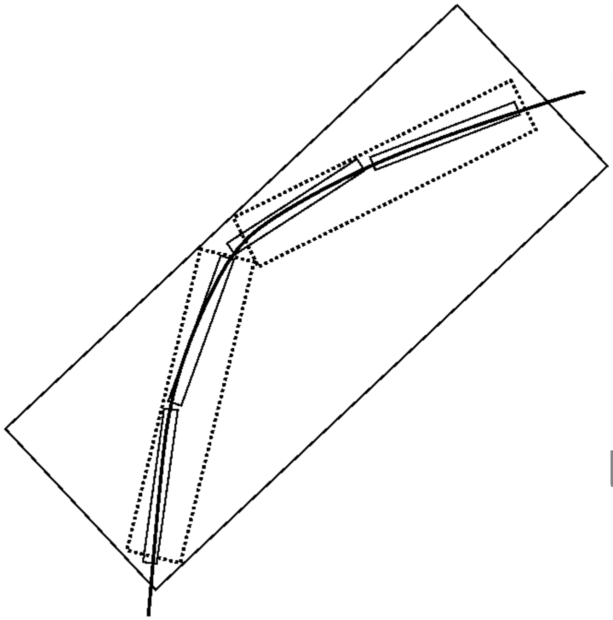
3876 bytes

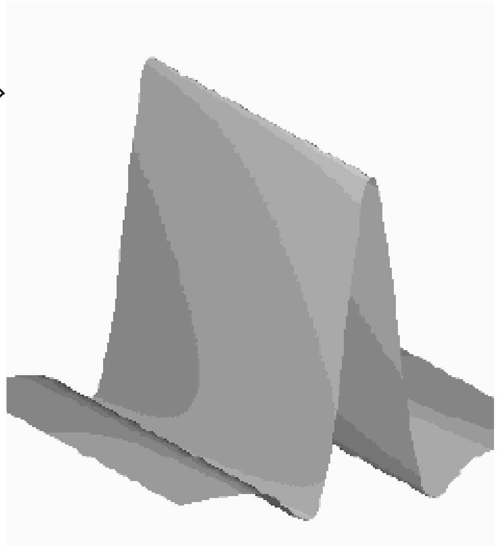
Wavelets and edges

• many wavelet coefficients are needed to account for edges ie singularities along lines or curves :



• need dictionaries of strongly anisotropic atoms :





ridgelets, curvelets, contourlets, bandelettes, etc.

Multiscale Transforms

Critical Sampling

(bi-) Orthogonal WT
Lifting scheme construction
Wavelet Packets
Mirror Basis

Redundant Transforms

Pyramidal decomposition (Burt and Adelson)
Undecimated Wavelet Transform
Isotropic Undecimated Wavelet Transform
Complex Wavelet Transform
Steerable Wavelet Transform
Dyadic Wavelet Transform
Nonlinear Pyramidal decomposition (Median)

New Multiscale Construction

Contourlet
Bandelet
Finite Ridgelet Transform
Platelet
(W-)Edgelet
Adaptive Wavelet

Ridgelet
Curvelet (Several implementations)
Wave Atom

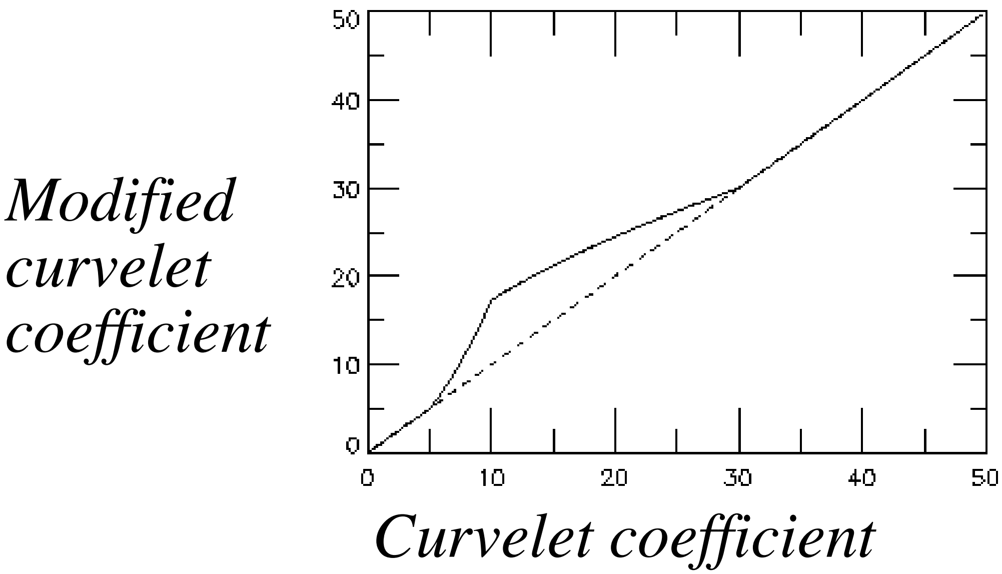
CONTRAST ENHANCEMENT USING THE CURVELET TRANSFORM

J.-L Starck, F. Murtagh, E. Candes and D.L. Donoho, ***"Gray and Color Image Contrast Enhancement by the Curvelet Transform"***,

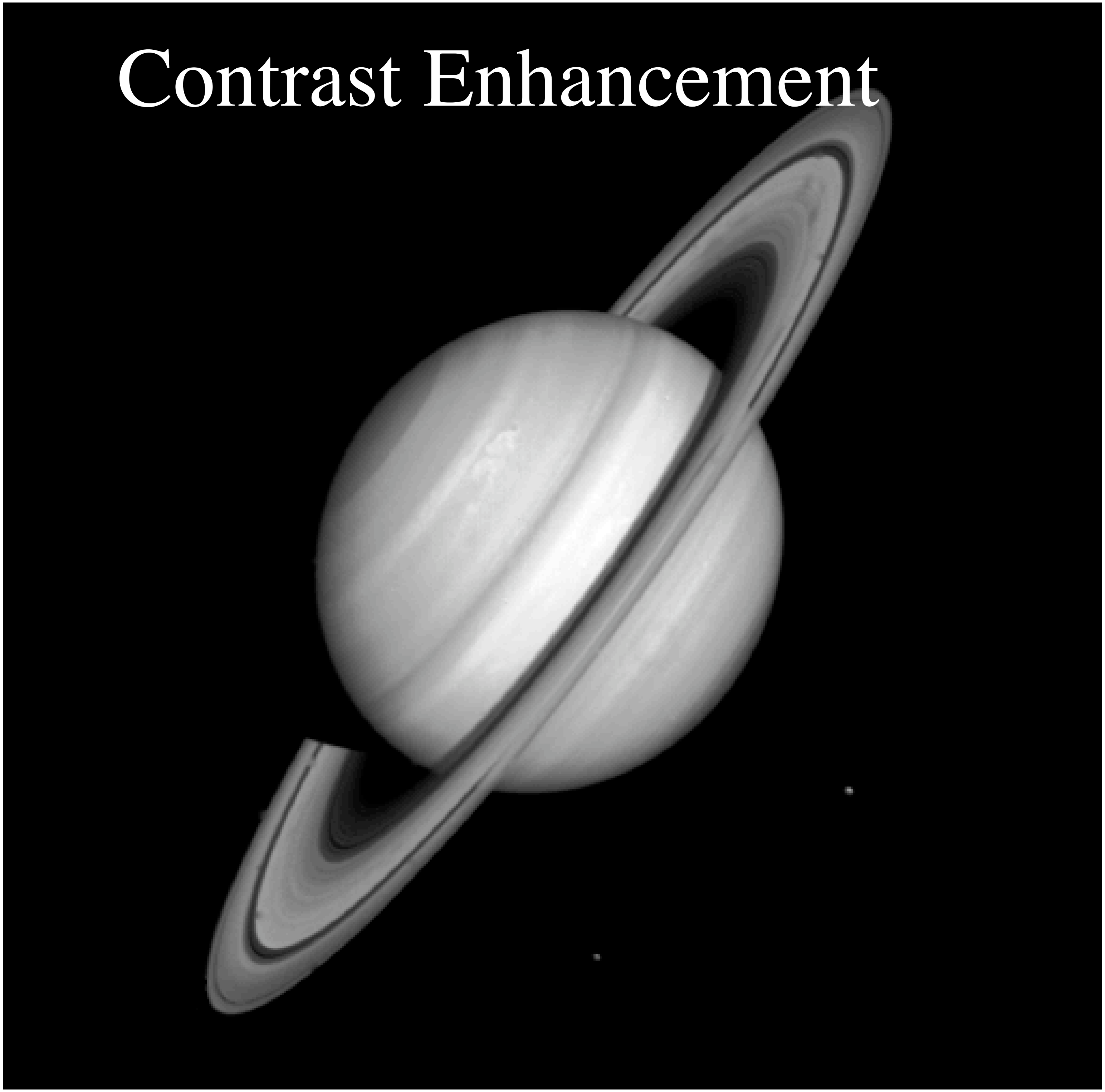
IEEE Transaction on Image Processing, 12, 6, 2003.

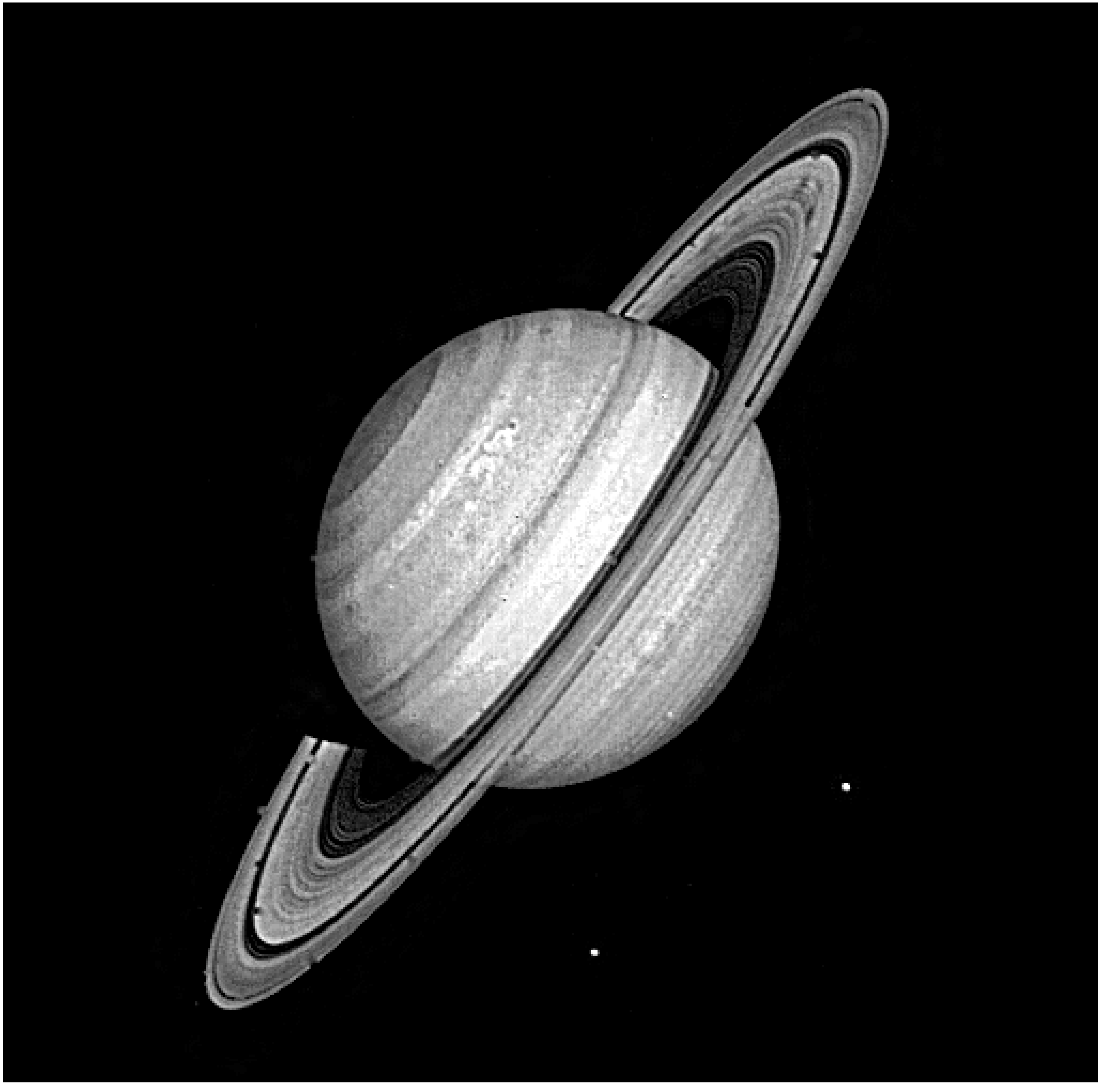
$$\tilde{I} = C_R\left(y_c\left(C_T I\right)\right)$$

$$
\begin{cases}
y_c(x,\sigma) = 1 & \text{if} \quad x < c\sigma \\
y_c(x,\sigma) = \dfrac{x - c\sigma}{c\sigma}\left(\dfrac{m}{c\sigma}\right)^p + \dfrac{2c\sigma - x}{c\sigma} & \text{if} \quad x < 2c\sigma \\
y_c(x,\sigma) = \left(\dfrac{m}{x}\right)^p & \text{if} \quad 2c\sigma \le x < m \\
y_c(x,\sigma) = \left(\dfrac{m}{x}\right)^s & \text{if} \quad x > m
\end{cases}
$$



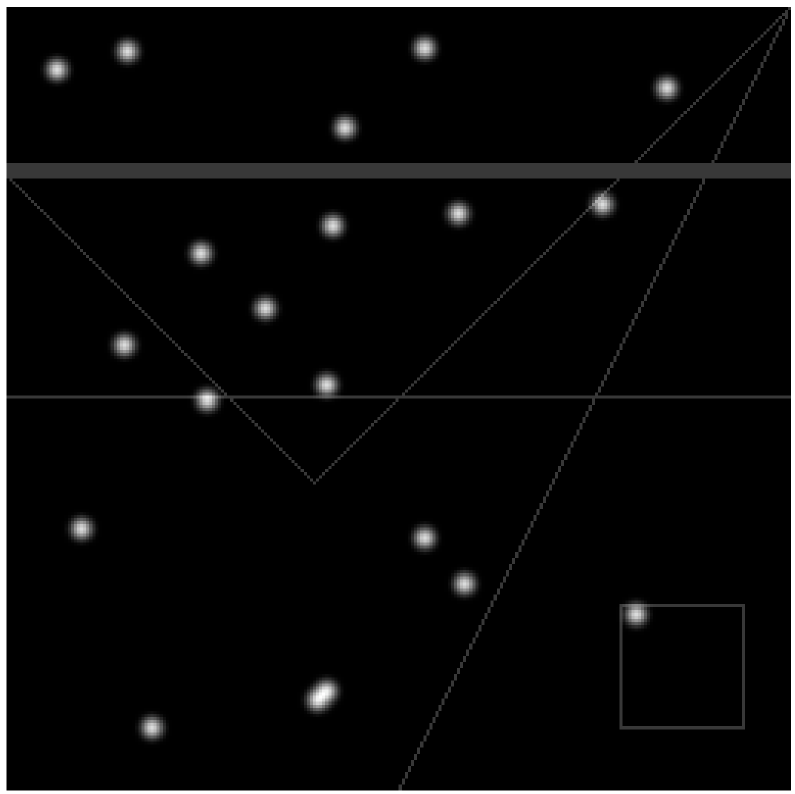
Contrast Enhancement



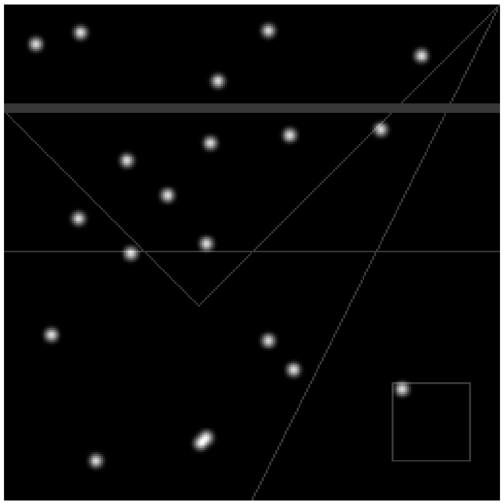


A difficult issue

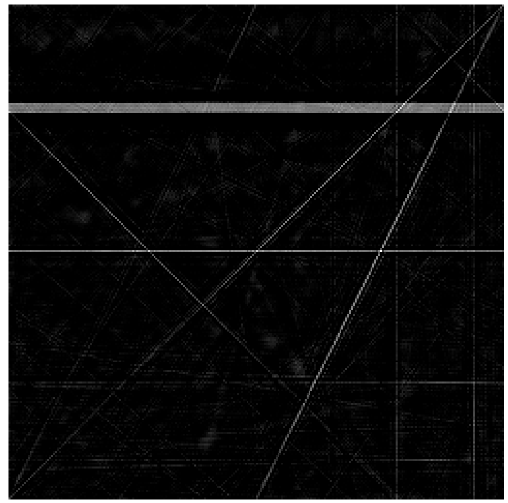
Is there any representation that well represents the following image ?



Going further

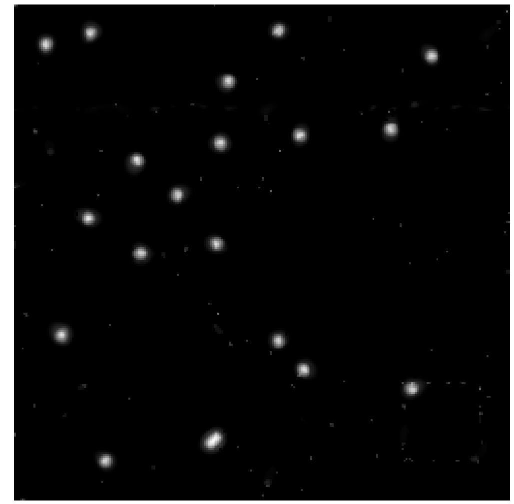


=



Lines

+



Gaussians

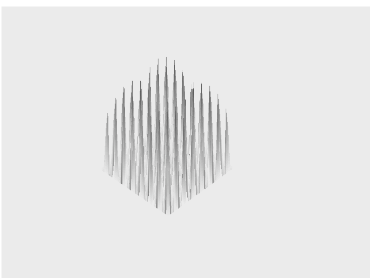
Curvelets

Wavelets

REDUNDANT REPRESENTATIONS

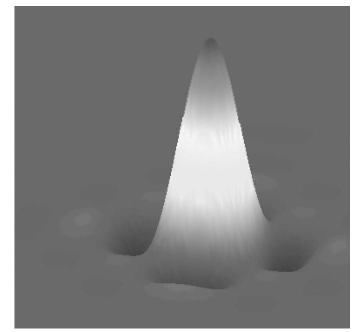
How to choose a representation ?

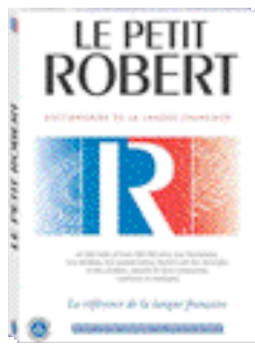
Basis → Dictionary



Local DCT

Wavelets





Curvelets

Others

Sparse Representation in a Redundant Dictionary

Given a signal s, we assume that it is the result of a sparse linear combination of atoms from a known dictionary D.

A dictionary D is defined as a collection of waveforms $(\phi_\gamma)_{\gamma\in\Gamma}$, and the goal is to obtain a representation of a signal s with a linear combination of a small number of basis such that:

$$s = \sum_\gamma \alpha_\gamma \phi_\gamma$$

Or an approximate decomposition:

$$s = \sum_\gamma \alpha_\gamma \phi_\gamma + R$$

Formally, the sparsest coefficients are obtained by solving the optimization problem:

$$\text{(P0)} \quad \text{Minimize} \quad \|\alpha\|_0 \quad \text{subject to} \quad s = \phi\alpha$$

It has been proposed (*to relax and*) to replace the l_0 norm by the l_1 norm (Chen, 1995):

$$\text{(P1)} \quad \text{Minimize} \quad \|\alpha\|_1 \quad \text{subject to} \quad s = \phi\alpha$$

It can be seen as a kind of convexification of (P0).

It has been shown (Donoho and Huo, 1999) that for certain dictionary, it there exists a highly sparse solution to (P0), then it is identical to the solution of (P1).

We consider now that the dictionary is built of a set of L dictionaries related to multiscale transforms, such wavelets, ridgelet, or curvelets.

Considering L transforms, and α_k the coefficients relative to the kth transform:

$$\phi = [\phi_1, \ldots, \phi_L], \quad \alpha = \{\alpha_1, \ldots, \alpha_L\}, \quad s = \phi\alpha = \sum_{k=1}^{L} \phi_k \alpha_k$$

Noting $T_1, \ldots T_L$ the L transform operators, we have:

$$\alpha_k = T_k s_k, \qquad s_k = T_k^{-1} \alpha_k, \qquad s = \sum_{k=1}^{L} s_k$$

A solution α is obtained by minimizing a functional of the form:

$$J(\alpha) = \left\| s - \sum_{k=1}^{L} T_k^{-1} \alpha_k \right\|_2^2 + \|\alpha\|_p$$

Different Problem Formulation

$$J(s_1,\ldots,s_L) = \left\| s - \sum_{k=1}^{L} s_k \right\|_2^2 + \lambda \sum_{k=1}^{L} \left\| T_k s_k \right\|_p$$

.We do not need to keep all transforms in memory.

. There are less unknown (because we use non orthogonal transforms).

.We can easily add some constraints on a given component

Morphological Component Analysis (MCA)

"Redundant Multiscale Transforms and their Application for Morphological Component Analysis", Advances in Imaging and Electron Physics, 132, 2004.

$$J(s_1,\ldots,s_L) = \left\| s - \sum_{k=1}^{L} s_k \right\|_2^2 + \lambda \sum_{k=1}^{L} \| T_k s_k \|_p + \sum_{k=1}^{L} \gamma_k C_k(s_k)$$

$C_k(s_k)$ = constraint on the component s_k

Compare to a standard matching or basis pursuit:

- We do not need to keep all transforms in memory.
- There are less unknown (because we use non orthogonal transforms).
- We can easily add some constraints on a given component

The MCA Algorithm

The MCA algorithm relies on an iterative scheme: at each iteration, MCA picks in alternately in each basis the most significant coefficients of a residual term:

. Initialize all s_k to zero

. Iterate t=1,...,Niter

- Iterate k=1,..,L

Update the kth part of the current solution by fixing all other parts and minimizing:

$$J(s_k) = \left\| s - \sum_{i=1, i \neq k}^{L} s_i - s_k \right\|_2^2 + \lambda_t \left\| T_k s_k \right\|_1$$

Which is obtained by a simple soft/hard thresholding of : $s_r = s - \sum_{i=1, i \neq k}^{L}$

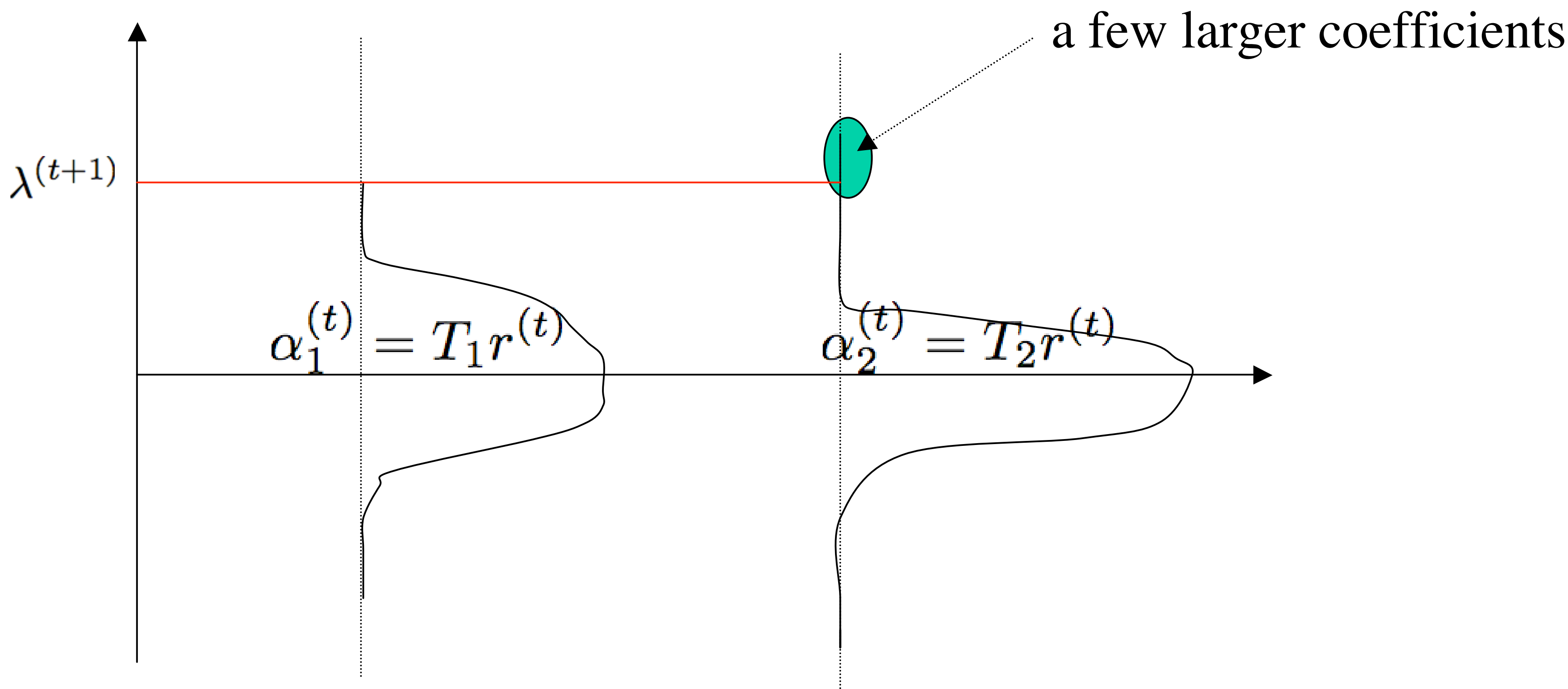
- Decrease λ_t

How to optimally tune the thresholds ?

- The thresholds play a key role as they manage the way coefficients are selected and thus determine the sparsity of the decomposition.

- As K transforms per iteration are necessary :
  the least number of iterations, the faster the decomposition.

$$r^{(t)} = s - s_1^{(t)} - s_2^{(t)}$$



a few larger coefficients
$\lambda^{(t+1)}$
$\alpha_1^{(t)} = T_1 r^{(t)}$
$\alpha_2^{(t)} = T_2 r^{(t)}$

In practice : an empirical approach: The « MOM » strategy

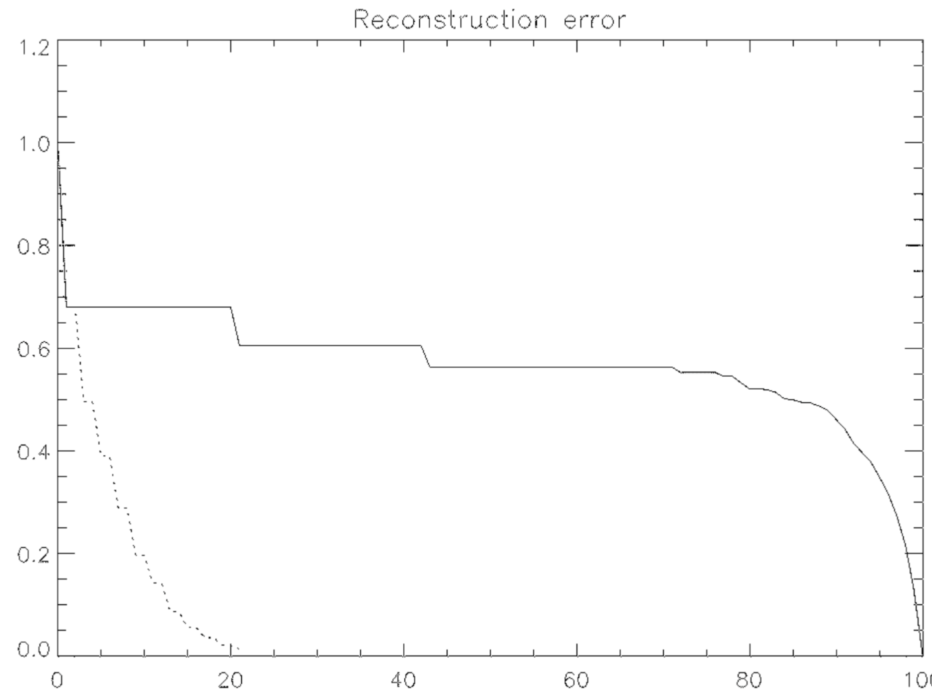
In practice, we would like to use an adaptative tuning strategy.
For a union of 2 orthogonal bases, the threshold is selected such that:

$$\min\{||r^{(k)}\mathbf{\Phi}_1||_\infty, ||r^{(k)}\mathbf{\Phi}_2||_\infty\} < \lambda < \max\{||r^{(k)}\mathbf{\Phi}_1||_\infty, ||r^{(k)}\mathbf{\Phi}_2||_\infty\}$$

That's why this strategy is called « Min Of Max » (MOM)

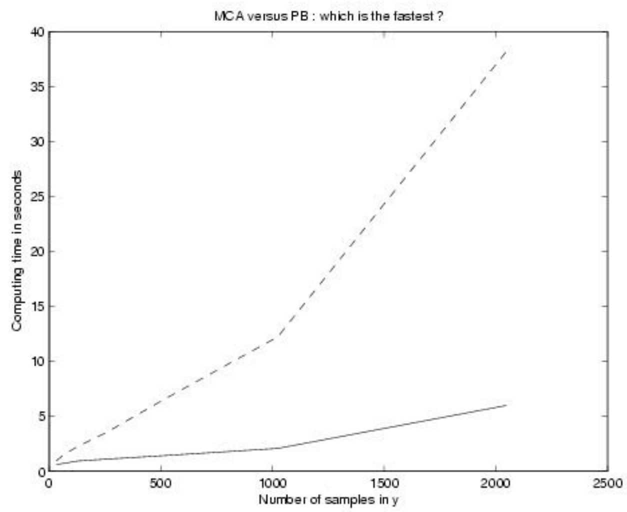
J. Bobin, J.-L. Starck, J. Fadili, Y. Moudden, and D.L. Donoho, "Morphological Component Analysis: new Results", submitted.

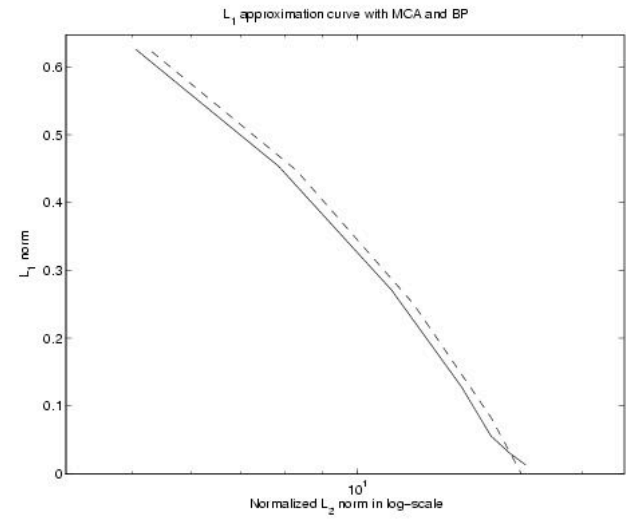
Mom in action

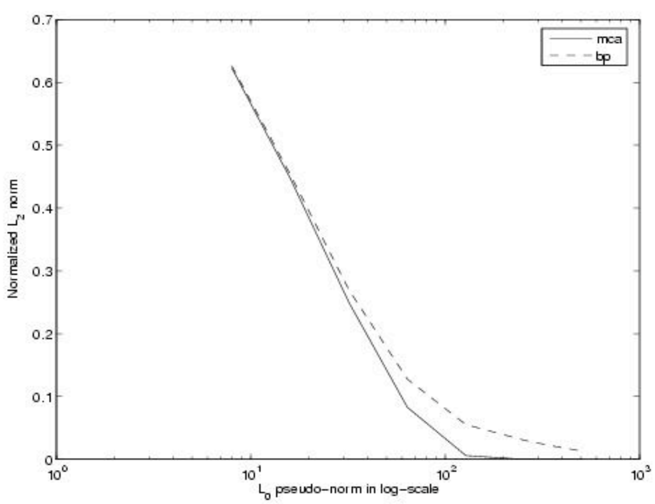


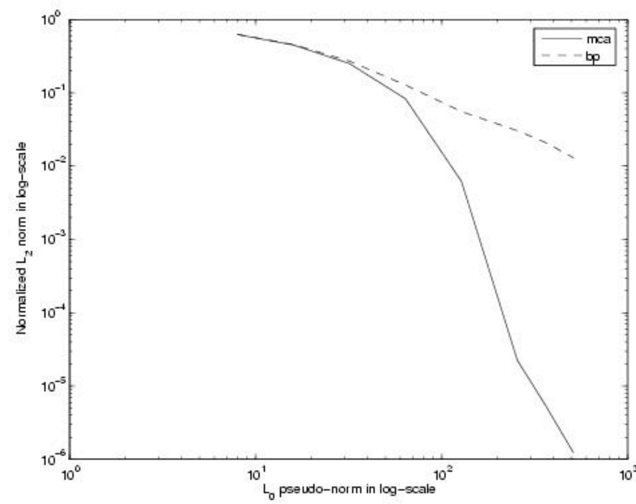
$$\Phi = \text{Curvelets} + \text{Global DCT}$$

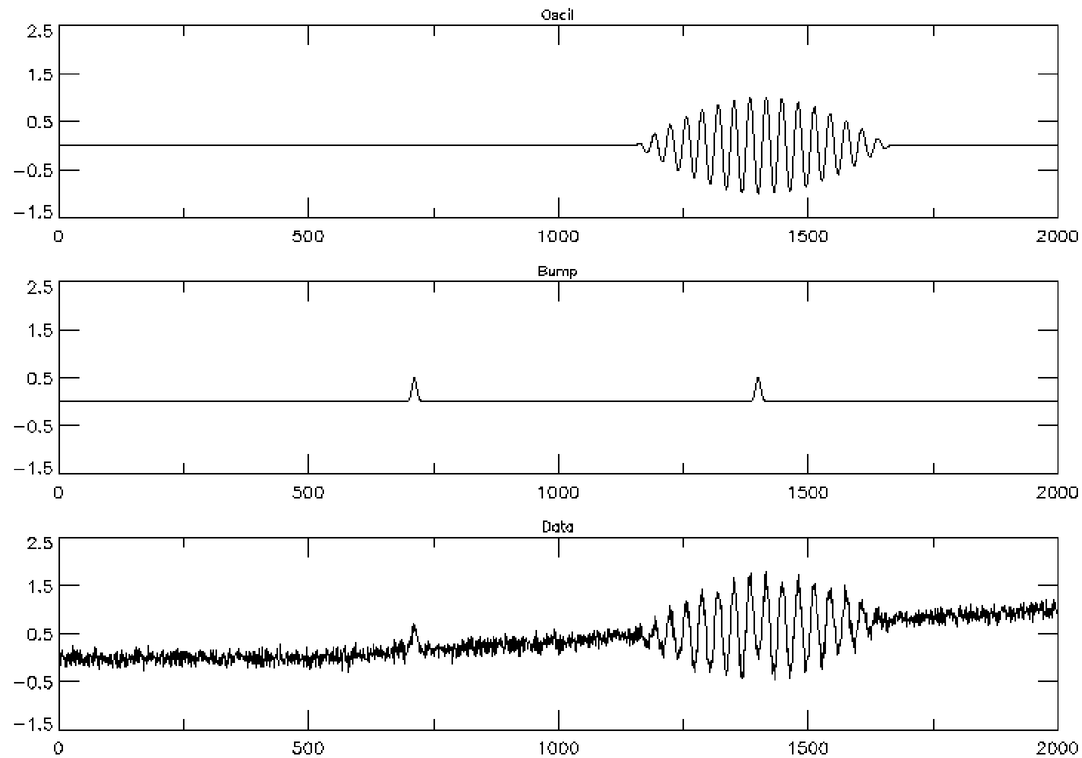
MCA versus Basis Pursuit



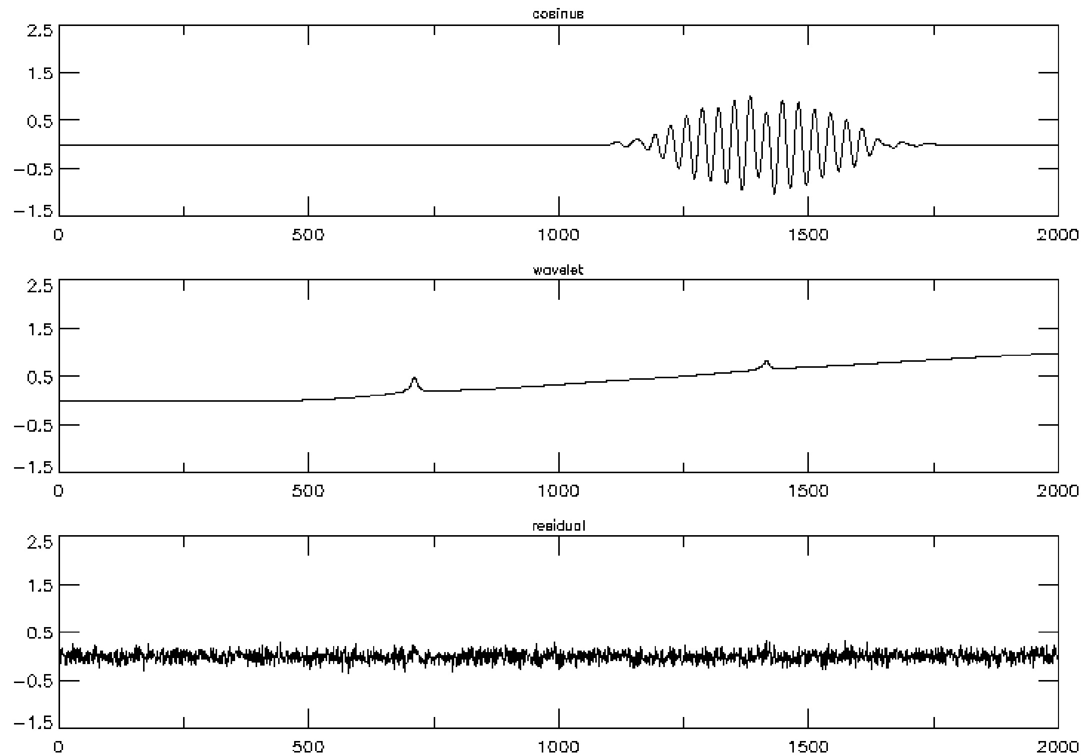




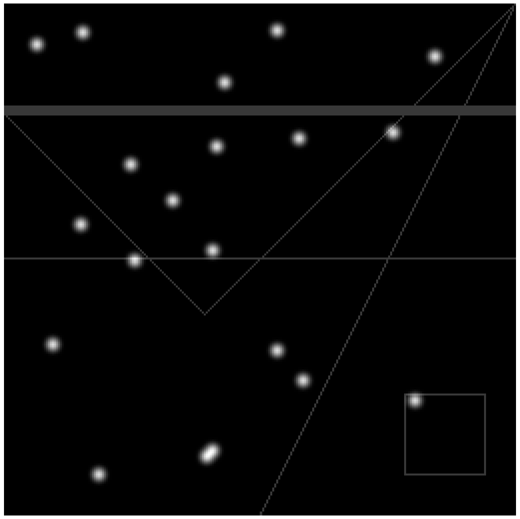




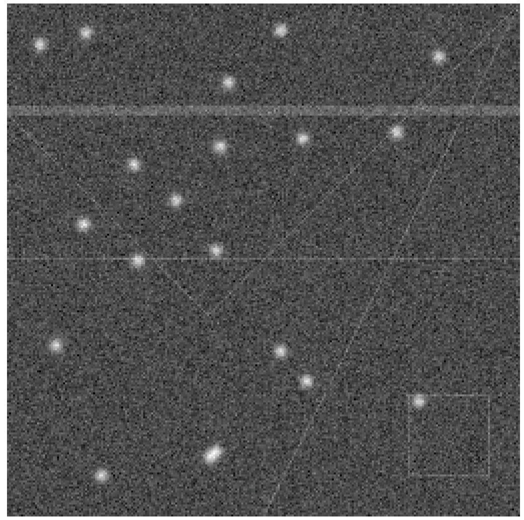
From top to bottom, oscillating component, component with bumps, and simulated data



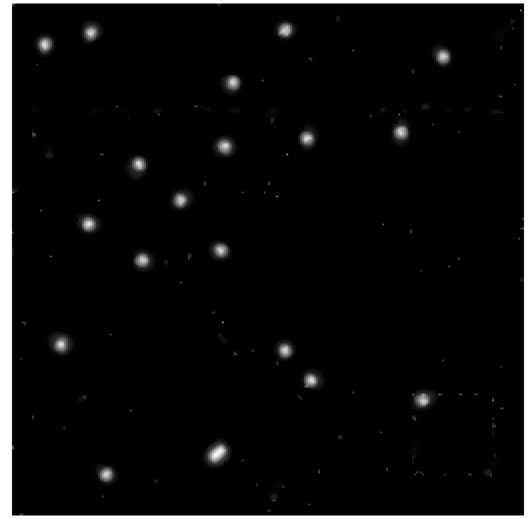
From top to bottom, reconstructed oscillating component, reconstructed component with bumps, and residual.



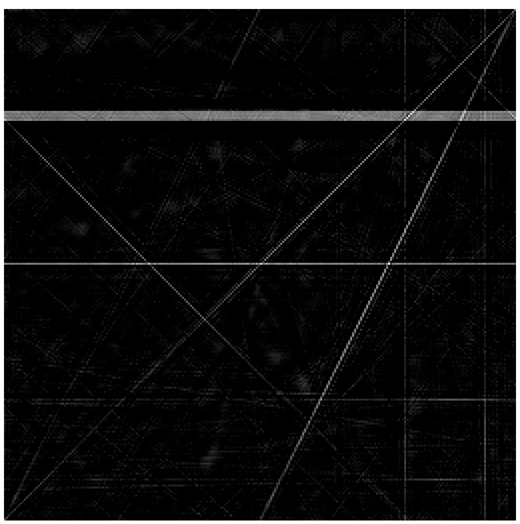
a) Simulated image (Gaussians+lines)



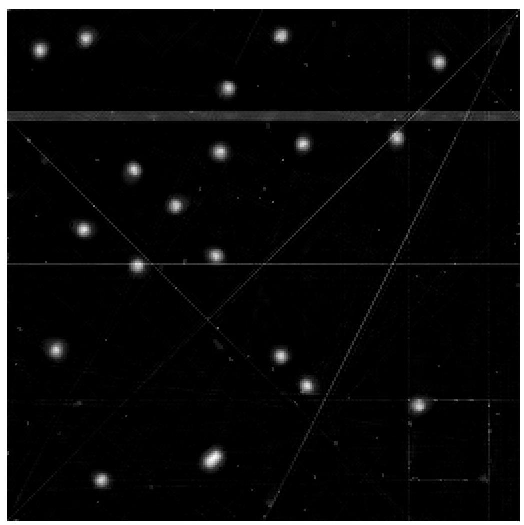
b) Simulated image + noise



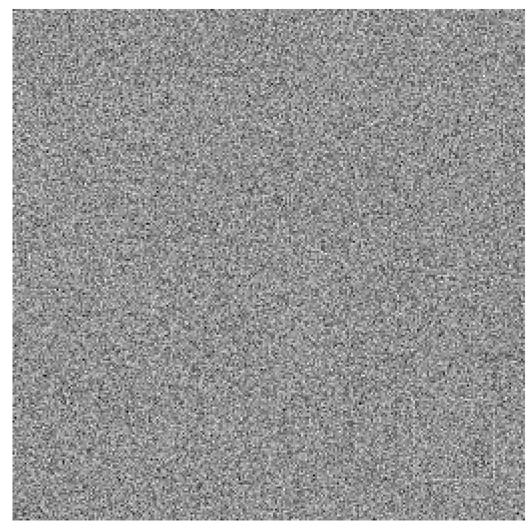
c) A trous algorithm



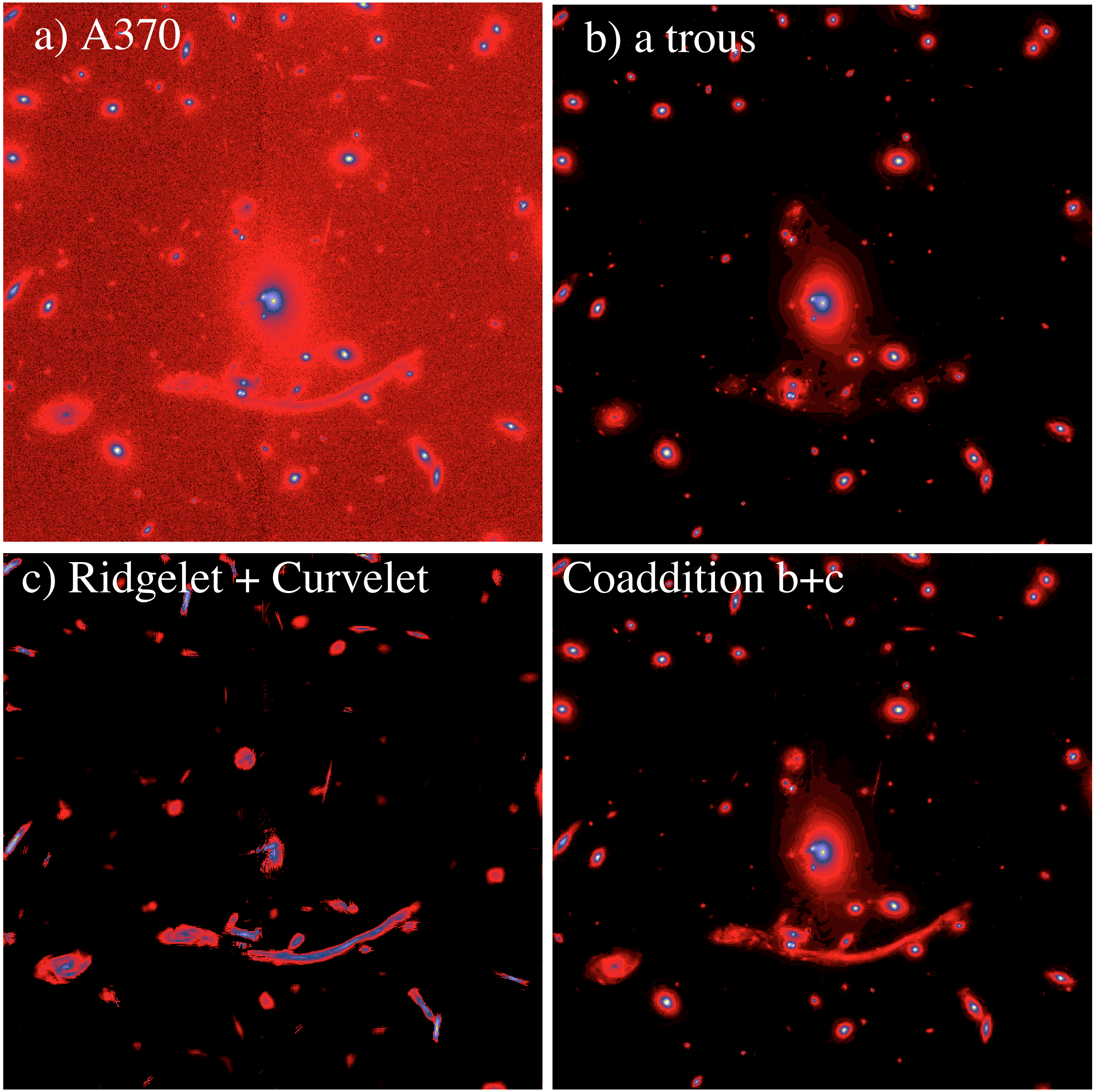
d) Curvelet transform

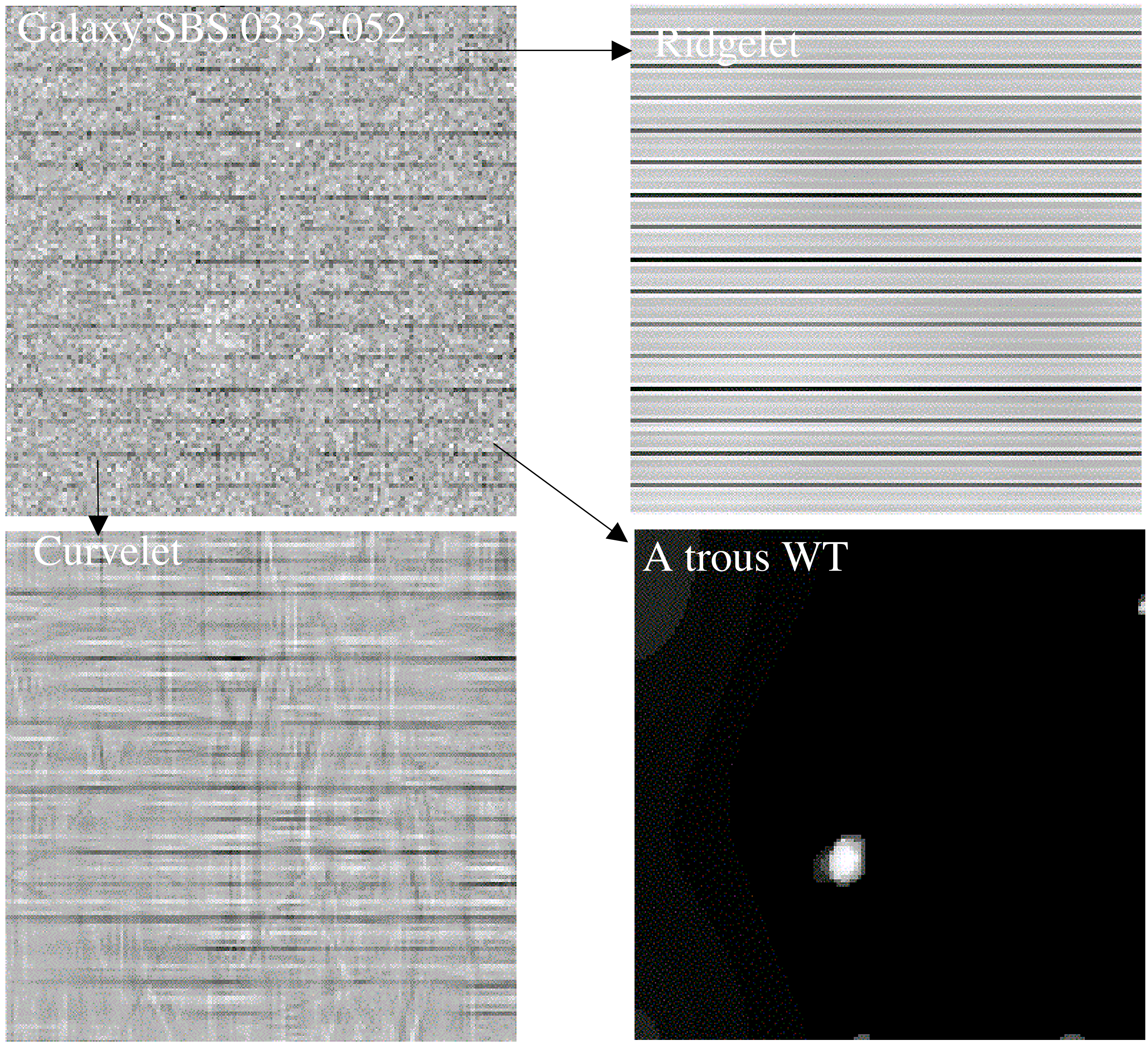


e) coaddition c+d

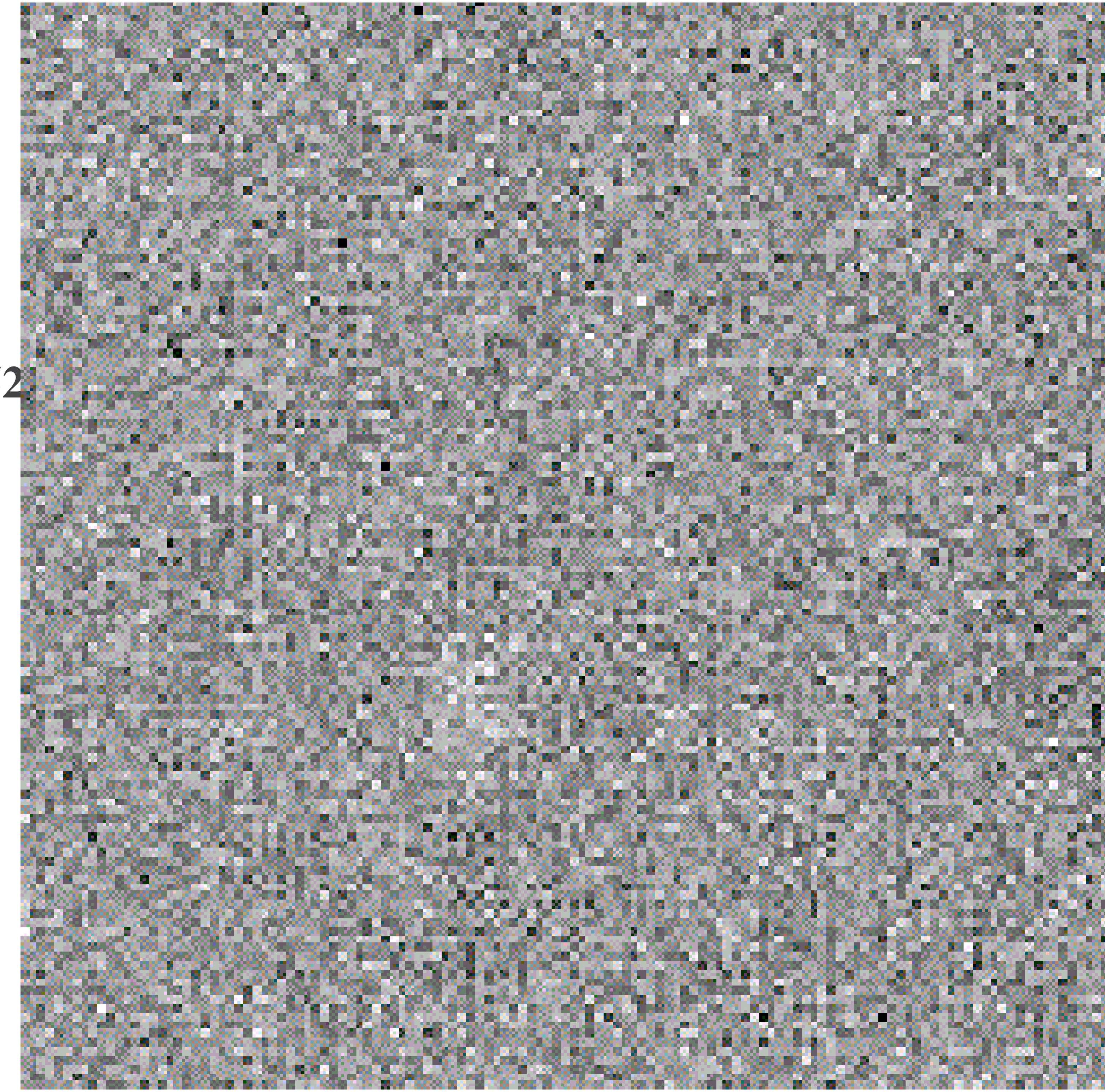


f) residual = e-b


a) A370
b) a trous
c) Ridgelet + Curvelet
Coaddition b+c


Galaxy SBS 0335-052
Ridgelet
Curvelet
A trous WT

Galaxy SBS 0335-052
10 micron
GEMINI-OSCIR



Separation of Texture from Piecewise Smooth Content

<u>The separation task</u>: decomposition of an image into a texture and a natural (piecewise smooth) scene part.

Dictionaries Choice

For the texture description (i.e. $\mathbf{T}_t$ dictionary), the DCT seems to have good properties. If the texture is not homogeneous, a local DCT should be preferred.

The curvelet transform represents well edges in an images, and should be a good candidate in many cases. The un-decimated wavelet transform could be used as well. In our experiments, we have chosen images with edges, and decided to apply the texture/signal separation using the DCT and the curvelet transform.

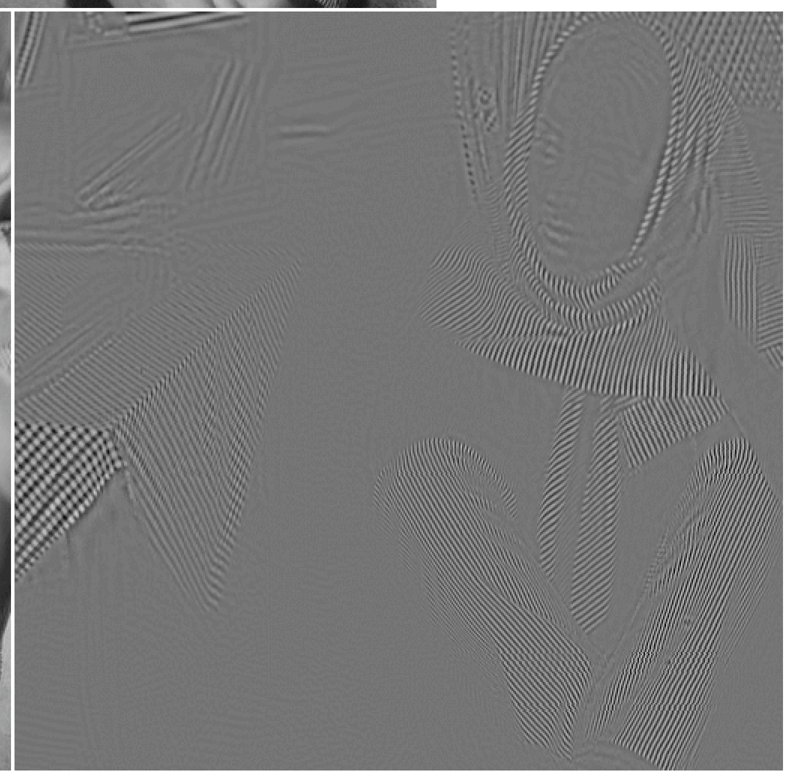
Numerical Consideration

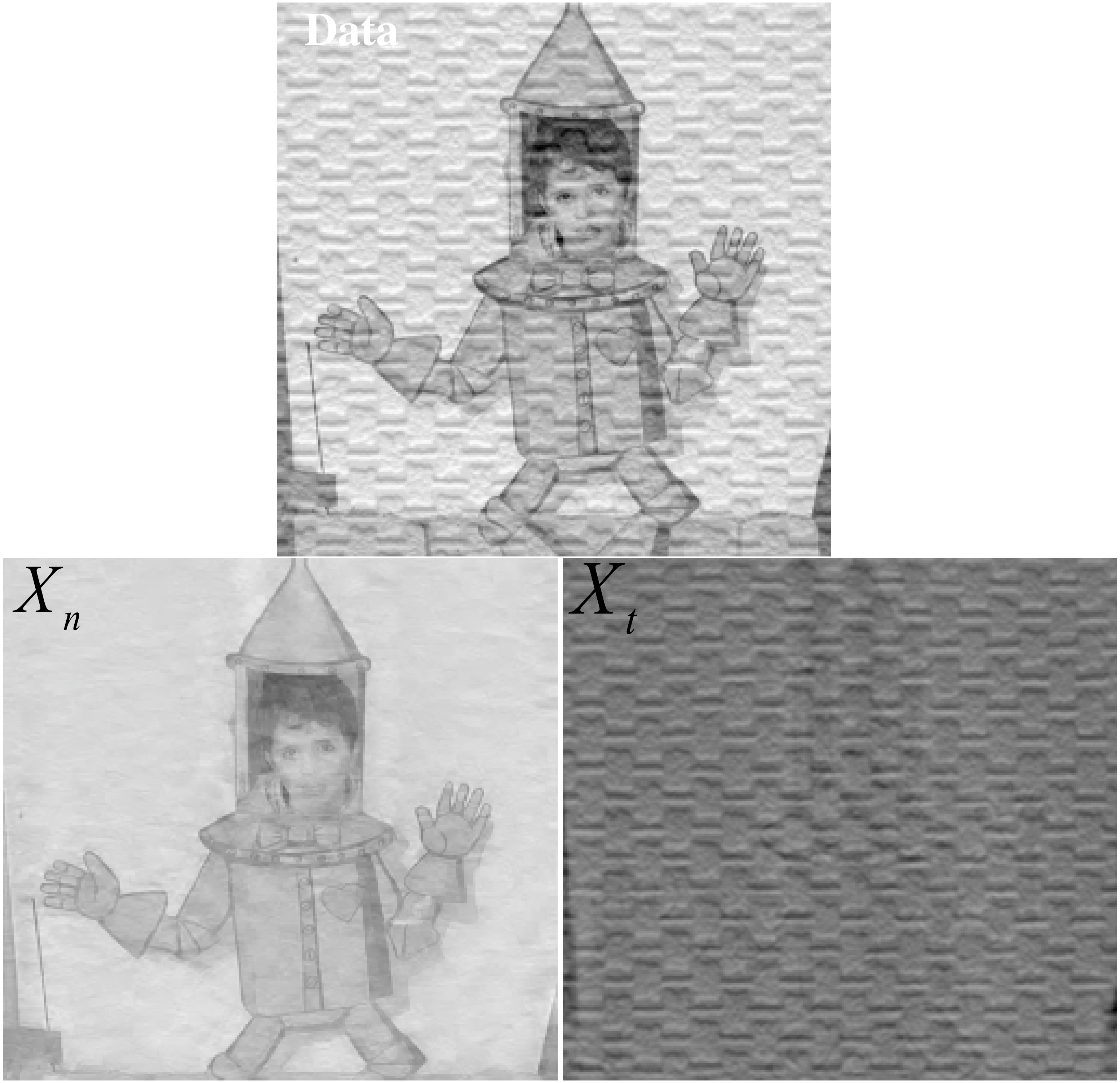
The DCT is denoted $\mathcal{D}$ and its inverse by $\mathcal{D}^{-1}$ (with a clear abuse of notations). The curvelet transform is denoted it by $\mathcal{C}$ and its inverse by $\mathcal{C}^{-1}$. We have two unknowns - $\underline{X}_t$ and $\underline{X}_n$ - the texture and the piecewise smooth images. The optimization problem to be solved is

$$\min_{\{\underline{X}_t, \underline{X}_n\}} \quad \|\mathcal{D}\underline{X}_t\|_1 + \|\mathcal{C}\underline{X}_n\|_1 + \lambda \|\underline{X} - \underline{X}_t - \underline{X}_n\|_2^2 + \gamma TV\left\{\underline{X}_n\right\}.$$

J.-L. Starck, M. Elad abd D.L. Donoho, "Image Decomposition Via the Combination of Sparse Representation and a Variational Approach", IEEE Transaction on Image Processing, 14, 10, pp 1570--1582, 2005.






Data
X_n
X_t

Edge Detection

Interpolation of Missing Data

$$J(s_1,\ldots,s_L) = \left\| M(s - \sum_{k=1}^{L} s_k) \right\|_2^2 + \lambda \sum_{k=1}^{L} \left\| T_k s_k \right\|_p$$

Where M is the mask: M(i,j) = 0 ==> missing data
M(i,j) = 1 ==> good data

If the data are composed of a piecewise smooth component + texture

$$J(X_t, X_n) = \left\| M(X - X_t - X_n) \right\|_2^2 + \lambda(\left\| \mathbf{C} X_n \right\|_1 + \left\| \mathbf{D} X_t \right\|_1) + \gamma \, \mathrm{TV}(X_n)$$

•*M. Elad, J.-L. Starck, D.L. Donoho, P. Querre, "Simultaneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA)", ACHA, Vol. 19, pp. 340-358, November 2005.*

•*M.J. Fadili, J.-L. Starck, "Sparse Representations and Bayesian Image Inpainting" , SPARS'05, Vol. I, Rennes, France, Nov., 2005.*

•*M.J. Fadili, J.-L. Starck and F. Murtagh, "Inpainting and Zooming using Sparse Representations", submitted.*

. Initialize all s_k to zero

. Iterate j=1,...,Niter

- Iterate k=1,..,L

- Update the kth part of the current solution by fixing all other parts and minimizing:

$$J(s_k) = \left\| M(s - \sum_{i=1, i \neq k}^{L} s_i - s_k) \right\|_2^2 + \lambda \| T_k s_k \|_1$$

Which is obtained by a simple soft thresholding of :

$$s_r = M(s - \sum_{i=1, i \neq k}^{L} s_i)$$

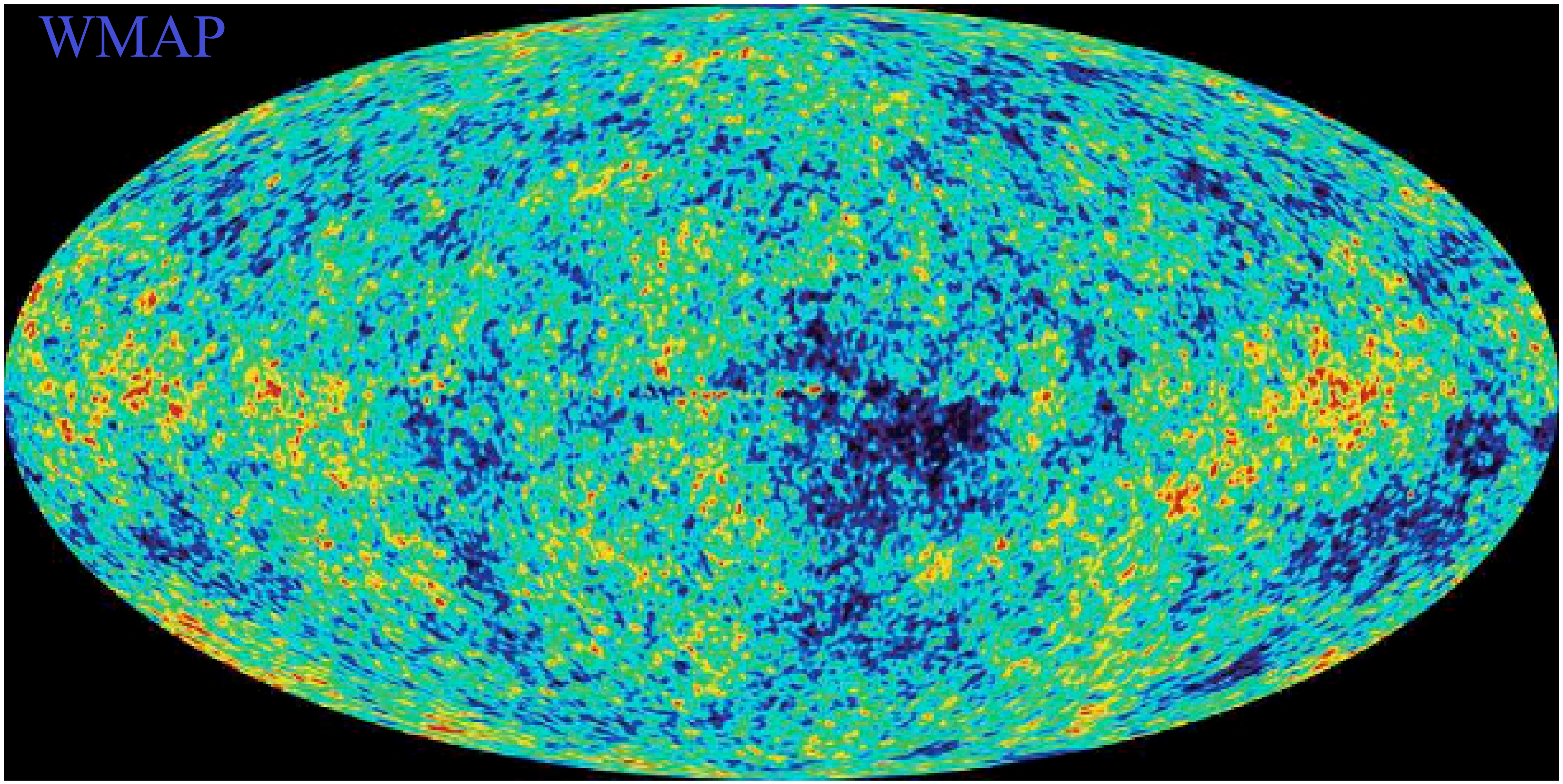
20%

50%

80%

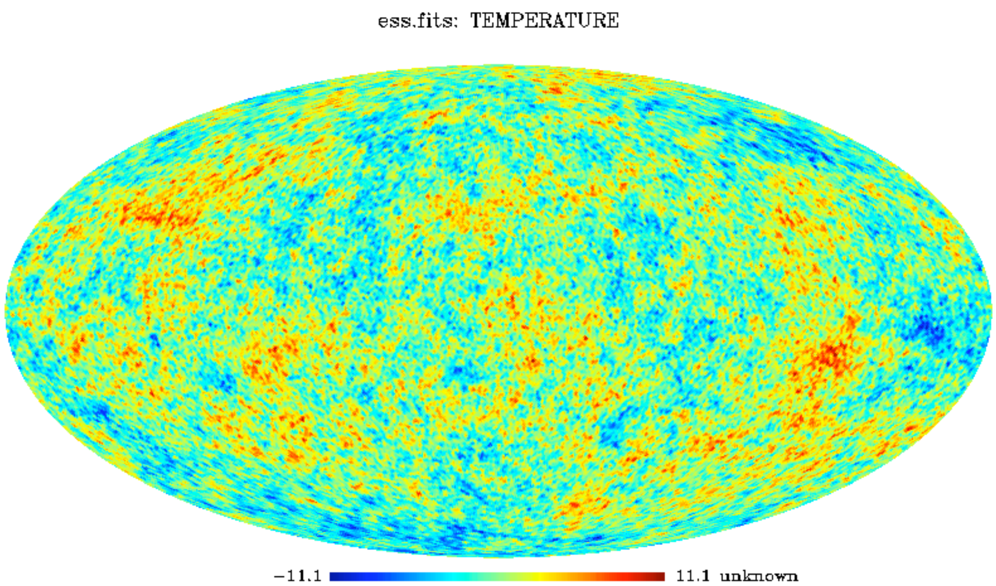
Inpainted with the curvelet dictionary (80% data missing)

Application in Cosmology



The cosmic Microwave Background is a relic radiation (with a temperature equals to 2.726 Kelvin) emitted 13 billion years ago when the Universe was about 370000 years old.

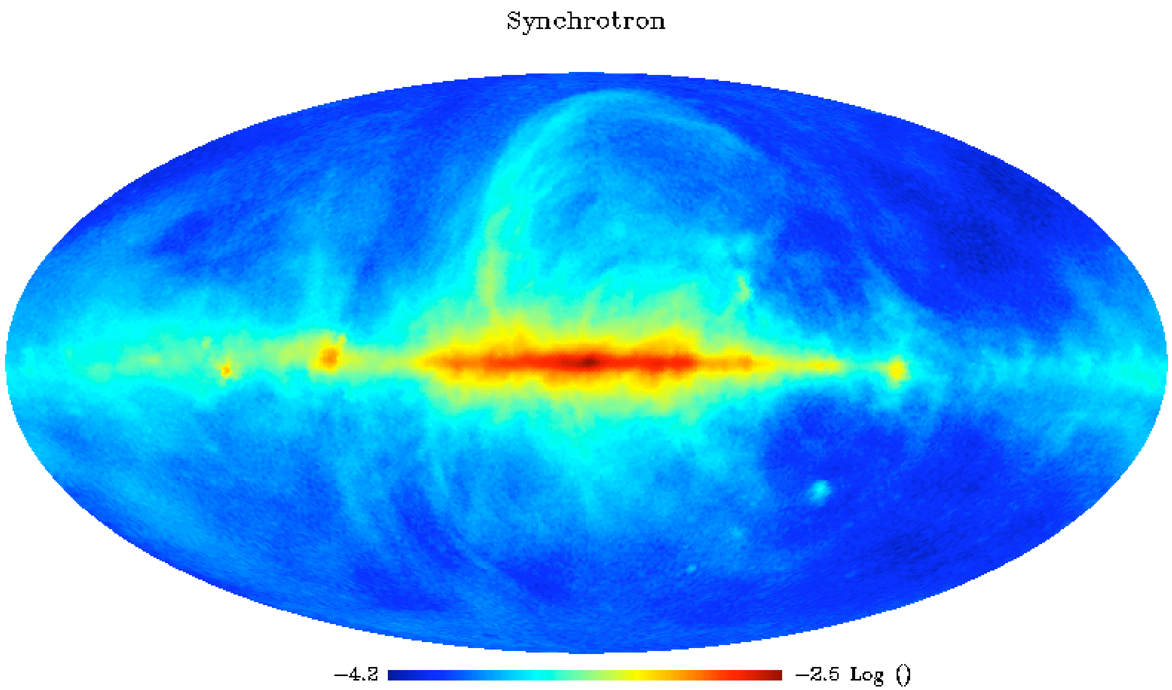
Wavelet, Ridgelet and Curvelet on the Sphere :

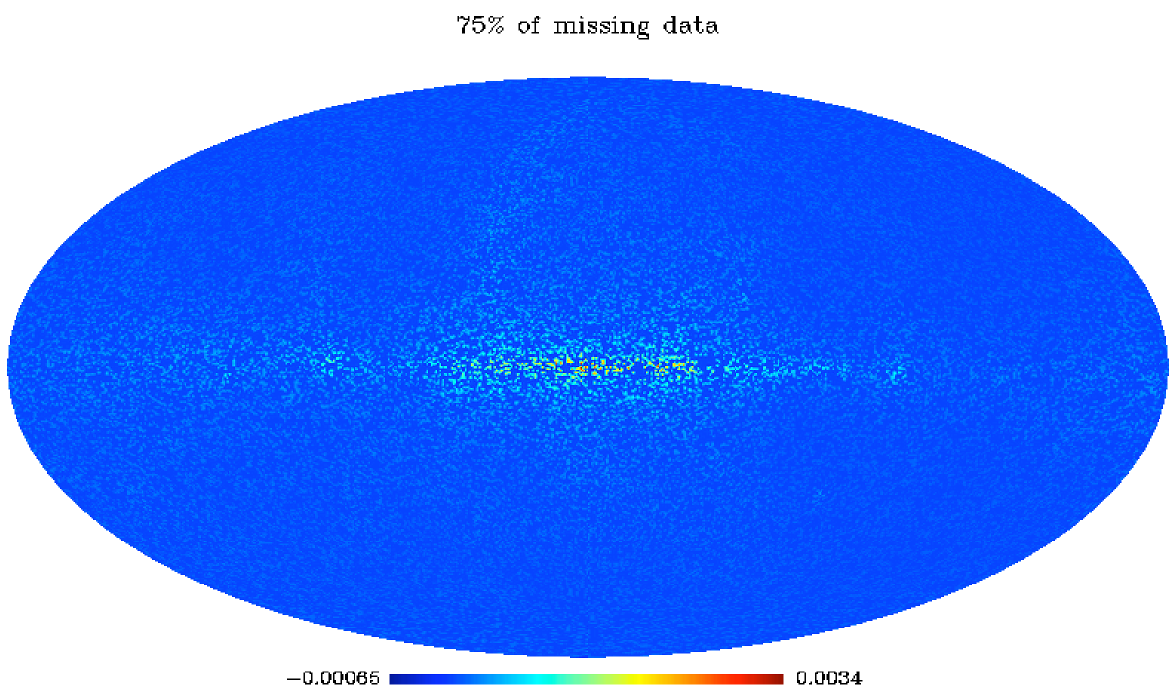


Wavelets, Ridgelets and Curvelets on the Sphere, *Astronomy & Astrophysics*, 446, 1191-1204, 2006.

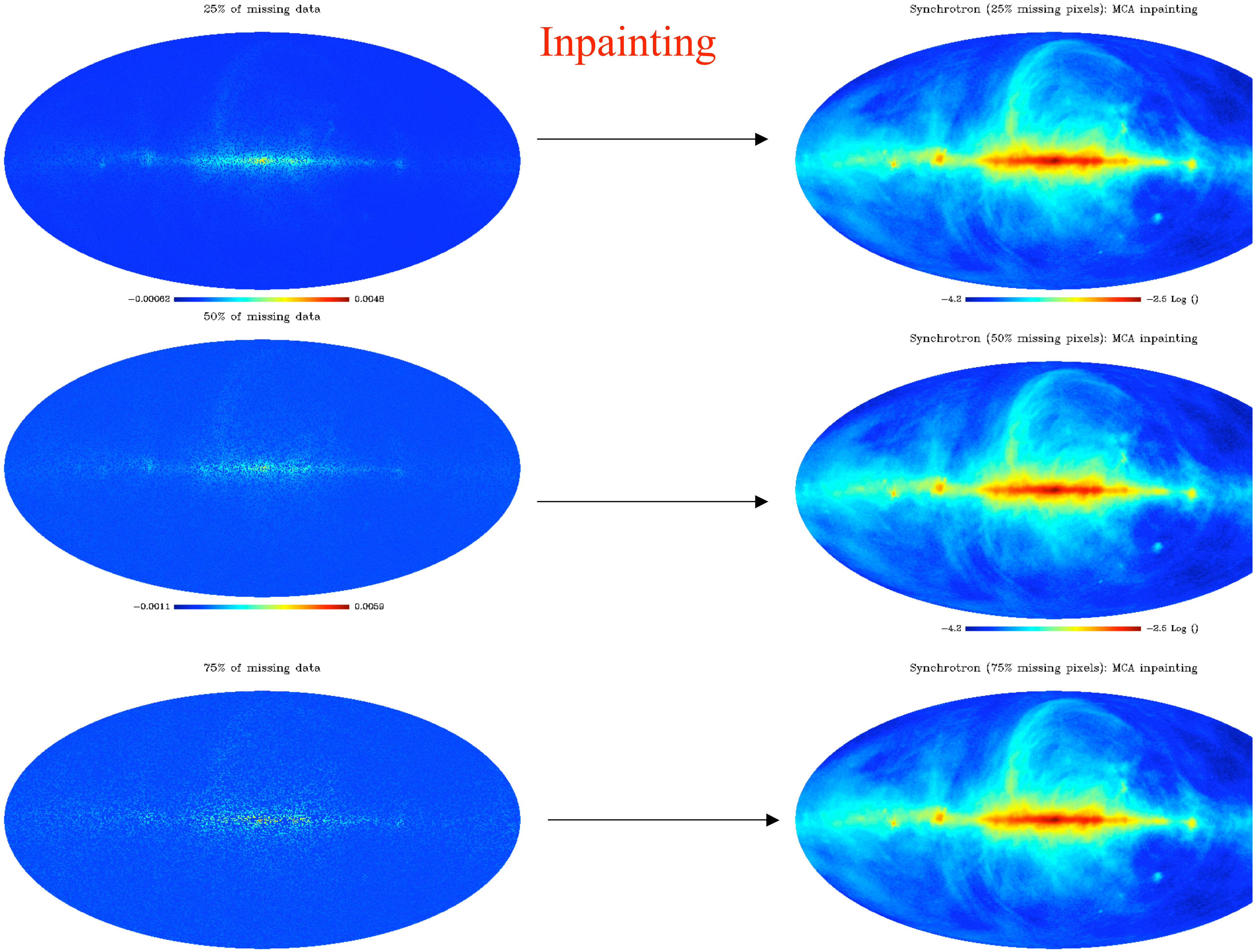
MR/S software **available** at: http://jstarck.free.fr/mrs.html

Multiscale transforms, Gaussianity tests
Denoising using Wavelets and Curvelets
Astrophysical Component Separation (ICA on the Sphere)


Synchrotron
−4.2
−2.5 Log ()

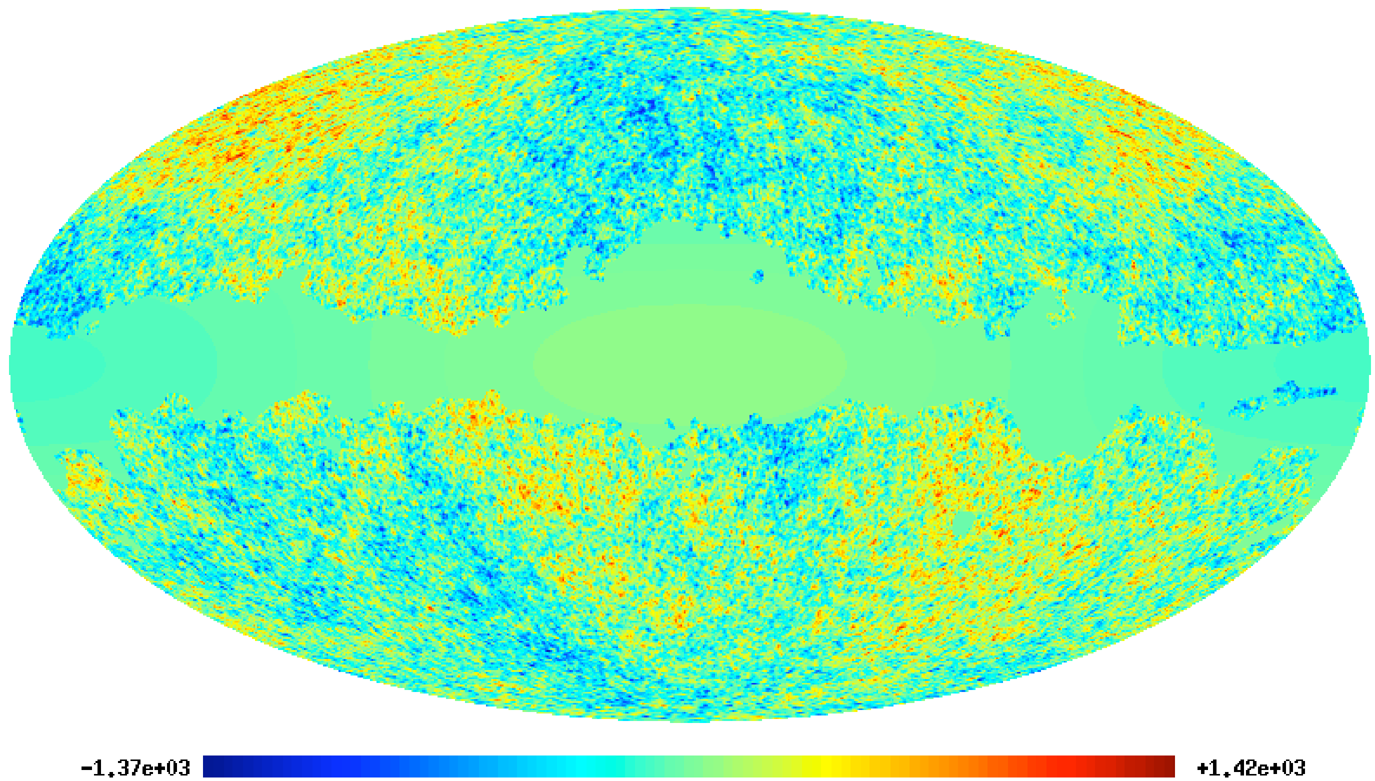

75% of missing data
−0.00065
0.0034

Inpainting



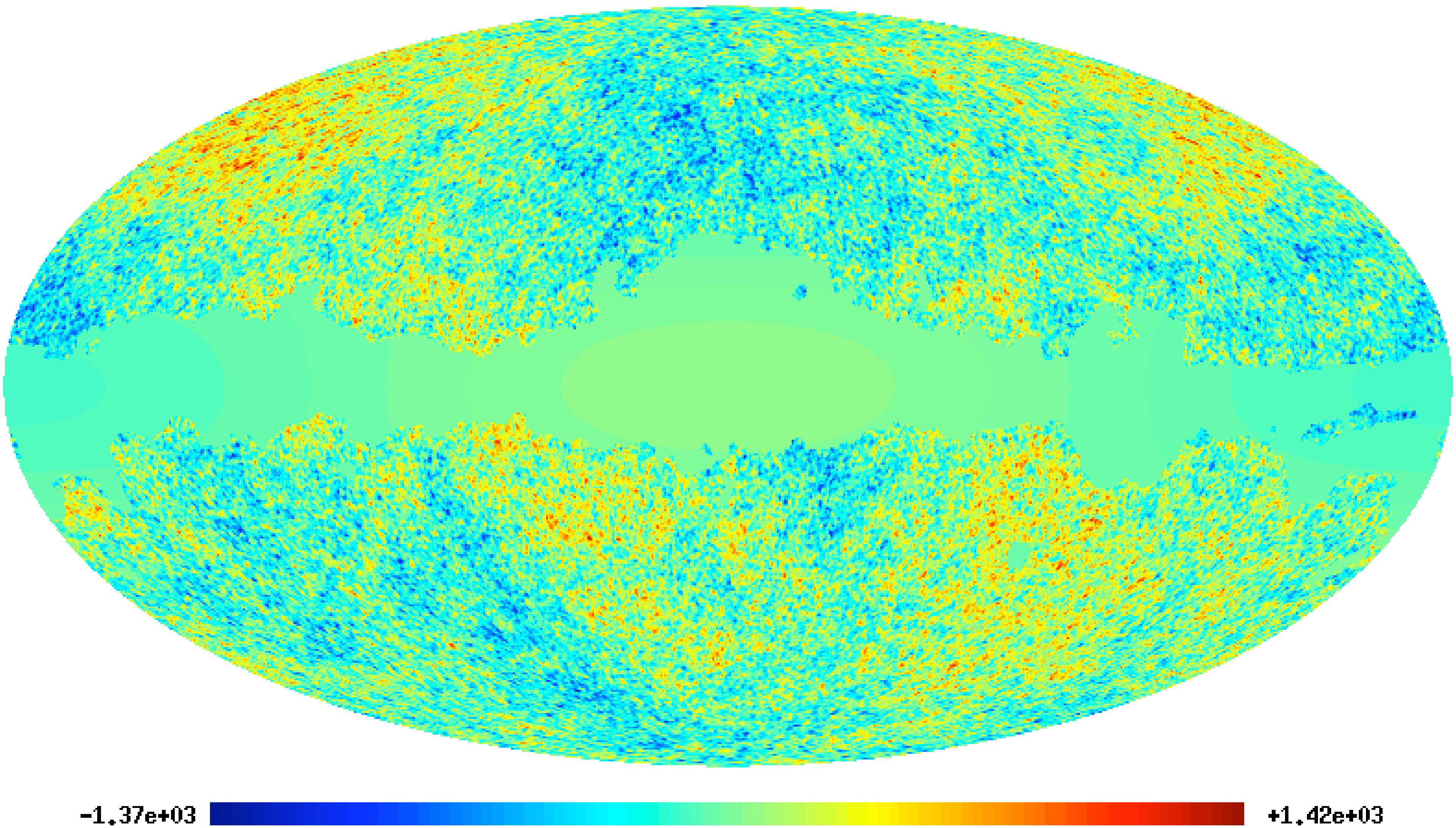
25% of missing data
−0.00062
0.0048
50% of missing data
−0.0011
0.0059
75% of missing data
Synchrotron (25% missing pixels): MCA inpainting
−4.2
−2.5 Log ()
Synchrotron (50% missing pixels): MCA inpainting
−4.2
−2.5 Log ()
Synchrotron (75% missing pixels): MCA inpainting

WHY INPAINTING IS USEFUL FOR THE CMB ?



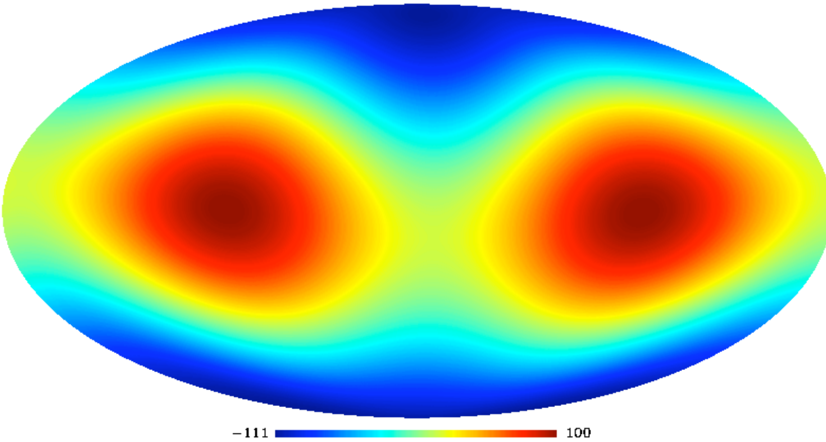
- Gaussianity test.
- Power estimation with the minimum of correlation.
- Any analysis where the mask is a problem.

Abrial et al, "Inpainting on the Sphere", Astronomical Data Analysis Conference IV, September 18-20, Marseille, 2006.

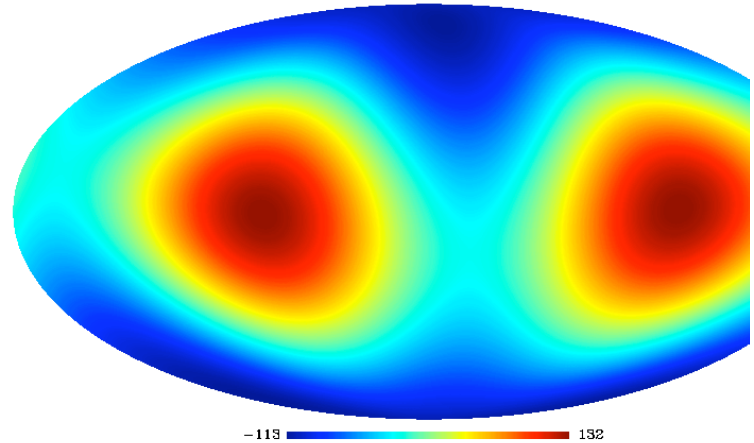


Abrial et al, "Inpainting on the Sphere", Astronomical Data Analysis Conference IV, September 18-20, Marseille, 2006.

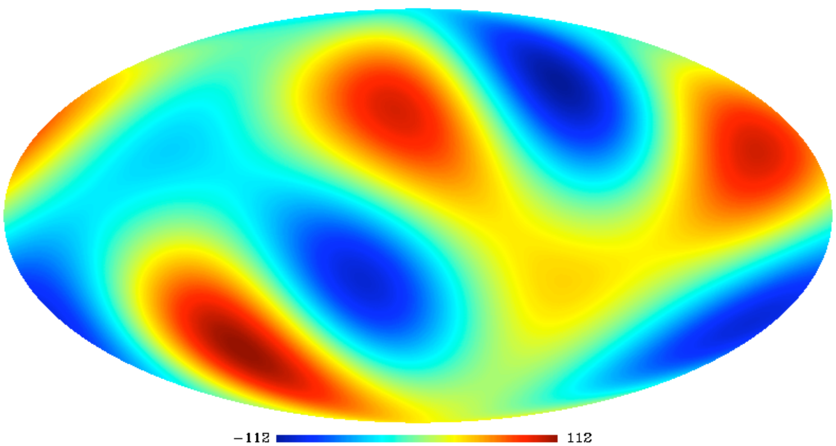
Simulated Data: l=2



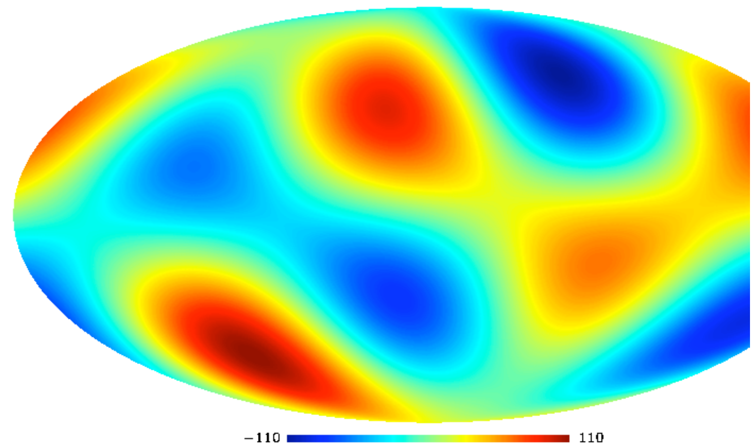
Simulated data (inpainting): l=2



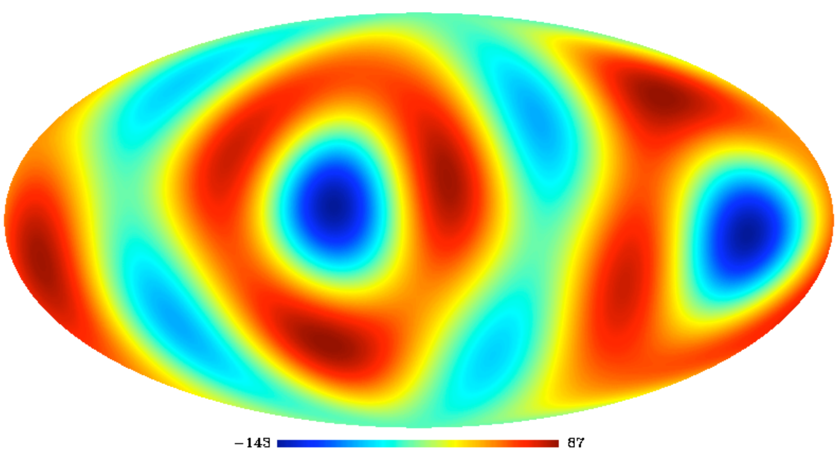
Simulated Data: l=3



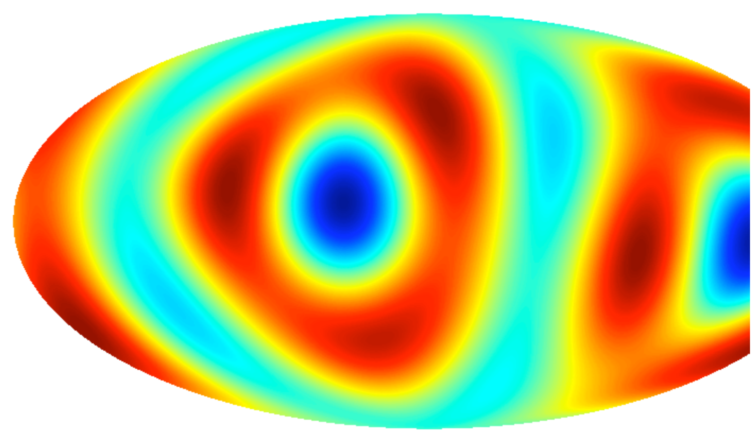
Simulated data (inpainting): l=3



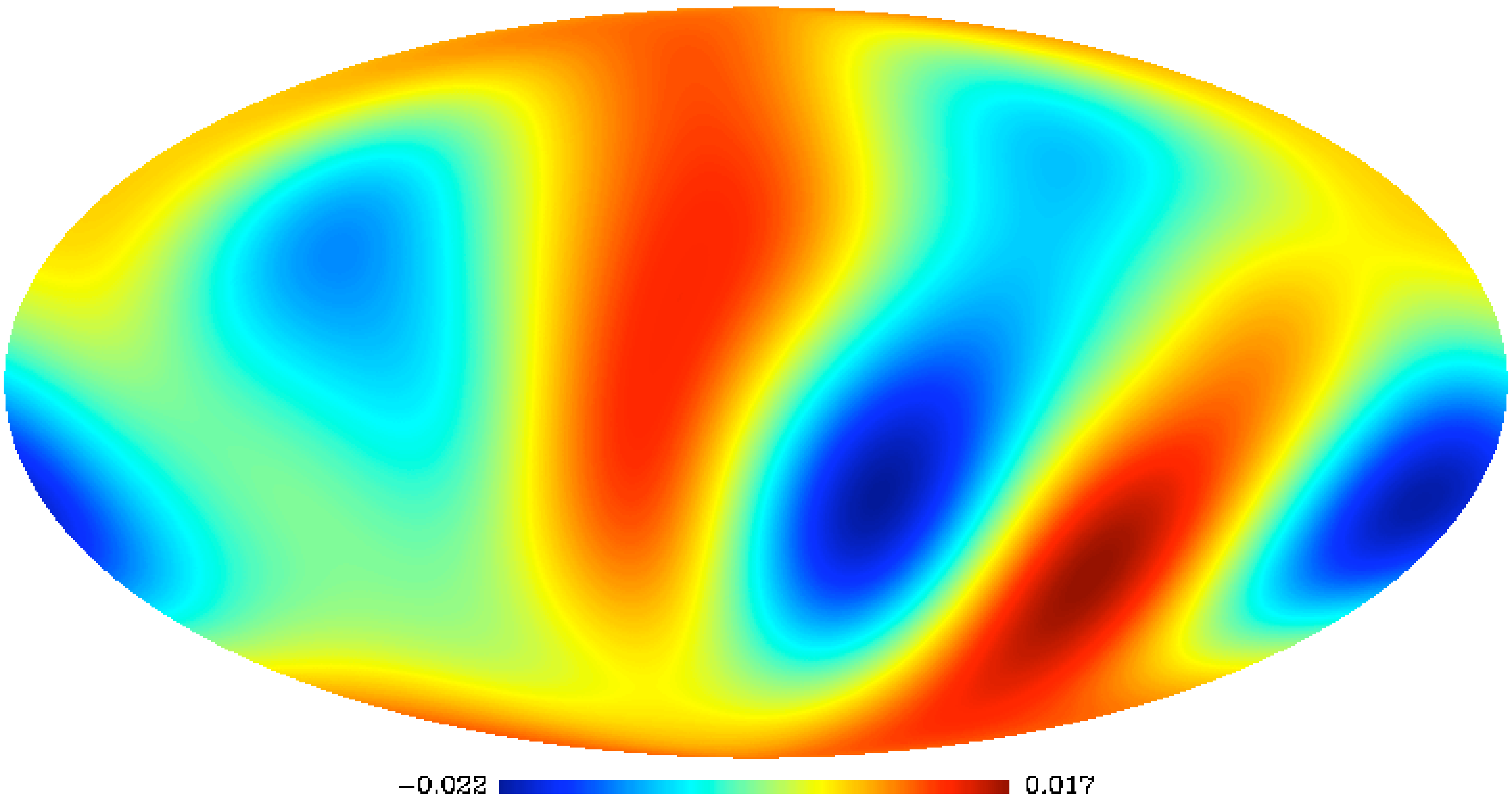
Simulated Data: l=4



Simulated data (inpainting): l=4



WMAP inpainting Scale 7



Multichannel MCA (MMCA)

$$X = AS \quad \text{or} \quad X_i = \sum_{k=1}^{K} a_{i,k} s_k, \;\; \exists T_k \text{ such that } \alpha_k = T_k s_k \text{ is sparse}$$

According to the MCA paradigm, each source is morphologically different from the others. Each source s_k is then well sparse in a specific basis Φ_k. Thus MMCA aims at solving the following minimization problem:

$$\min_{A,s_1,\ldots,s_k} = \sum_{l=1}^{m} \left\| X_l - \sum_{k=1}^{K} A_{k,l} s_k \right\|_2^2 + \lambda \sum_{k=1}^{K_i} \left\| T_k s_k \right\|_p$$

Both the source matrix S and the mixing matrix A are estimated alternately for fixed values of λ_k from a Maximum A Posteriori.

Defining a multichannel residual D_k : $\mathbf{D}_k = \mathbf{X} - \sum_{k' \neq k} a^{k'} s_{k'}$

the parameters are **alternately** estimated such that :

$$J(s_k) = \left\| D_k - s_k \right\|_2^2 + \lambda_n \left\| T_k s_k \right\|_p$$

J. Bobin et al, "Morphological Diversity and Source Separation", IEEE Transaction on Signal Processing, Vol 13, 7, pp 409--412, 2006.

The MMCA Algorithm

. Initialize all s_k to zero

. Iterate t=1,...,Niter

- Iterate k=1,..,L

  Update the kth part of the current solution by fixing all other parts and minimizing:

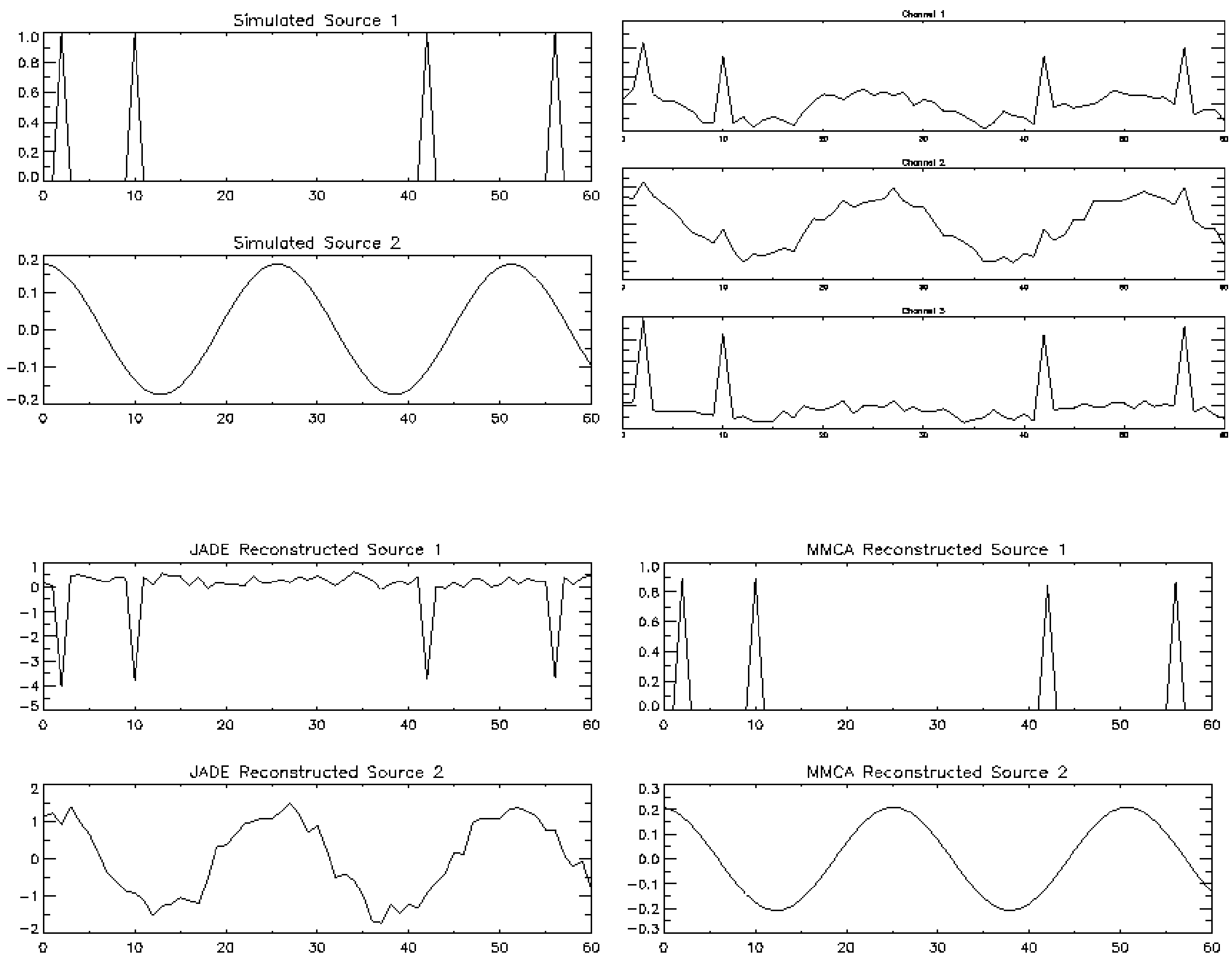
  $$J(s_k) = \left\| D_k - s_k \right\|_2^2 + \lambda_t \left\| T_k s_k \right\|_1 \quad \text{with} \quad D_k = a^{k^T}(X - \sum_{i=1, i \neq k}^{L} a^i s_i$$

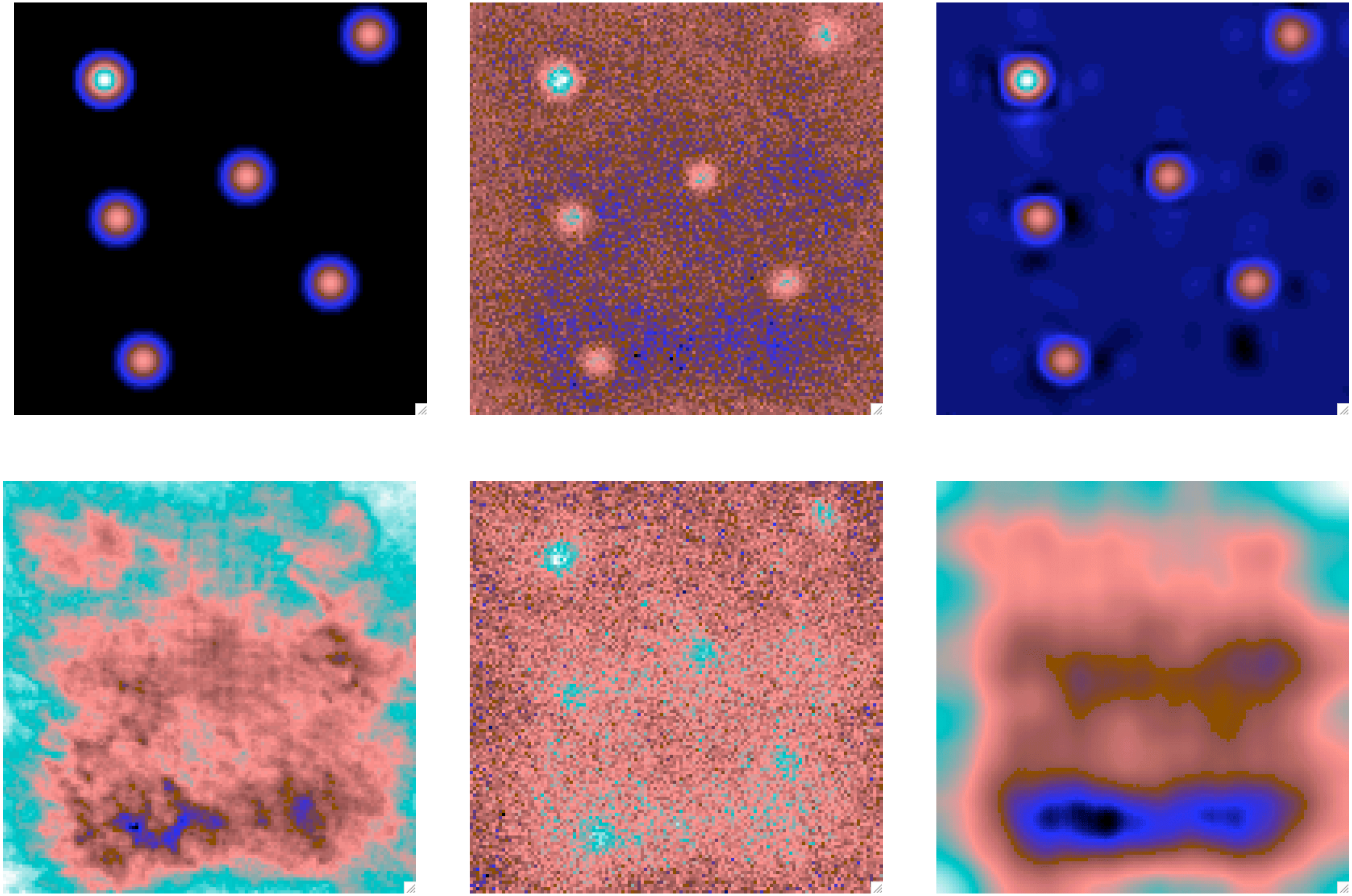
  which is obtained by a simple hard/soft thresholding of D_k

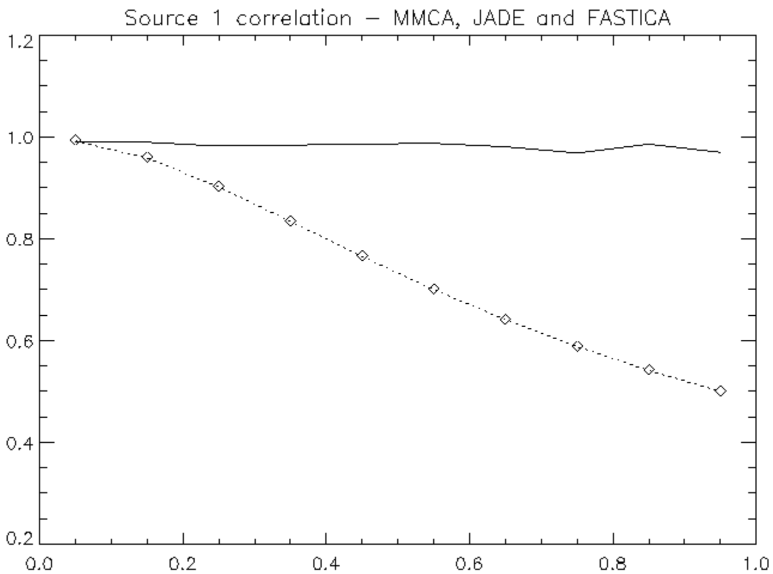
- estimation of a^k assuming all s_l and $a^l_{l \neq k}$ fixed

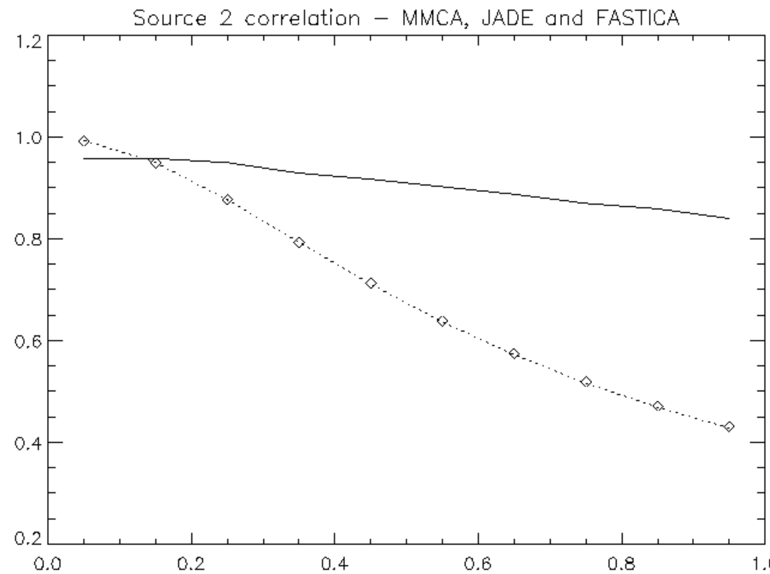
  $$a^k = \frac{1}{s_k {s_k}^T} D_k {s_k}^T$$

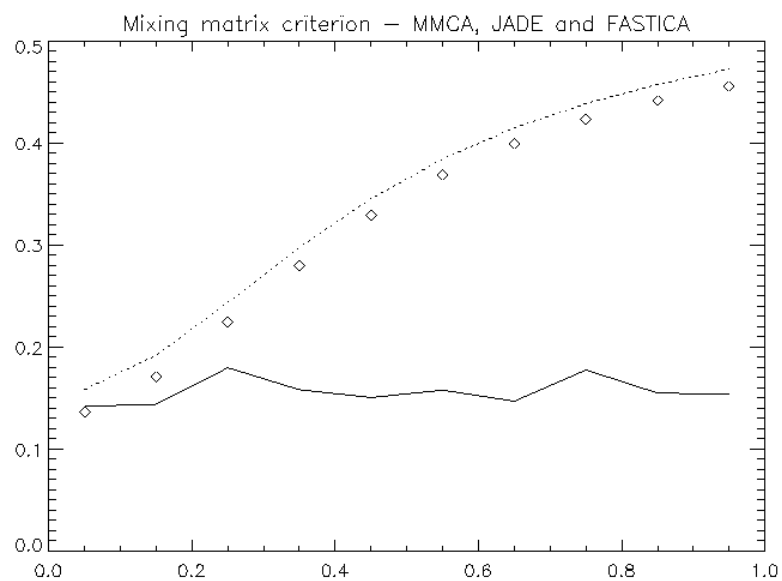
- Decrease λ_t


Simulated Source 1
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
Simulated Source 2
0.2
0.1
0.0
−0.1
−0.2
Channel 1
Channel 2
Channel 3
JADE Reconstructed Source 1
1
0
−1
−2
−3
−4
−5
JADE Reconstructed Source 2
2
1
0
−1
−2
MMCA Reconstructed Source 1
MMCA Reconstructed Source 2
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3




Source 1 correlation – MMCA, JADE and FASTICA
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.2
0.4
0.6
0.8
1.0


Source 2 correlation – MMCA, JADE and FASTICA
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.2
0.4
0.6
0.8
1.0


Mixing matrix criterion – MMCA, JADE and FASTICA
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Generalized MCA (GMCA)

Source: $S=\left[s_1,\ldots,s_n\right]$ Data: $X=\left[x_1,\ldots,x_m\right]=AS$

We now assume that the sources are linear combinations of morphological components :

$$s_i=\sum_{k=1}^{K}c_{i,k} \quad \text{such that} \quad \alpha_{i,k}=T_{i,k}c_{i,k} \text{ sparse}$$

$$\Longrightarrow \qquad X_l=\sum_{i=1}^{n}A_{i,l}s_i=\sum_{i=1}^{n}A_{i,l}\sum_{k=1}^{K}c_{i,k}$$

$$\phi=\left[\left[\phi_{1,1},\ldots,\phi_{1,K}\right],\ldots,\left[\phi_{n,1},\ldots,\phi_{n,K}\right],\right], \quad \alpha=S\phi^t=\left[\left[\alpha_{1,1},\ldots,\alpha_{1,K}\right],\ldots,\left[\alpha_{n,1},\ldots,\alpha_{n,K}\right]\right]$$

GMCA aims at solving the following minimization:

$$\min_{A,c_{1,1},\ldots,c_{1,K},\ldots,c_{n,1},\ldots,c_{n,K}}=\sum_{l=1}^{m}\left\|X_l-\sum_{i=1}^{n}A_{i,l}\sum_{k=1}^{K}c_{i,k}\right\|_2^2+\lambda\sum_{i=1}^{n}\sum_{k=1}^{K}\left\|T_{i,k}c_{i,k}\right\|_p$$

The GMCA Algorithm

. Initialize all c_k to zero, $\lambda_1 = \max(\alpha), \delta = \max(\alpha)/\text{Niter}$

. Iterate t=1,...,Niter

- Iterate i=1,..,NbrSource

  Defining a multichannel residual $\mathbf{D_i}$: $D_i = X - \sum_{i' \neq i} a^{i'} s_{i'}$

  Iterate k=1,..,K_k

  - Least square estimate of $c_{i,k}$: $l_{i,k} = \frac{1}{a^{i^T} a^i} a^{i^T} (D_i - a^i \sum_{k' \neq k} c_{i,k'})$

  - Minimize: $J(\tilde{l}_{i,k}) = \left\| l_{i,k} - \tilde{l}_{i,k} \right\|_2^2 + \lambda_t \left\| T_{i,k} \tilde{l}_{i,k} \right\|_1$

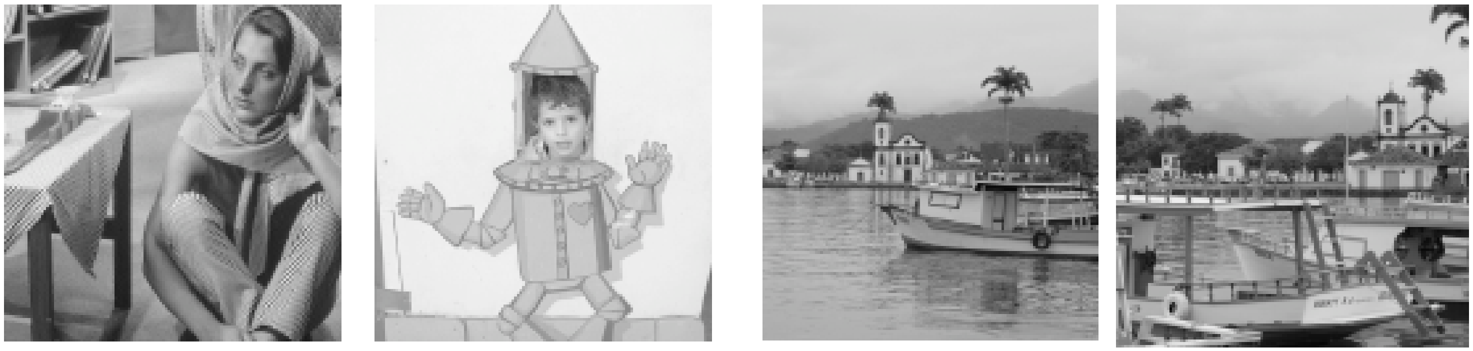
    which is obtained by a simple hard/soft thresholding of $l_{i,k}$

  $s_k = \sum_i l_{k,i}$

- $S = [s_1, \ldots, s_K]^t$
- Estimation of the matrix A: $A = XS^t (SS^t)^{-1}$
- Decrease $\lambda_{t+1} = \lambda_t - \delta$

A first result (1)

Original
Sources



Mixtures

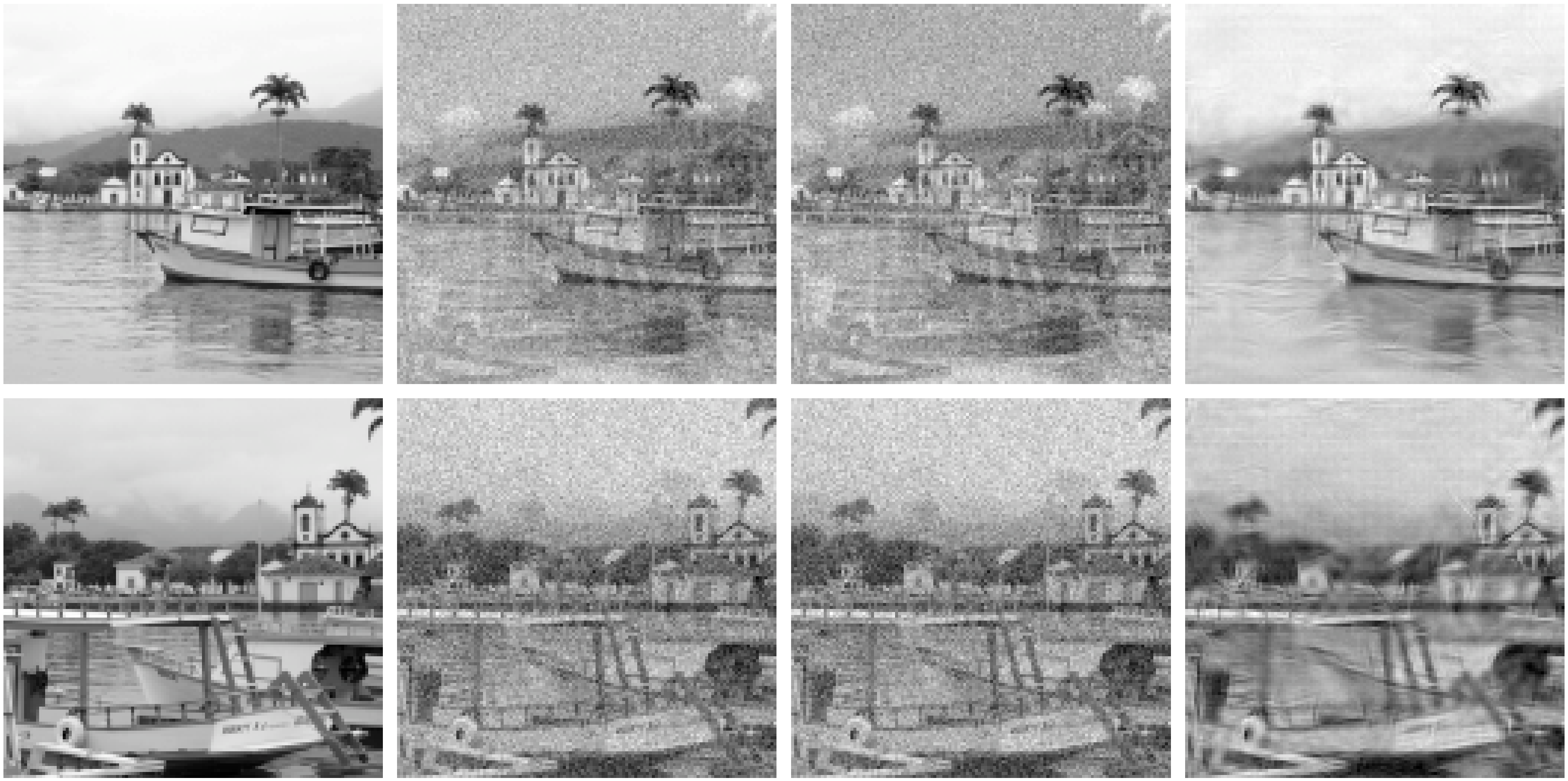
Noiseless experiment, 4 random mixtures, 4 sources

A first result (2)




2 mixtures SNR = 10.4dB

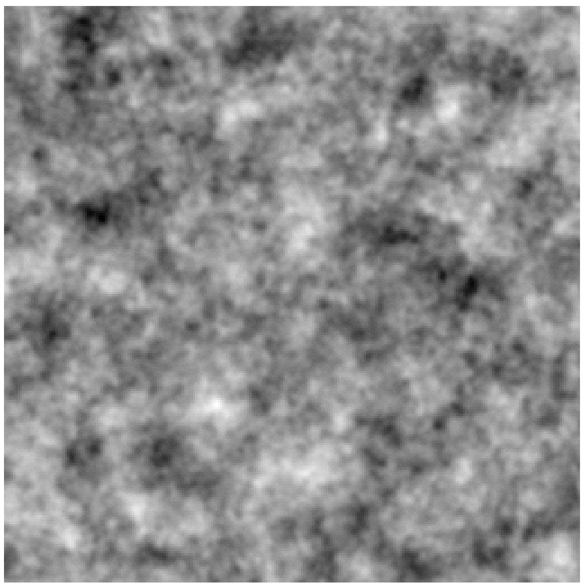
$\Phi =$ Curvelets + DCT



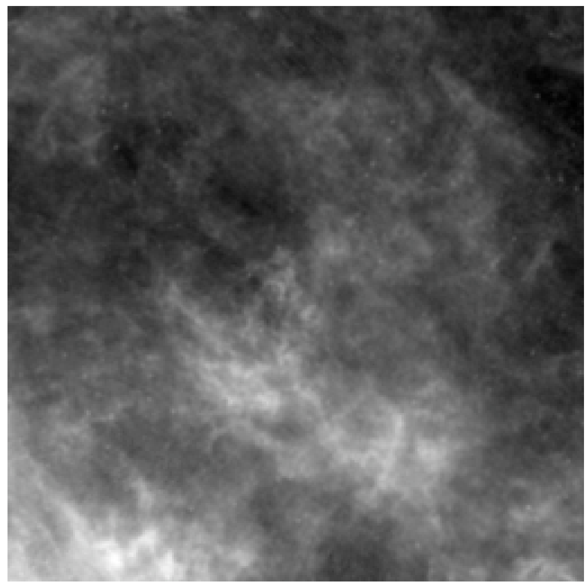
Sources Mixtures JADE

The source images: 300x300 pixels corresponding to a field of 12,5x12,5 degres.

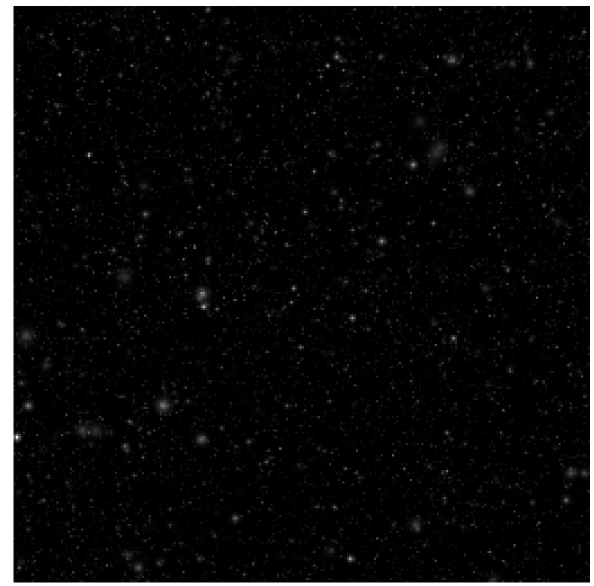
CMB



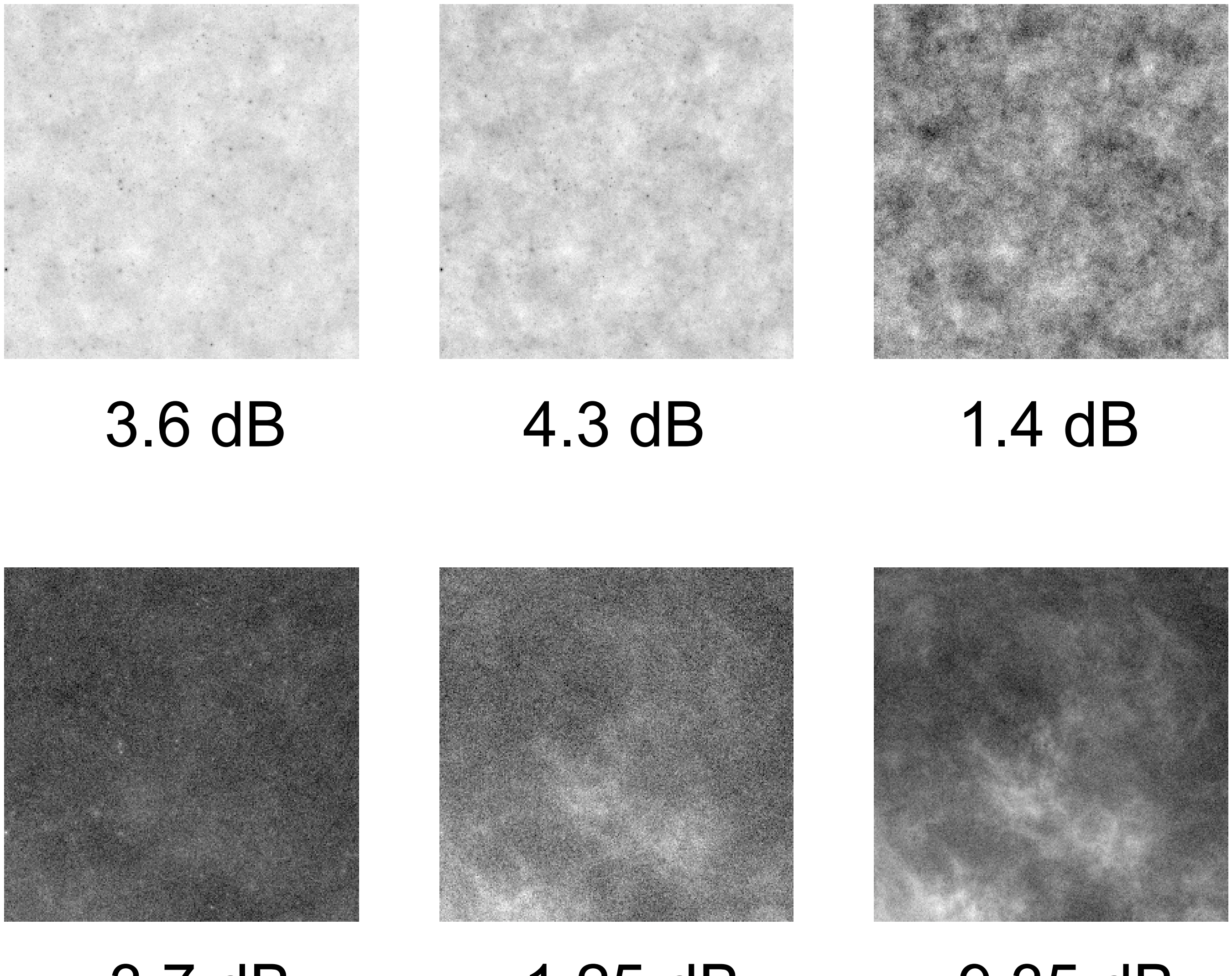
DUST



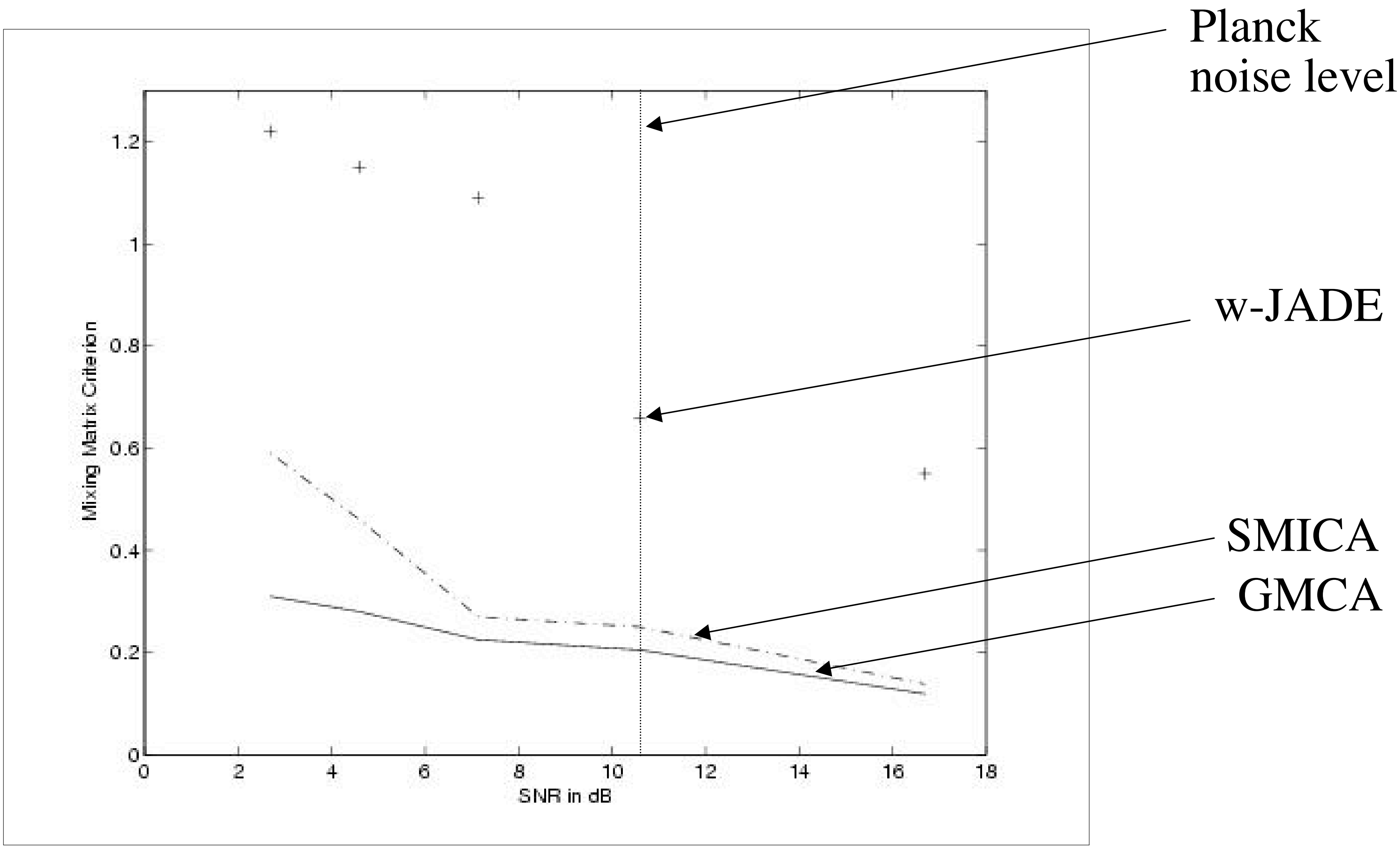
SZ

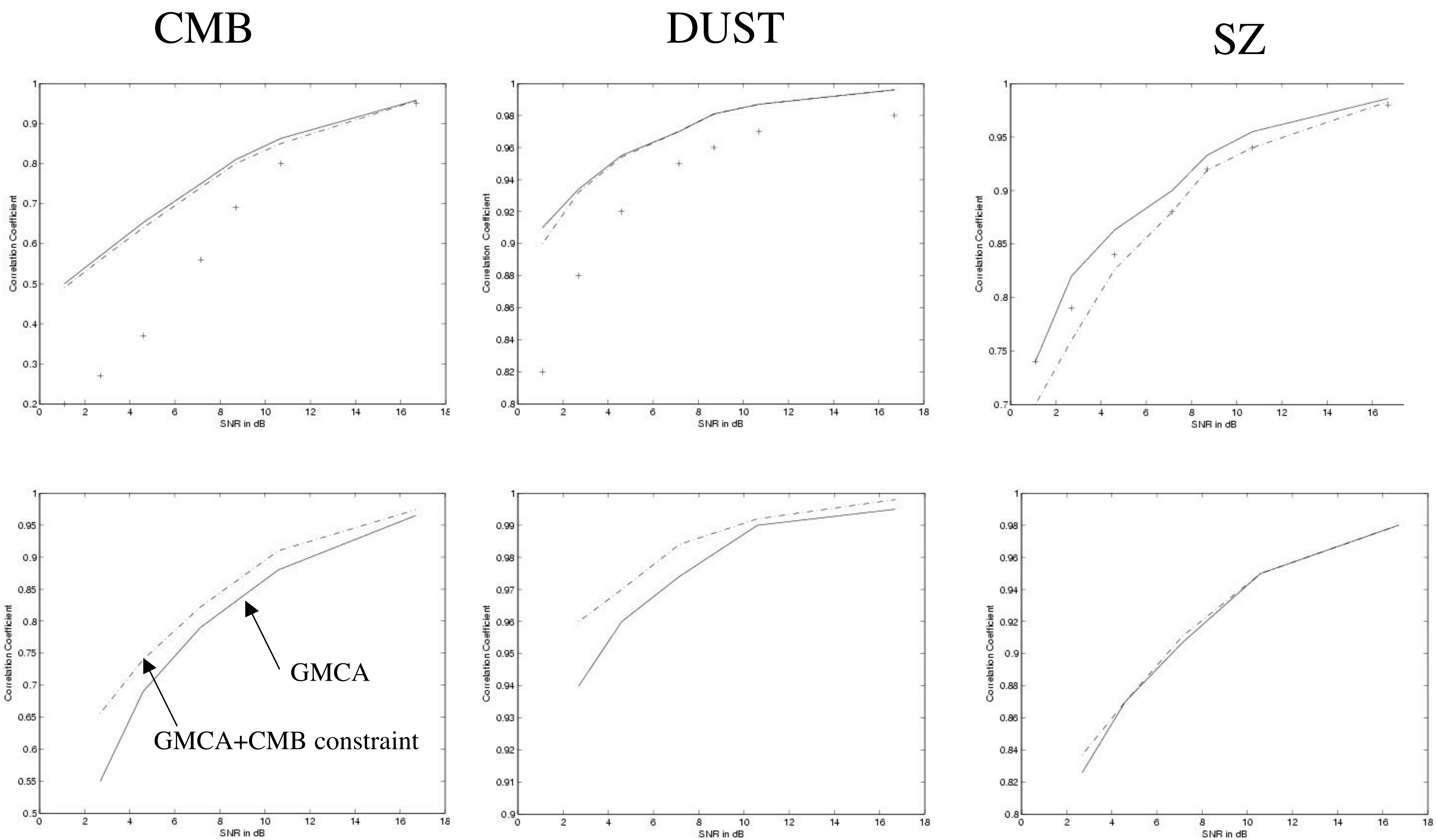


The six simulated HFI Channels
(100, 143, 217, 353, 545 and 857 GHz)



Mixing Matrix Estimation Error





Bobin et al, "CMB and SZ reconstruction using GMCA", Astronomical Data Analysis Conference IV, September 18-20, Marseille, 2006.

Conclusions

- MCA method can be useful in different applications such texture separation or inpainting.

.**Redundant Multiscale Transforms and their Application for Morphological Component Analysis,** *Advances in Imaging and Electron Physics, 132, 2004*.

. **Image Decomposition Via the Combination of Sparse Representation and a Variational Approach**, *IEEE Transaction on Image Processing*, *14, 10, pp 1570--1582, 2005.*

. **Simultaneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA),** ACHA, *19, pp. 340-358, 2005.*

- The MMCA algorithm brings a very strong and robust component separation as long as the MMCA hypothesis is verified (sources are sparsified in different bases) i.e. for morphologically diverse sources.

. **Morphological Diversity and Source Separation",** *IEEE Trans. on Signal Processing letters,* Vol 13, 7, pp 409--412, 2006.

- GMCA is more general, and can be applied for many applications.

More MCA experiments available at http://jstarck.free.fr/mca.html and Jalal Fadili's web page (http://www.greyc.ensicaen.fr/~jfadili).