Non local image processing

Jean-Michel Morel

Ecole Normale Supérieure de Cachan

Antoni Buades Capó

CNRS, Université Paris Descartes

Thanks for grants: CYCIT, ONR, CNES, DGA

APPLICATIONS OF NONLOCAL HEAT EQUATION

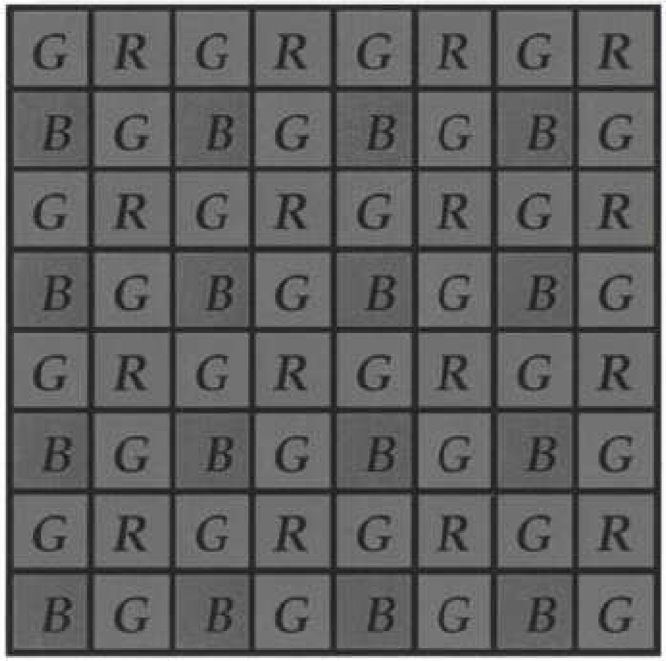
- A. Buades, B. Coll, and J.M M., A review of image denoising methods with a new one, Multiscale Modeling and Simulation, 4 (2), 2005

A noisy image…

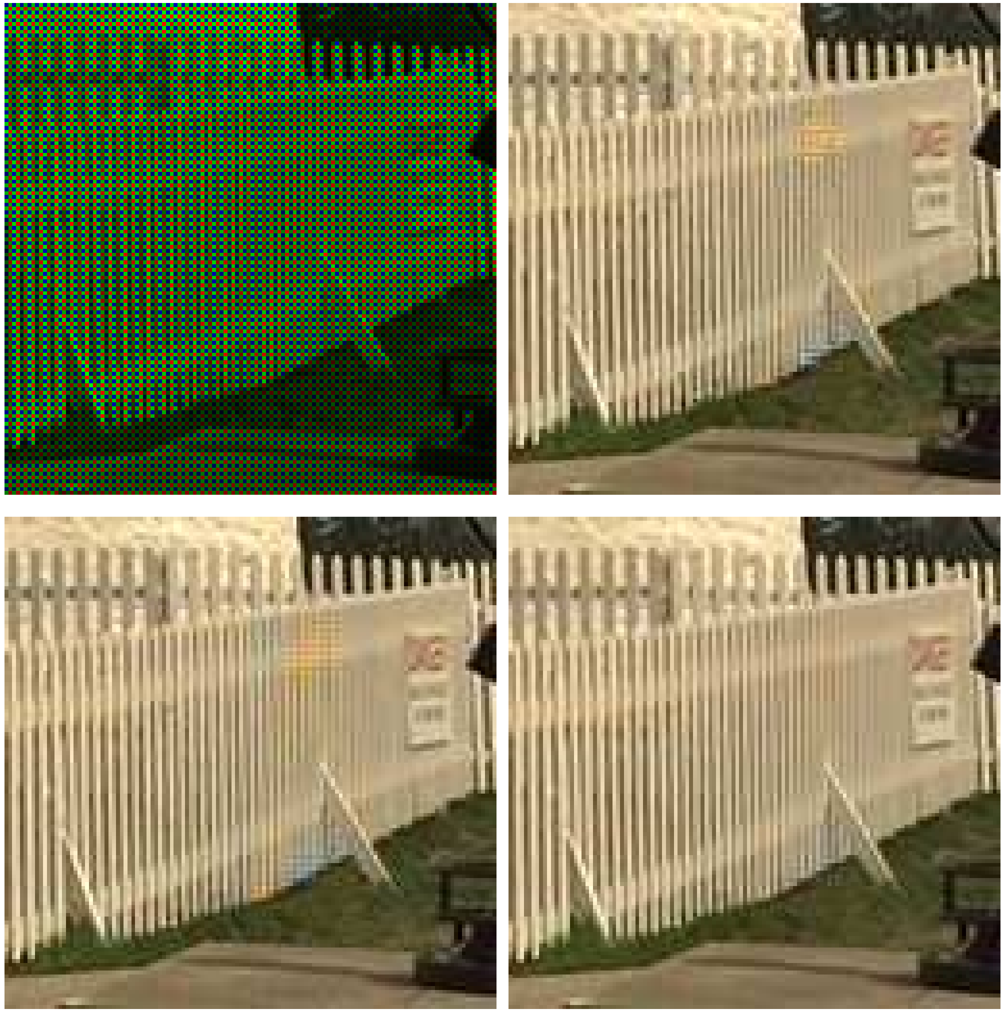
…enhanced with DxO Noise

DxO Noise is an original, low cost implementation of the "*Non Local Means*" noise removal approach, recently introduced by Professor Jean Michel Morel's team[1].

Color image demosaicing



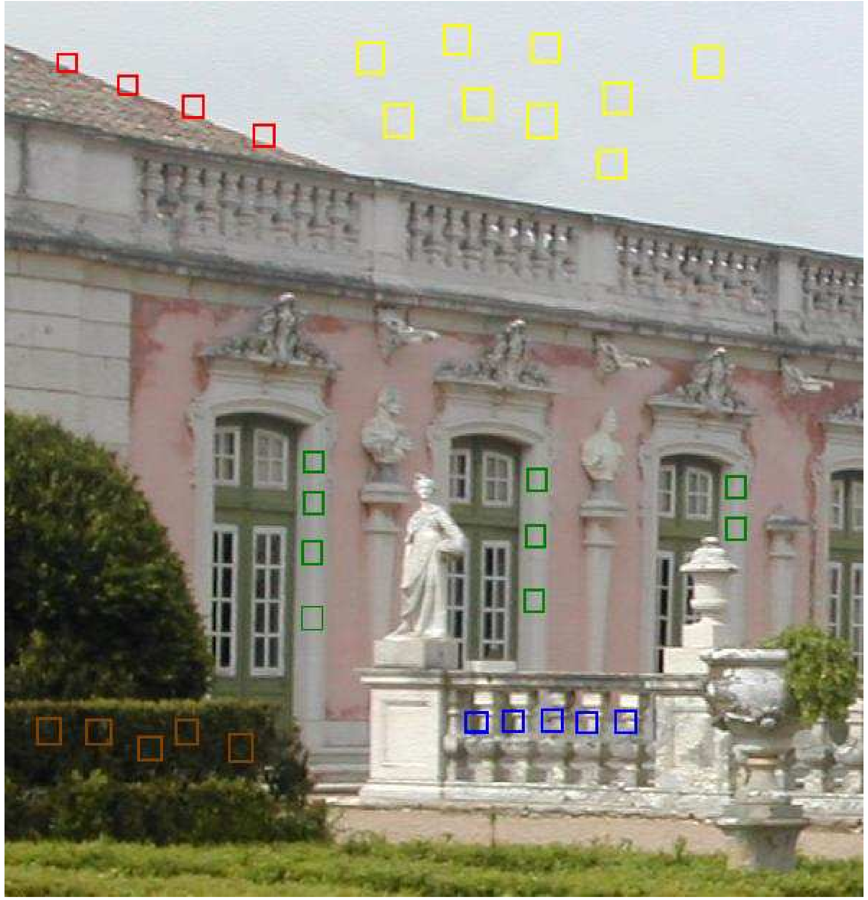
Demosaicing of color images



THE EFROS-LEUNG IMAGE AUTOSIMILARITY

A. Efros and Th. K. Leung Texture Synthesis by Non-parametric Sampling, IEEE International Conference on Computer Vision, Corfu, Greece, September 1999

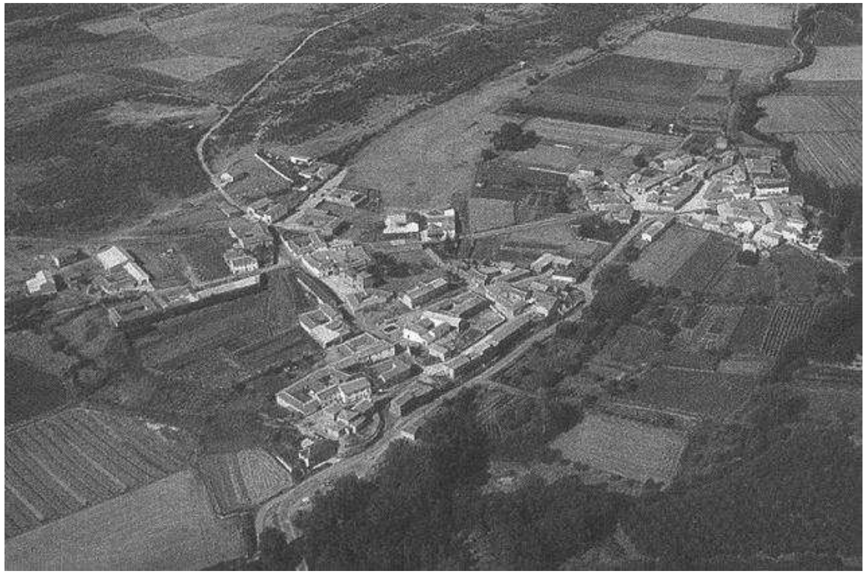
Image autosimilarity

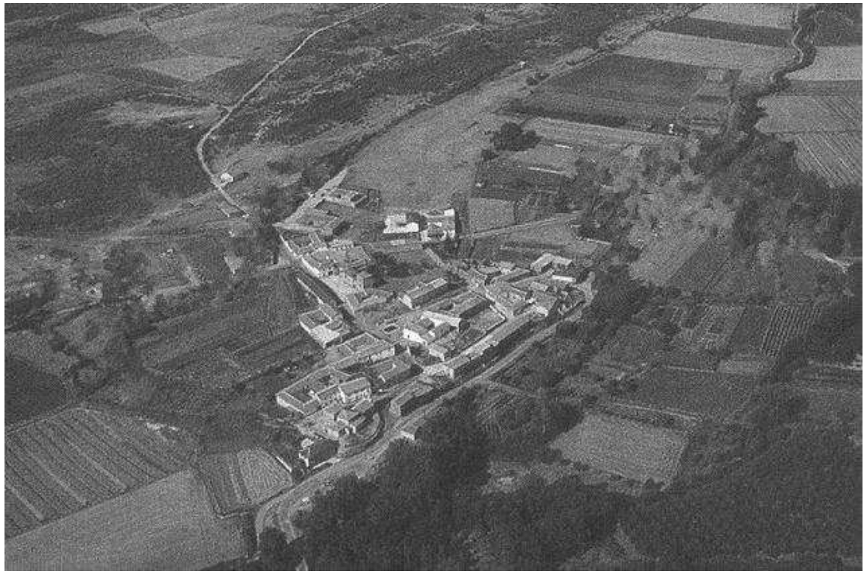


Groups of similar windows in a digital image, long range interaction. First used by Efros and Leung for texture synthesis. Fourier : too global, not geometrically adaptive. Wavelets: not adaptive enough. Main idea: the image generates its own non local model because self-similarities are non local.

- Example 1 : removing text from images by the Efros Leung algorithm ("inpainting")

- Example 2 : changing the content of aerial views





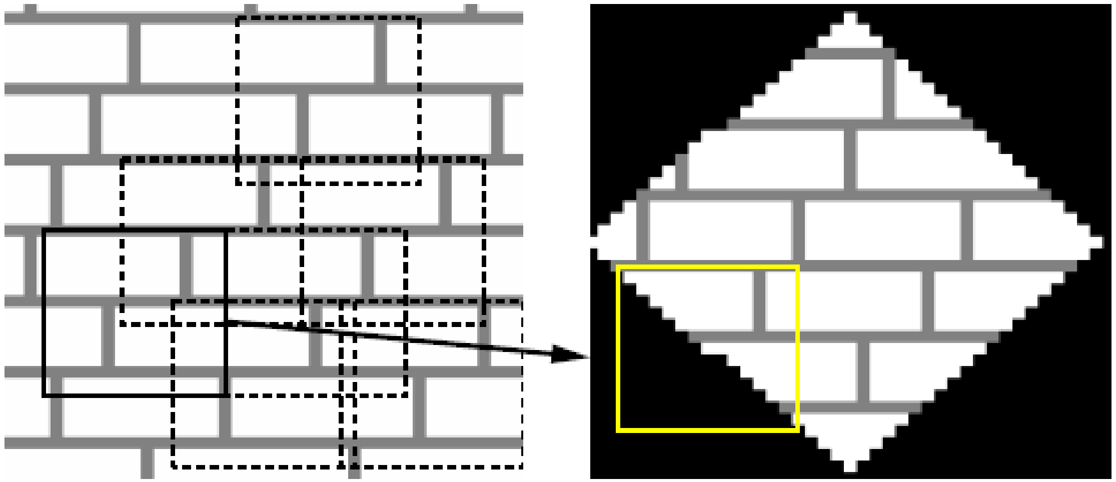
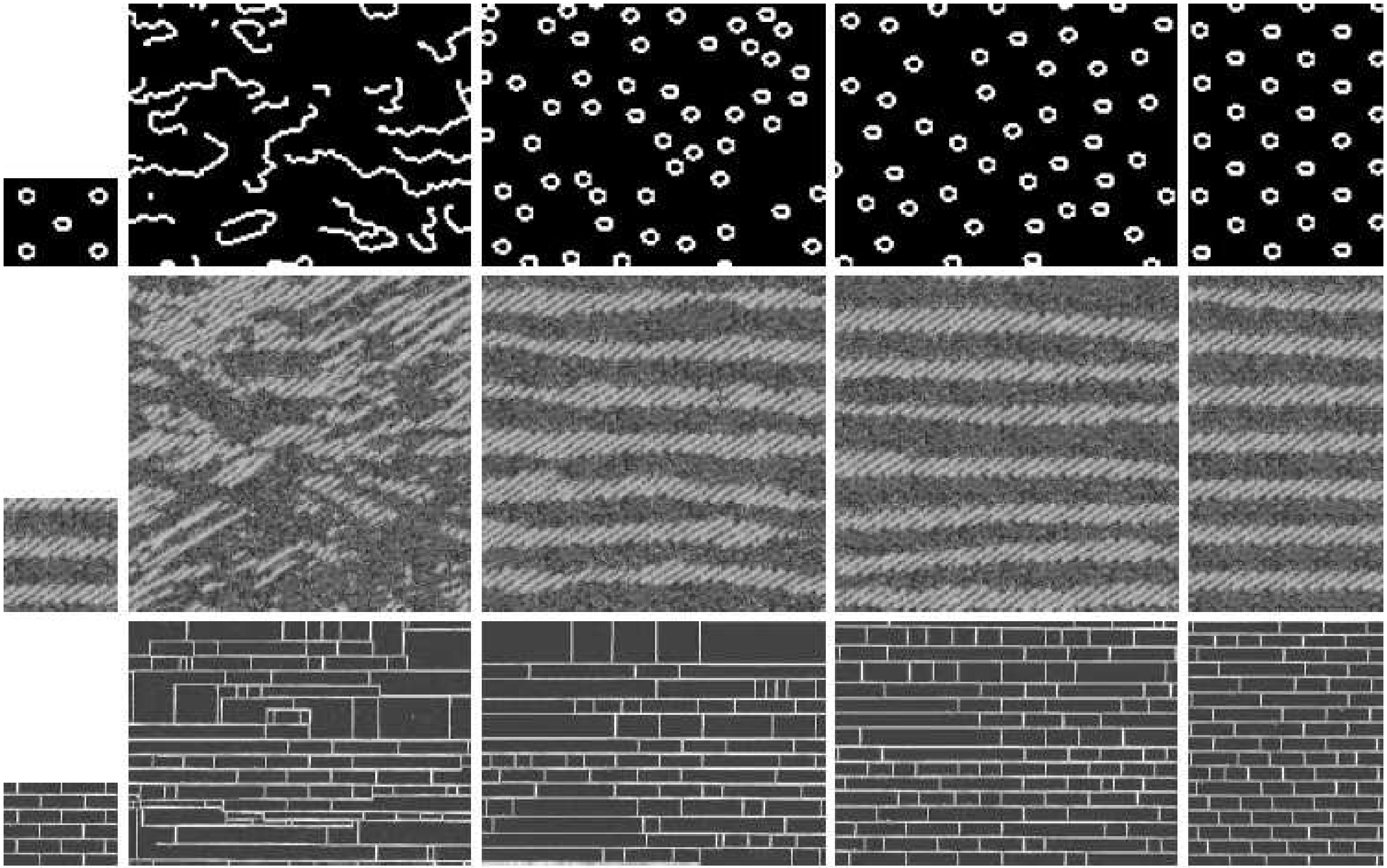


Figure 1. Algorithm Overview. Given a sample texture image (left), a new image is being synthesized one pixel at a time (right). To synthesize a pixel, the algorithm first finds all neighborhoods in the sample image (boxes on the left) that are similar to the pixel's neighborhood (box on the right) and then randomly chooses one neighborhood and takes its center to be the newly synthesized pixel.

Example 3: texture synthesis from samples



NOISE

The main assumption on noise

- **Main Hypothesis** In a digital image, the noise $n(i)$ at each pixel i only depends on the original pixel value $\Phi(i)$ and is additive, i.i.d. for all pixels $j \in J(i)$ with the same original value as i.
- $J(i)$ is the **neighborhood** of i. The challenge is finding $J(i)$ for every i. The simplest idea to do so is to *assume that all pixels with the same observed value $u(i)$ have the same noise model*: neighborhood filters.
- A more sophisticated use of the Hypothesis : *for a given pixel in an image, detect all pixels which have the same underlying model.*
- By the Hypothesis each j in $J(i)$ obeys a model $u(j) = v(i) + n(j)$ where $n(j)$ are i.i.d. It is then licit to perform a denoising of $u(i)$ by replacing it by

$$NFu(i) =: \frac{1}{|J(i)|} \sum_{j \in J(i)} u(j).$$

- By the variance formula for independent variables one then obtains $NFu(i) = v(i) + \tilde{n}(i)$ where

$$\mathrm{Var}(\tilde{n}(i)) = \frac{1}{|J(i)|} \mathrm{Var}(n(i)).$$

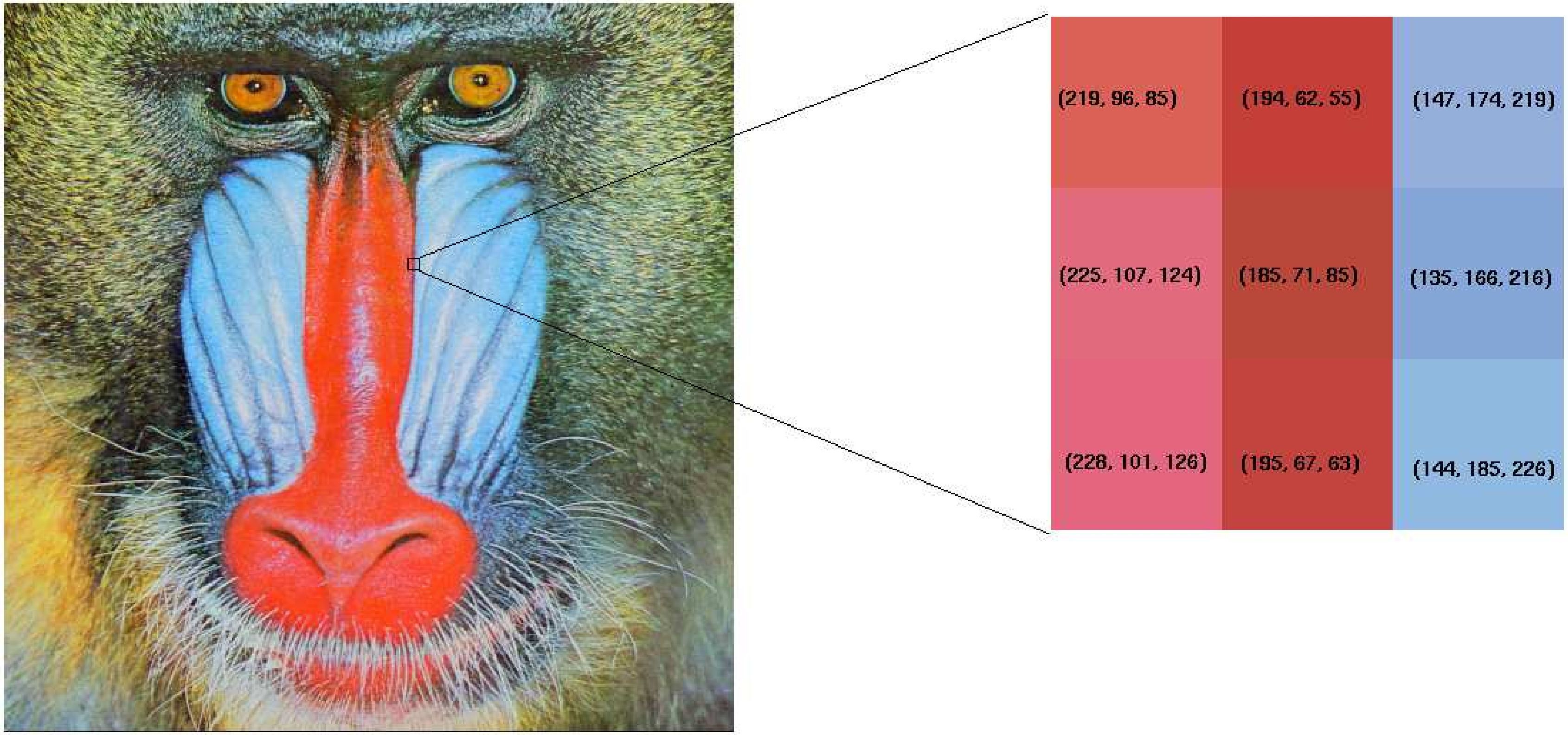
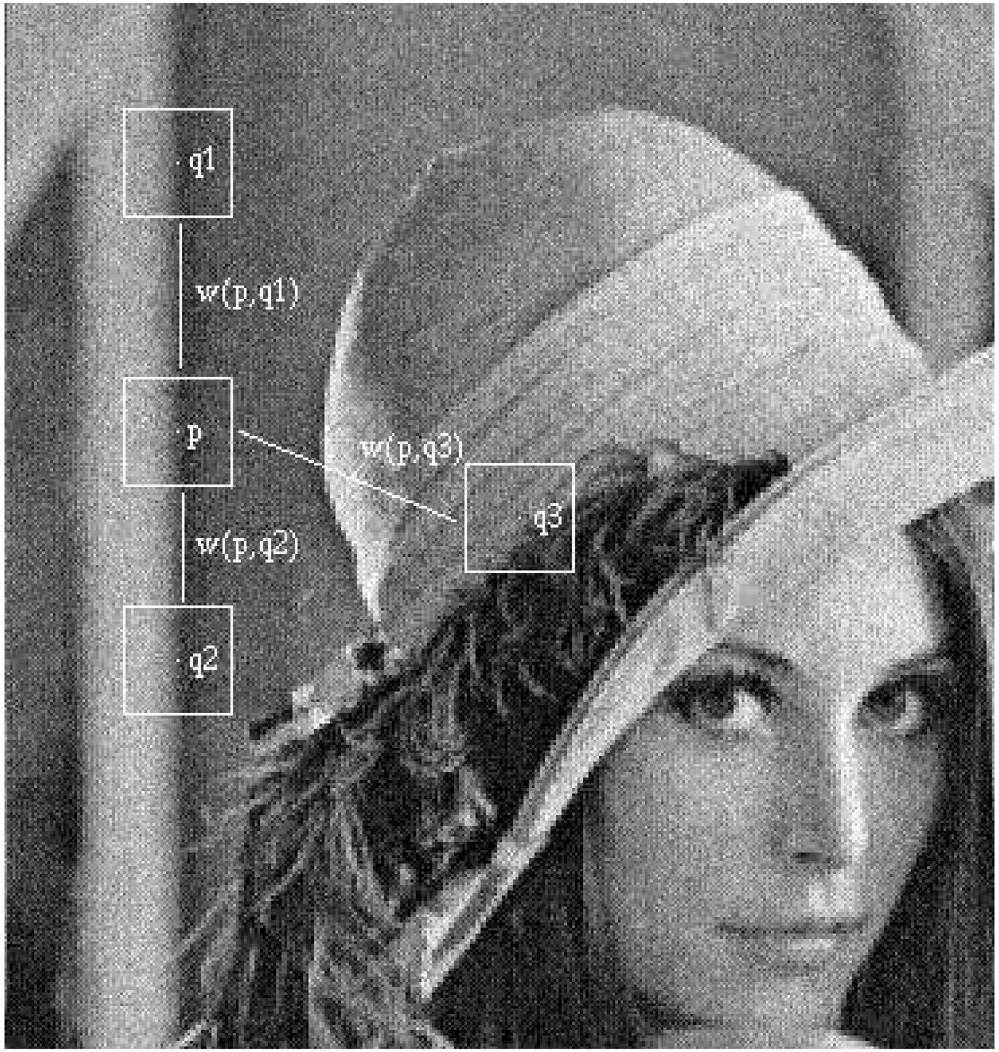


Figure 1: A. Buades, B. Coll, and J.M Morel, "Neighborhood filters and PDE's", Numerische Mathematik, 105 (1), 2006.


q1
w(p,q1)
p
w(p,q3)
q3
w(p,q2)
q2

FOUR NEIGHBORHOOD FILTERS FROM LOCAL TO NONLOCAL

- Gaussian mean : average of pixels in a whole Gaussian neighborhood.

$$G_\rho[u](\mathbf{x}) = \frac{1}{C(\mathbf{x})} \int_\Omega e^{-\frac{|\mathbf{x}-\mathbf{y}|^2}{\rho^2}} u(\mathbf{y}) d\mathbf{y}$$

where $C(\mathbf{x}) = C$ is the normalization parameter of the Gaussian parameter.

- Neighborhood filter. Average of pixels with a similar configuration in a whole Gaussian neighborhood.

$$NF_{\rho,h}[u](\mathbf{x}) = \frac{1}{C(\mathbf{x})} \int_\Omega e^{-\frac{|\mathbf{x}-\mathbf{y}|^2}{\rho^2}} e^{-\frac{|u(\mathbf{x})-u(\mathbf{y})|^2}{h^2}} u(\mathbf{y})\ d\mathbf{y};$$

- Anisotropic filter (mean curvature motion): Average of spatially close pixels in the direction of the level line

$$AF_h u(\mathbf{x}) = G_h * u_{|l(\vec{\xi})} = \int_{\mathbb{R}} G_h(t) u(\mathbf{x} + t\frac{Du^\perp}{|Du|}) dt,$$

- NL-means filter. Average of pixels with a similar configuration in a whole Gaussian neighborhood.

$$NL_{h,a}[u](\mathbf{x}) = \frac{1}{C(\mathbf{x})} \int_\Omega e^{-\frac{1}{h^2}\int_{\mathbb{R}^2} G_a(t)|u(\mathbf{x}+t)-u(\mathbf{y}+t)|^2 dt}\ u(\mathbf{y})\ d\mathbf{y},$$

where G_a is a Gaussian kernel of standard deviation a and h acts as a filtering parameter. Markovian hypothesis: Pixels with a similar neighborhood have a similar grey level value.

Diffusion kernel on graphs or discrete manifolds

- A. Buades, A. Chien, J.M Morel, and S. Osher : in preparation
- Leo Grady, Random Walks for Image Segmentation, IEEE Trans. on Pattern Analysis and Machine Intelligence, 28(11), 2006.
- S. Kindermann, S. Osher, P.W. Jones - Deblurring and denoising of images by nonlocal functionals, Multiscale Modeling and Simulation 2005
- G. Gilboa, J. Darbon, S. Osher, and T. Chan, Nonlocal Convex Functionals for Image Regularization 2006.
- A. Szlam, M. Maggioni, and R. R. Coifman, A general framework for adaptive regularization based on diffusion processes on graphs, Yale technichal report, (2006).
- D. Comaniciu, and P. Meer, Mean shift: a robust approach toward feature space analysis, Pattern Analysis and Machine Intelligence, 24(5), 2002.

Diffusion kernel on graphs or discrete manifolds

- $x \to p(x) \in \mathcal{M}$ associates to a pixel in the image the "patch" around it.
- Define a symmetric isotropic kernel on the discrete "patch" manifold $\mathcal{M}$

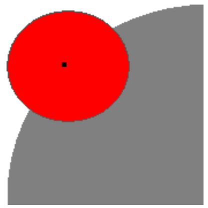
$$\forall p, q \in \mathcal{M}, W_0(p,q) = e^{-\frac{||p-q||^2}{2\sigma^2}}$$

- This operator defined on pixels thanks to the identification given by the mapping p : $\forall x, y, W_0(x,y) = W_0(p(x), p(y))$.
- Normalized filtering kernel: $W(x,y) = W(p(x), q(y)) = \frac{1}{D(p)} W_0(p,q)$; $D(p) = \sum_q W(p,q)$.
- The kernel W defines an operator on images g by $Wg(x) = \sum_y W(x,y)g(y)$. This is a low-pass filter because $W1 = 1$ and W is order preserving.
- High pass Laplace operator L and its symmetrized version L_0 defined by $L =: Id - W$ and $L_0 = D^{\frac{1}{2}} L D^{-\frac{1}{2}} = Id - D^{-\frac{1}{2}} W_0 D^{-\frac{1}{2}}$, (2.4) where $D = diag_x(D(x))$.
- L_0 is a discrete graph Laplacian. The iteration of symmetric NL-means is interpreted as a heat equation on the patch manifold (Gilboa, Osher) or as a Donoho-Johnstone threshold on the eigenvalues of this Laplacian (Coifman, Szlam, Peyré)

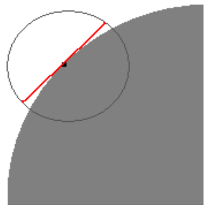
NONLOCAL HEAT KERNELS and EIGENVECTORS OF LAPLACIAN

•Diffusion neighborhoods

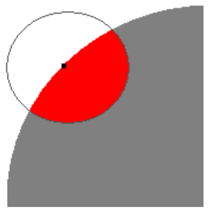
Image



Gaussian



Anisotropic filt.

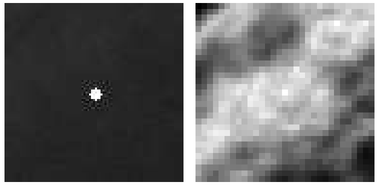


Neighborhood filter

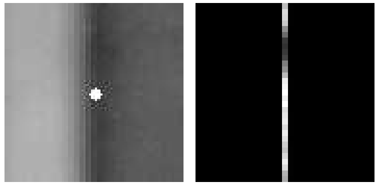
NLmeans

The NL-means: An extension of all previous methods or Gestalt Grouping

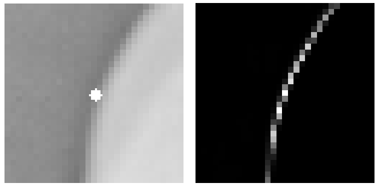
- Flat region. The large coefficients are spread out like a convolution.



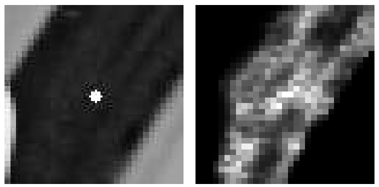
- Straight edge. The large coefficients are aligned like in a anisotropic filter.



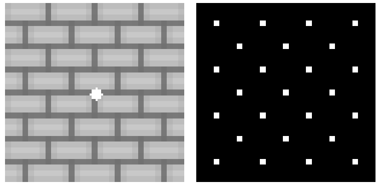
- Curved edge. The weights favor pixels belonging to the same contour.



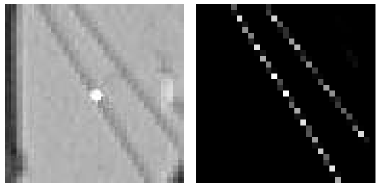
- Flat neighborhood. The average is made in the grey level neighborhood as the neighborhood filter.



- Periodic case. The large coefficients are distributed across the texture (non local).



- Repetitive structures. The weights favor similar configurations even they are far away (non local).



6 G. PEYRÉ

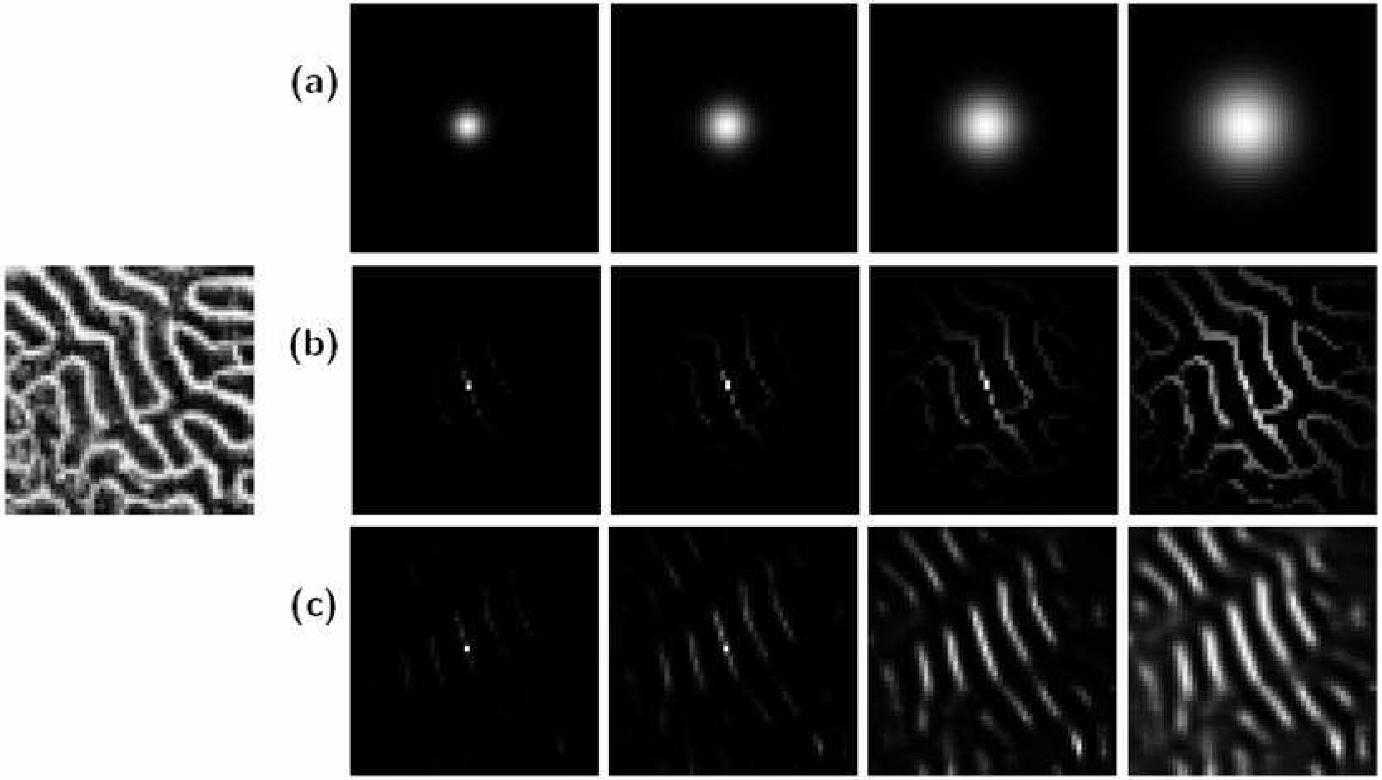


FIGURE 2.1. *Left: original image f. Right: heat diffusions with an increasing time for: (a) local embedding $x \mapsto x$, (b) semi-local embedding $x \mapsto (x, \lambda f(x))$, (c) non-local embedding $x \mapsto p_x(f)$.*

Figure 2: G. Peyré, Manifold models for signals and images, preprint, (2007).

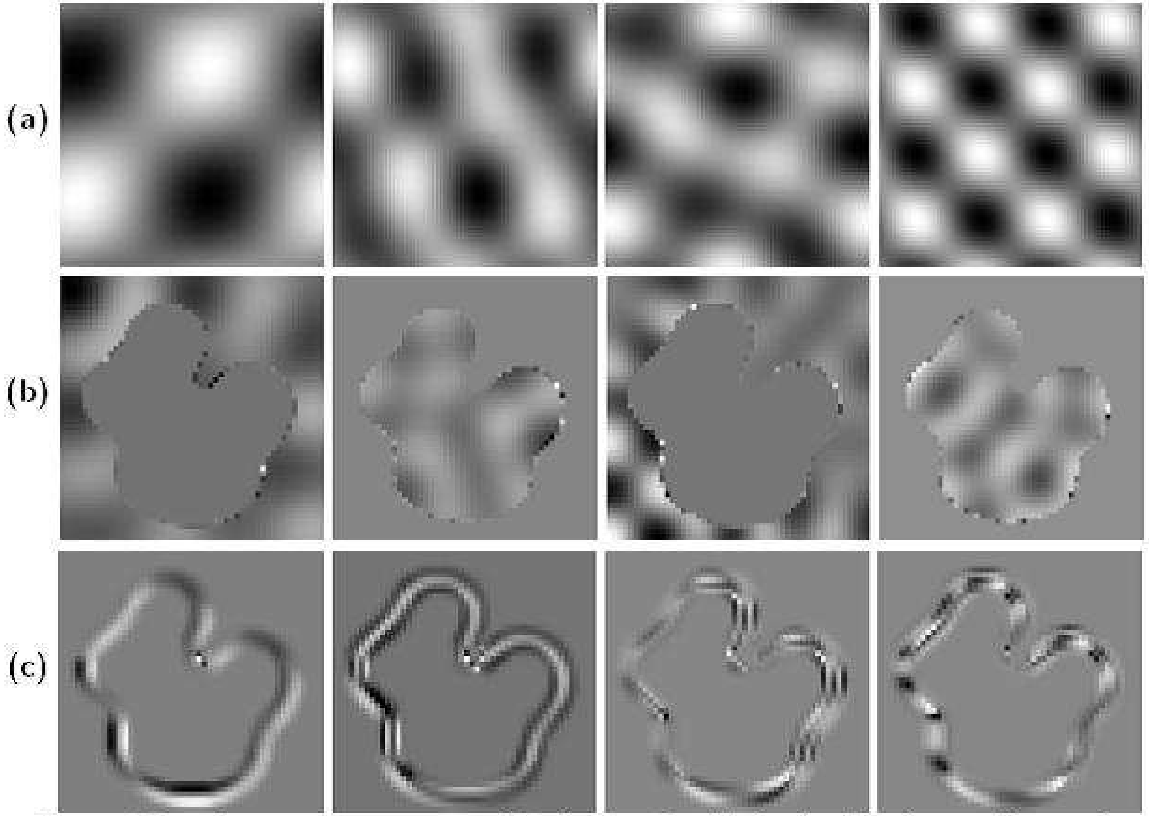
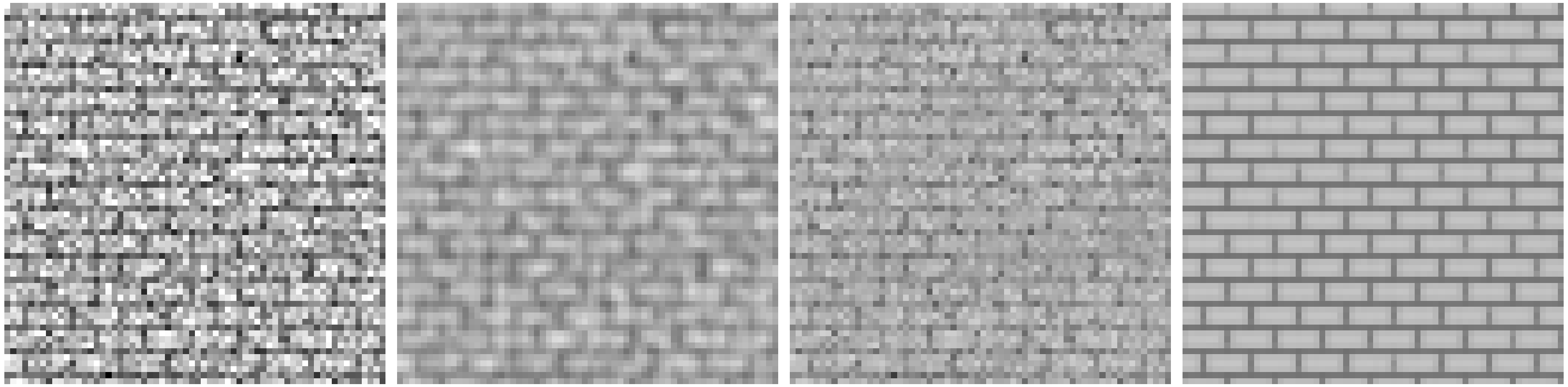


FIGURE 2.2. *Some eigenvectors u_ω of Laplacians for (a) the local Laplacian (Fourier basis), b) the semi-local Laplacian, (c) the non-local Laplacian.*

Figure 3: G. Peyré, Manifold models for signals and images, preprint, (2007).

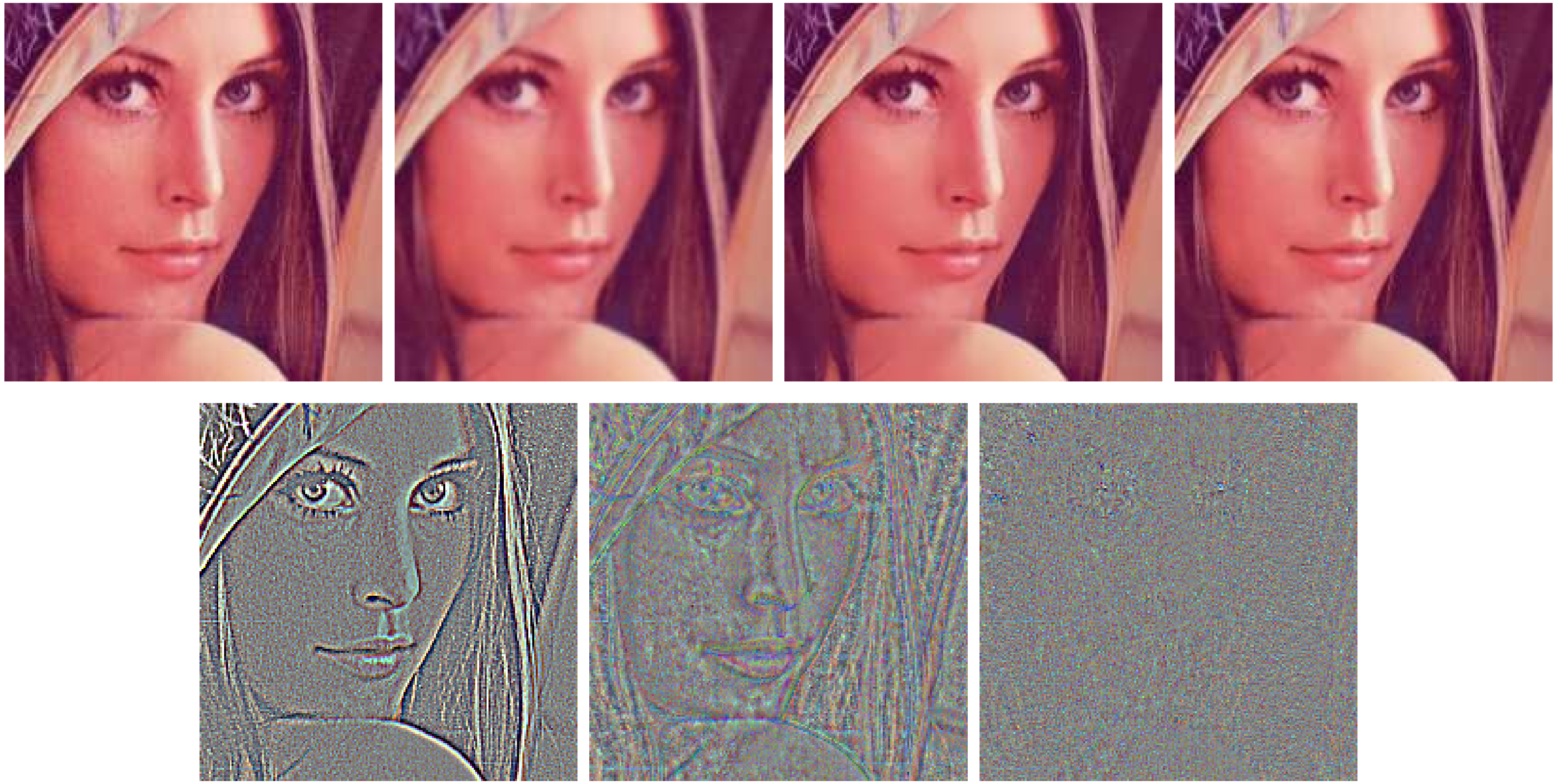
Comparison: Visual quality: Input, Gauss, Neighborhood classic, Non local means

- Comparison on a periodic image

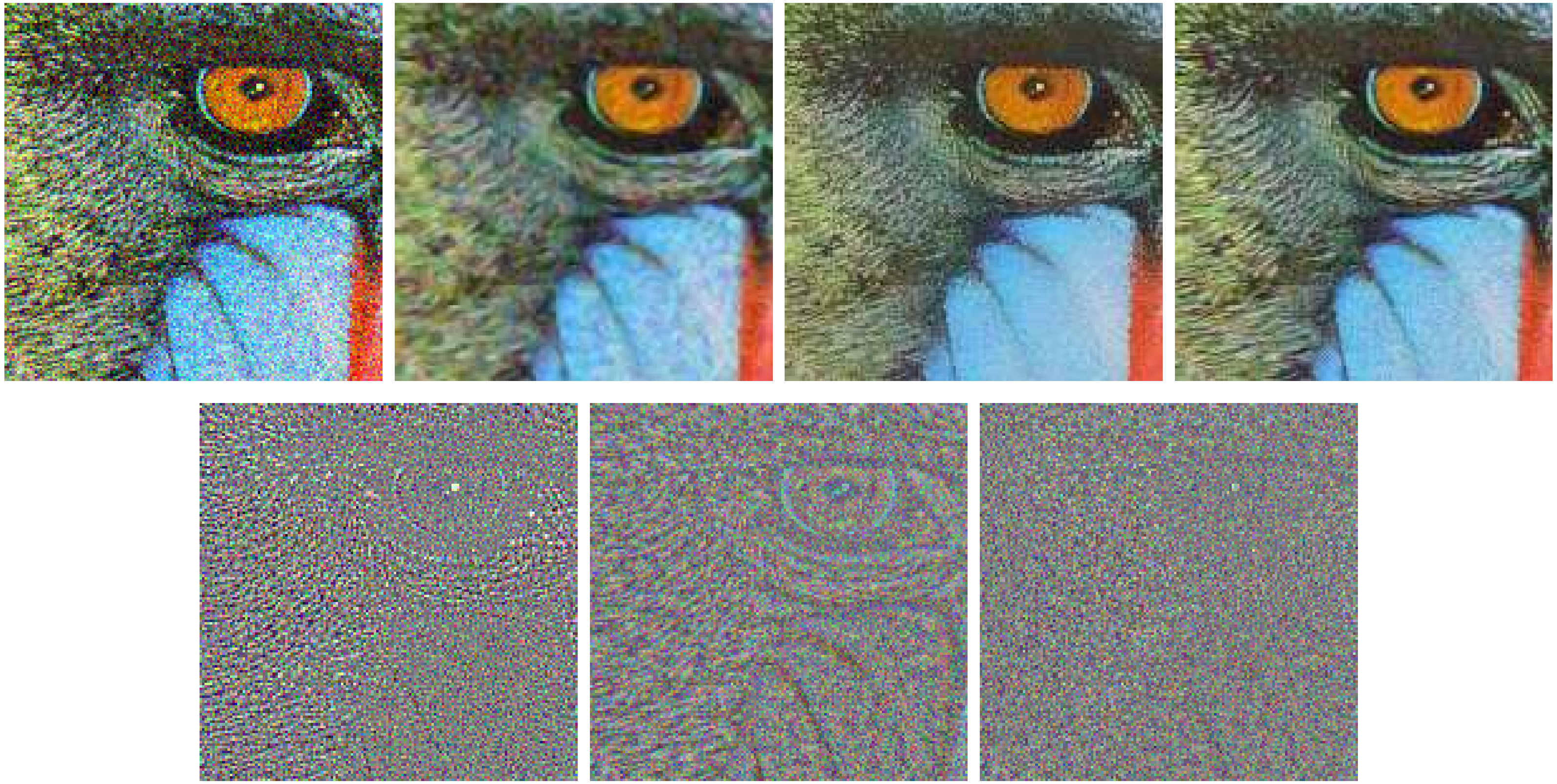


Visual Comparison

- Restored images and removed noise by Gaussian convolution, sigma filter and NL-means.



- Restored images and removed noise by the anisotropic filter, the sigma filter and the NL-means.



TWO MAIN DENOISING PRINCIPLES

Denoising

We define a denoising method D_h as a decomposition

$$v = D_h v + n(D_h, v), \tag{1}$$

where h is a filtering parameter which usually depends on the standard deviation of the noise σ.

- **Preservation of original information.** Features in $n(D_h, v) = v - D_h v$ are removed from v. We call this difference *method noise* when v is non or slightly noisy.

  Principle 1 *For every denoising algorithm, the method noise must be zero if the image contains no noise and should be in general an image of independent zero-mean random variables.*

- **No artifacts** The transformation of a white noise into any correlated signal creates structure and artifacts.

  Principle 2 *A denoising algorithm must transform a white noise image into a white noise image (with lower variance).*

Other classic algorithms to be compared

- Minimization of the total variation

$$TVF_\lambda(v) = \arg\min_u \int_\Omega |Du| + \lambda \int |v-u|^2 \tag{1}$$

- Wavelet thresholding

$$HWT = \sum_{\{(j,k) \,|\, |\langle v, \psi_{j,k}\rangle| > \tau\}} \langle v, \psi_{j,k}\rangle \psi_{j,k}$$

where $\mathcal{B} = \{\psi_{j,k}\}_{(j,k)}$ is a wavelet basis and τ the threshold.

Preservation of original information. We recall that we defined the method noise as the image difference $n(D_h, v) = v - D_h v$ when v is non or slightly noisy.

Principle 1 *For every denoising algorithm, the method noise must be zero if the image contains no noise and should be in general an image of independent zero-mean random variables.*

Theorem 1 *The convolution with a gaussian kernel G_h is such that*

$$u - G_h * u = -h^2 \Delta u + o(h^2),$$

for h small enough.

Theorem 2 *The image method noise of an anisotropic filter AF_h is*

$$u(\boldsymbol{x}) - AF_h u(\boldsymbol{x}) = -\frac{1}{2} h^2 |Du| curv(u)(\boldsymbol{x}) + o(h^2),$$

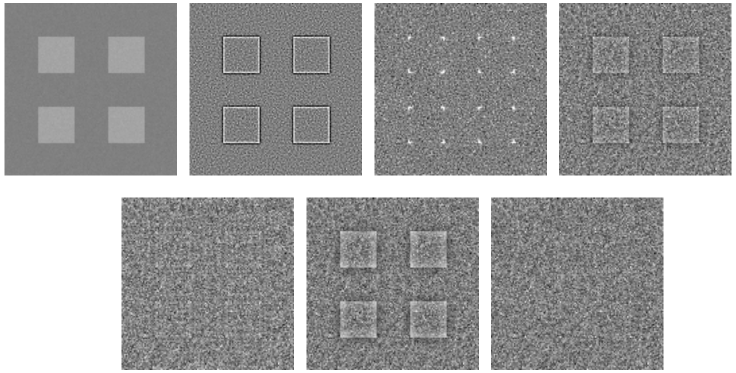
Theorem 3 *The method noise of the Total Variation minimization is*

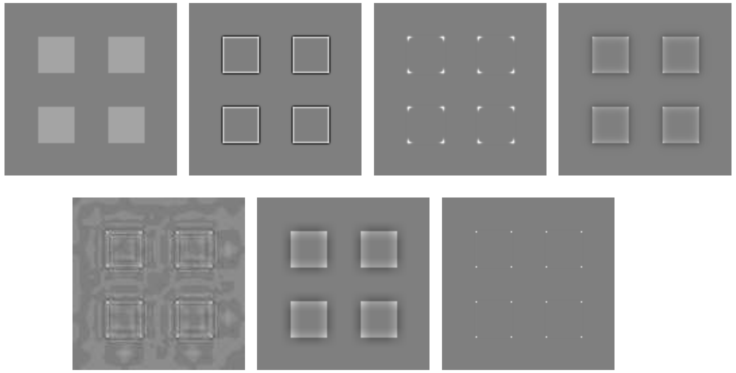
$$u(\boldsymbol{x}) - TVF_\lambda(u)(\boldsymbol{x}) = -\frac{1}{2\lambda} curv(TVF_\lambda(u))(\boldsymbol{x}).$$

Theorem 4 *The method noise of a hard thresholding $HWT_\mu(u)$ is*

$$u - HWT_\mu(u) = \sum_{\{(j,k) || \langle u, \psi_{j,k} \rangle | < \tau\}} \langle u, \psi_{j,k} \rangle \psi_{j,k}$$

Method Noise Method noise of six denoising methods





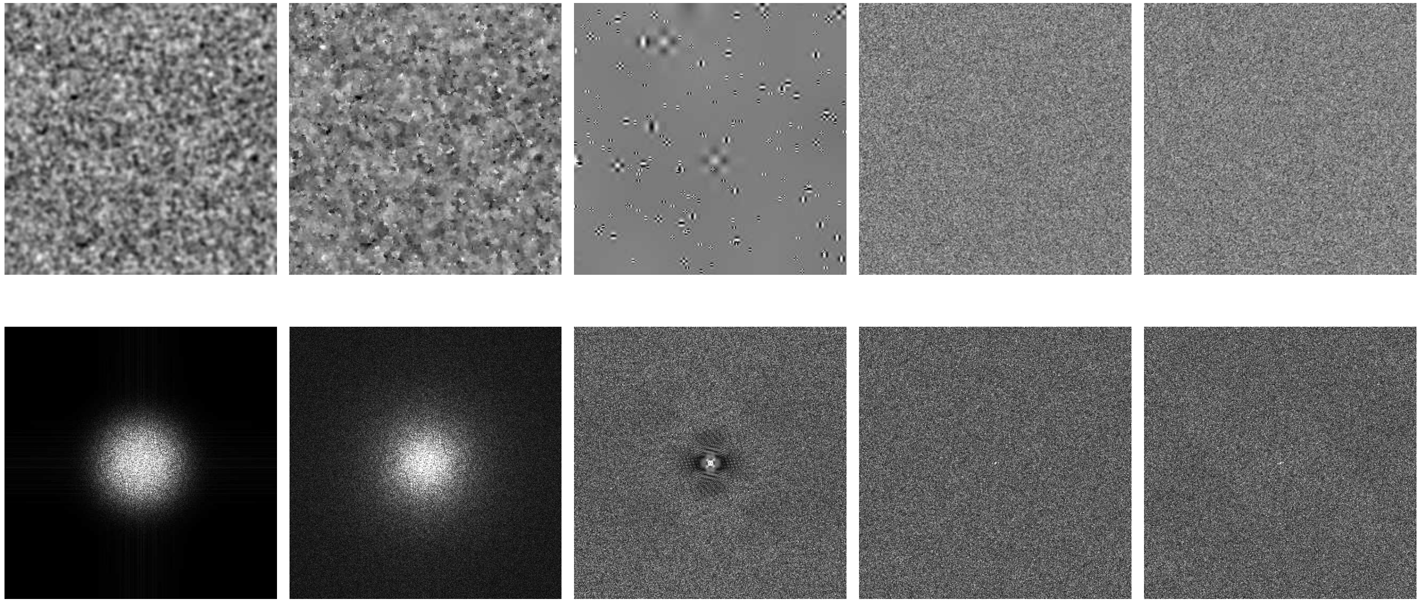
Gaussian, Anisotropic filtering, TV, stationary wavelet, neighborhood filter, NL-means.

In the noisy case, parameters are fixed in order to remove exactly an energy σ^2 ($\sigma = 2.5$). The same parameters have been used in the second experiment on the real image.

No artifacts The transformation of a white noise into any correlated signal creates structure and artifacts.

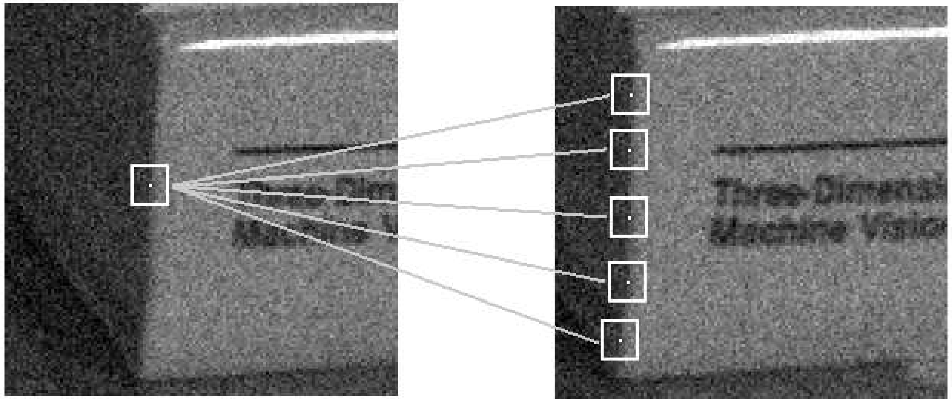
> **Principle 2** *A denoising algorithm must transform a white noise image into a white noise image (with lower variance).*

Application of the denoising algorithms to a noise sample and its Fourier transforms. Gauss, TV, wavelets, neighborhood filter, NL-means



Films and NL-means

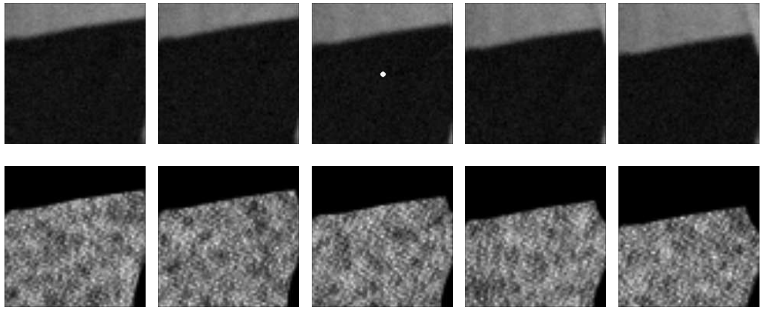
- The NL-means algorithm does not need to calculate the trajectories. It simply looks for the resembling pixels, no matter where they lie in the movie.

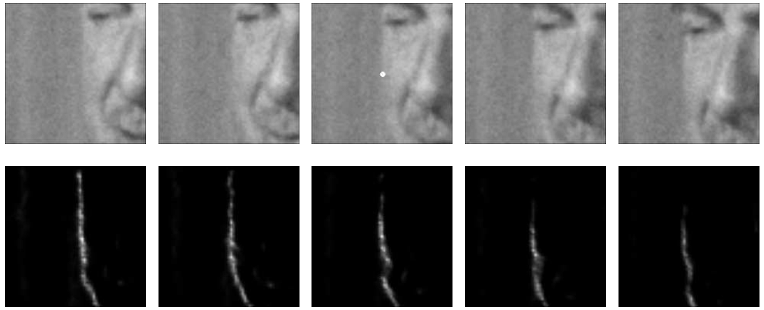


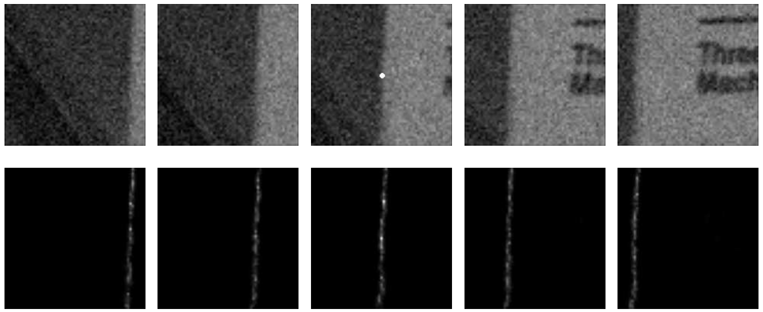
Why do not average all?

- Straightforward extension as a spatiotemporal filter.

Probability distributions in movement



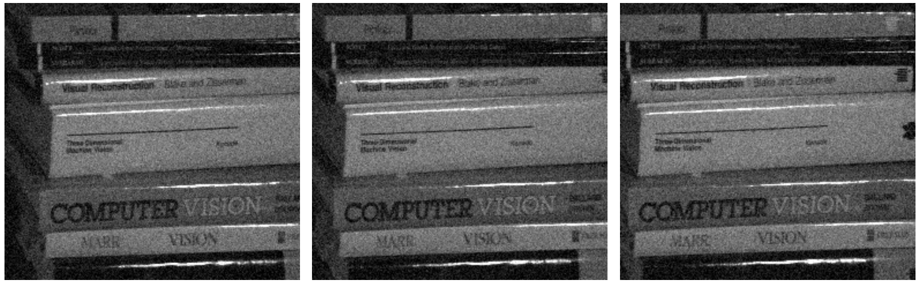




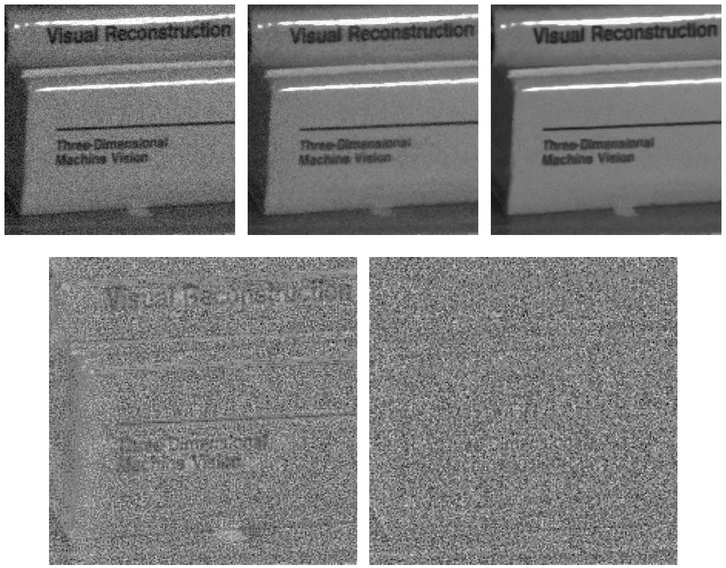
The algorithm looks for the pixels with a more similar configuration even they have moved. This algorithm is adapted to moving pictures without the need of an explicit motion estimation.

Comparison

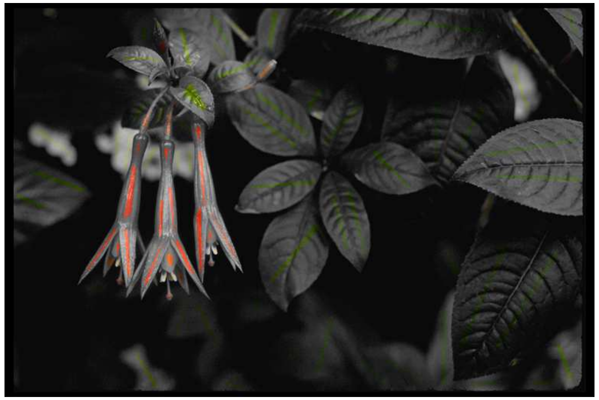
- Three consecutive frames of a noisy image sequence. The noisy sequence has been obtained by the addition of a Gaussian additive white noise ($\sigma = 15$) to the original sequence.



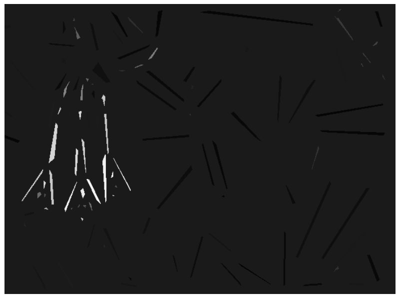
- Comparison experiment between the motion compensated neighborhood filter and the NL-means. The motion estimation has been obtained by the block matching algorithm.

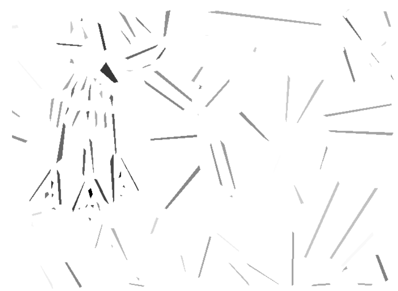


COLORIZATION AND SEGMENTATION BY DIFFUSION



which is decomposed as in Y, U, V as (seeds in color).



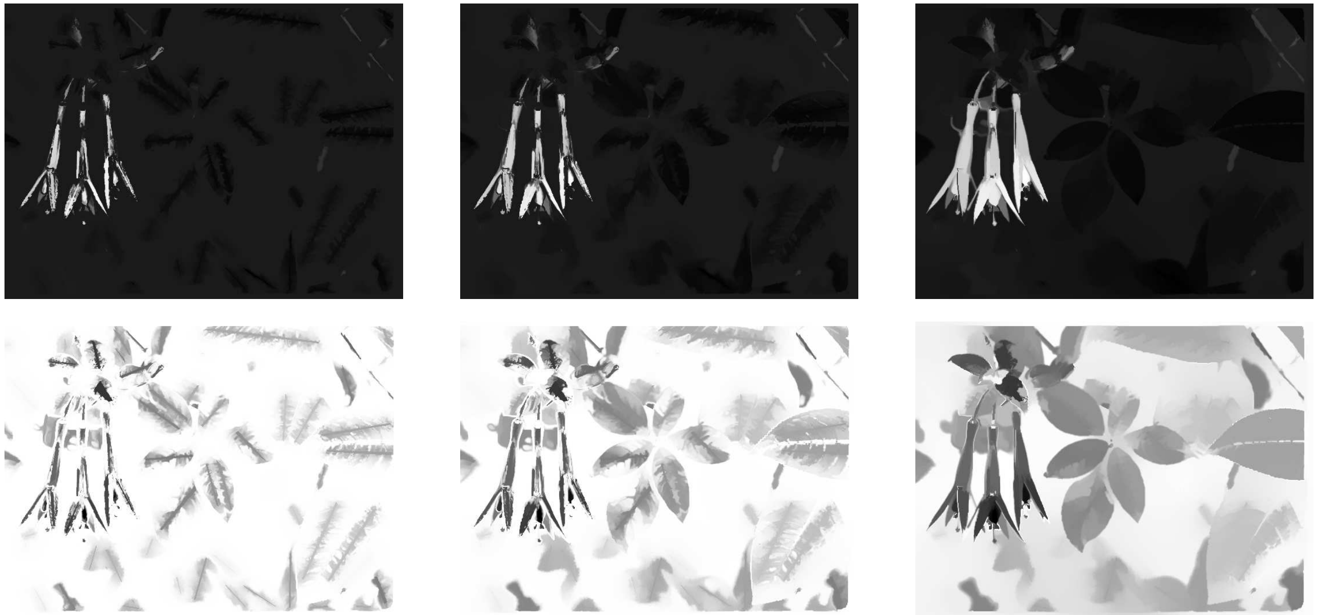


- Non local heat equation with neighborhood fixed by the grey level image and color seeds Dirichlet condition.
- Replace each pixel (x, y) with no initial color by the weighted mean

$$\hat{u}(x, y) = \frac{1}{|B|} \sum_{(v,w) \in B} u(v, w).$$

where B is the set of pixels with grey level similar to (x, y).

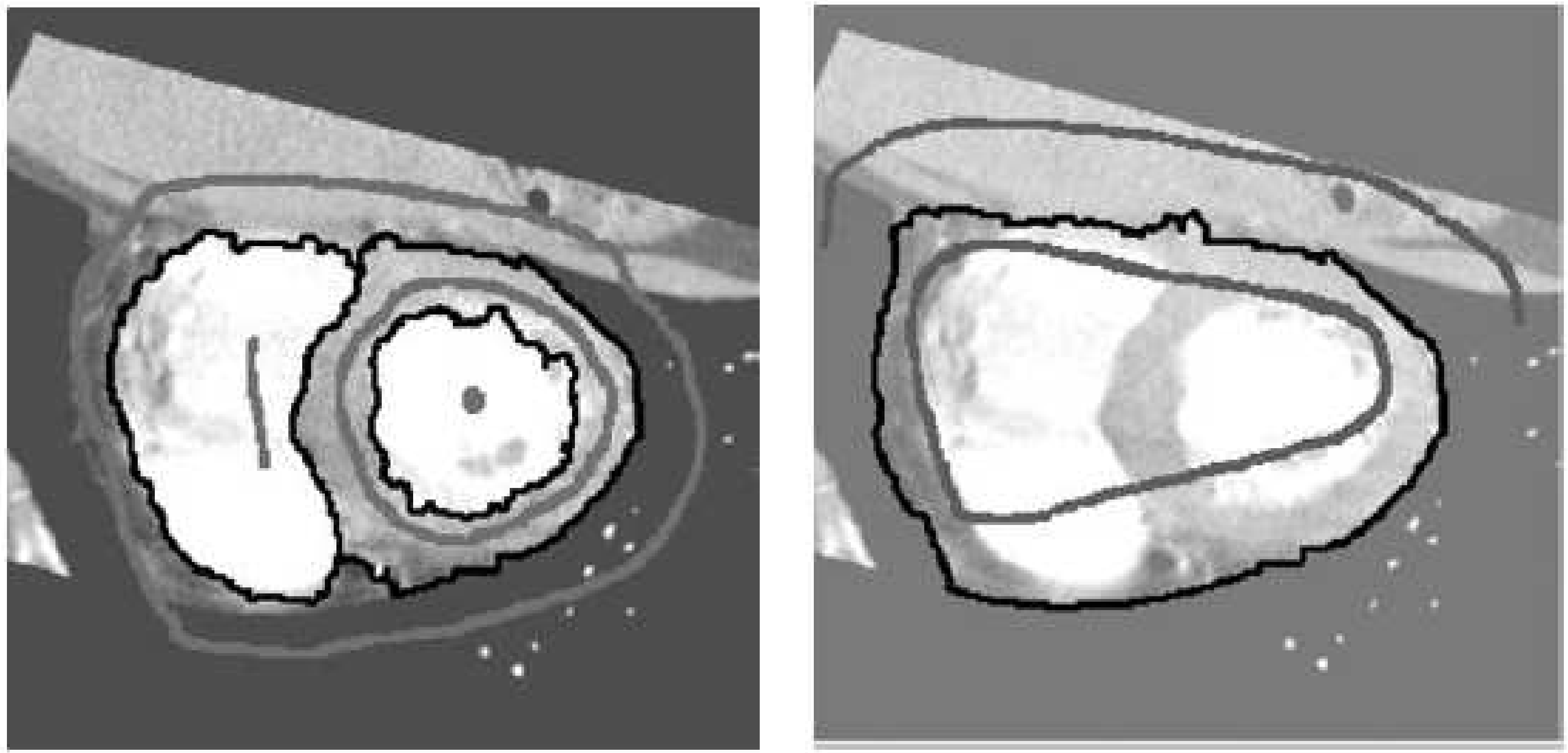
- Iterate



The single point comparison of the neighborhood filter can lead to the mixing of objects color. Middle: neighborhhod filter (Grady), right : NL-means.

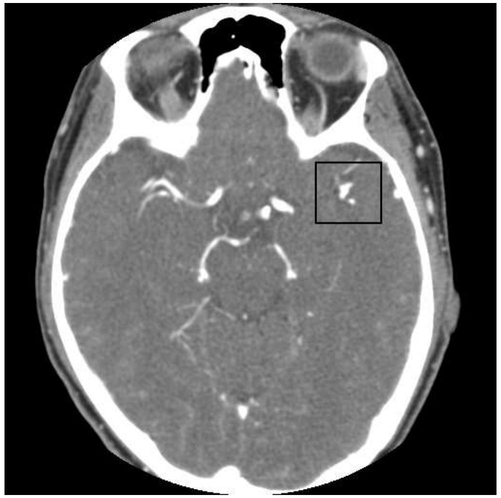
MEDICAL IMAGING

Grady and Funka-Lea multi-label segmentation for medical images



Histogram concentration and enhancing 3D.

- Algorithm easily generalized to 3D.
- Application to a CT 3D image of the head where blood vessels should be segmented.



We diffuse the 3D image conditionally to the weights given by the initial data.

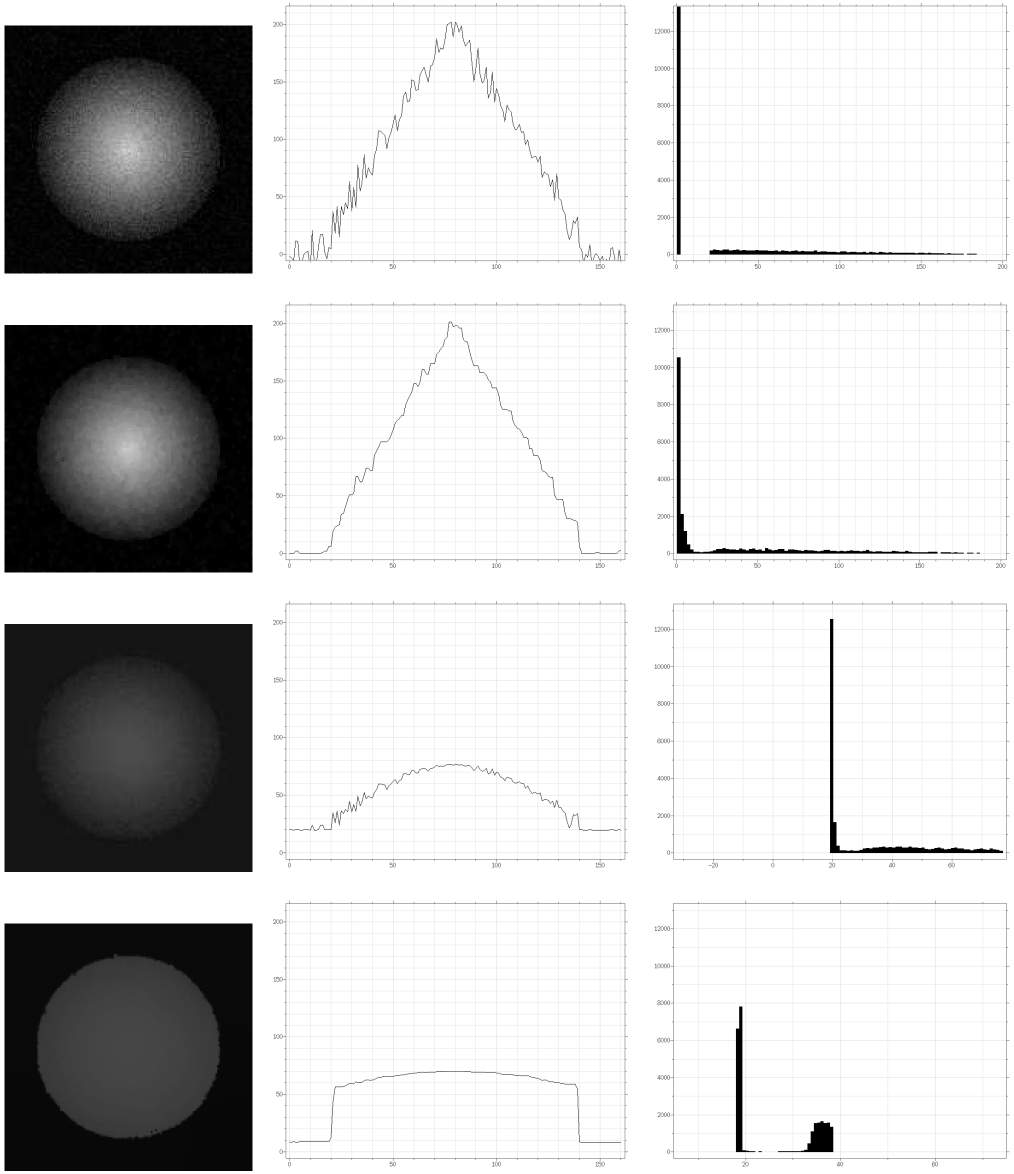


Figure 4: Experience on a 2D image. From top to bottom: original image, iterative application of a Gaussian mean, iterative application of a median filter, proposed method with neighborhood filter weights and proposed method with NL-means weights. On the middle central line of each image. On the right, histogram of the filtered image

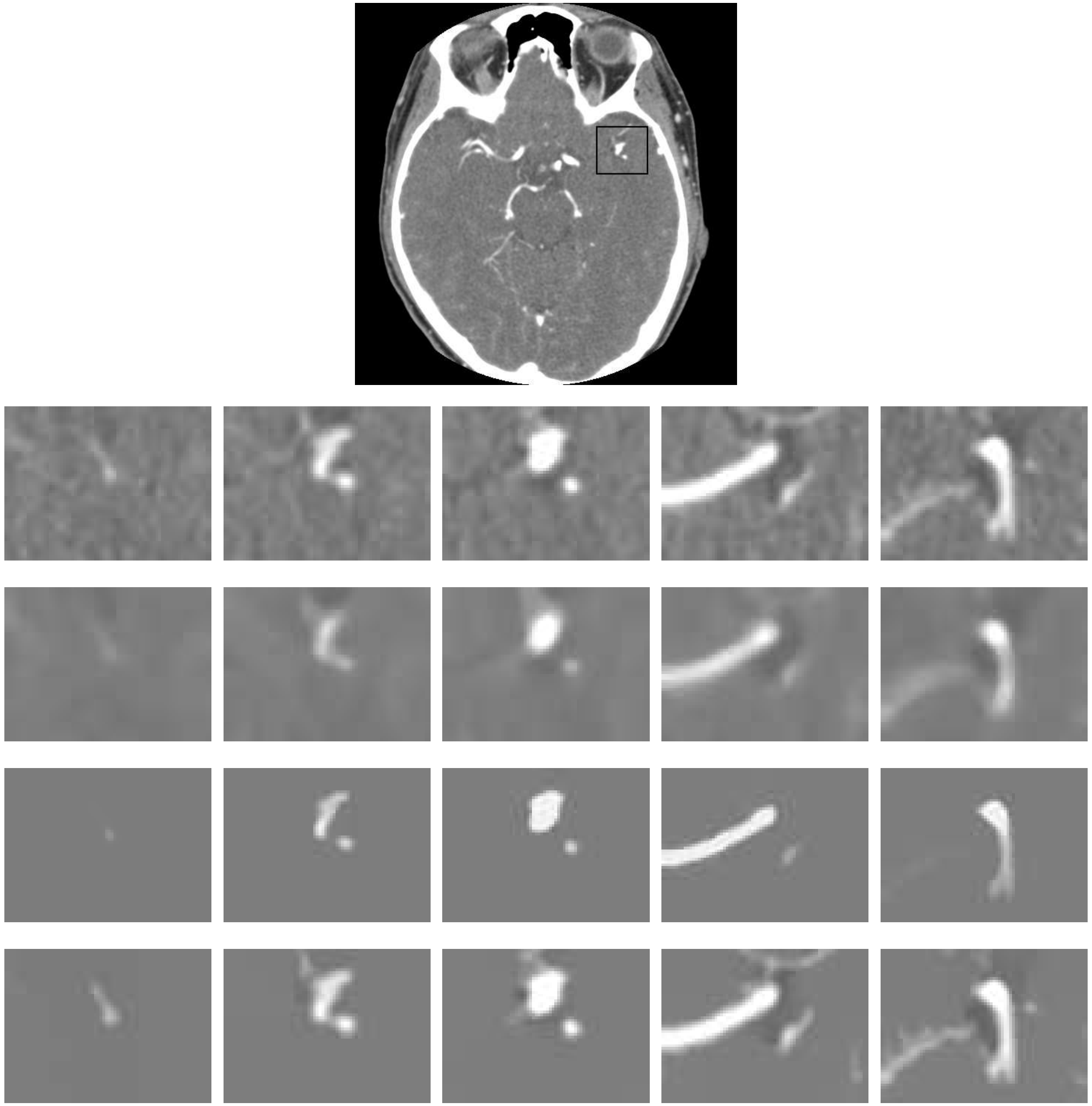


Figure 5: Application to a CT 3D image of the head where blood vessels should be segmented. Top: One single image of the CT image with marked interested area. Middle: Display of interested area for several slices of the 3D image. Second row: filtered slices by using median filter. Third row: neighborhood filter. Fourth row: 3D nonlocal heat equation.

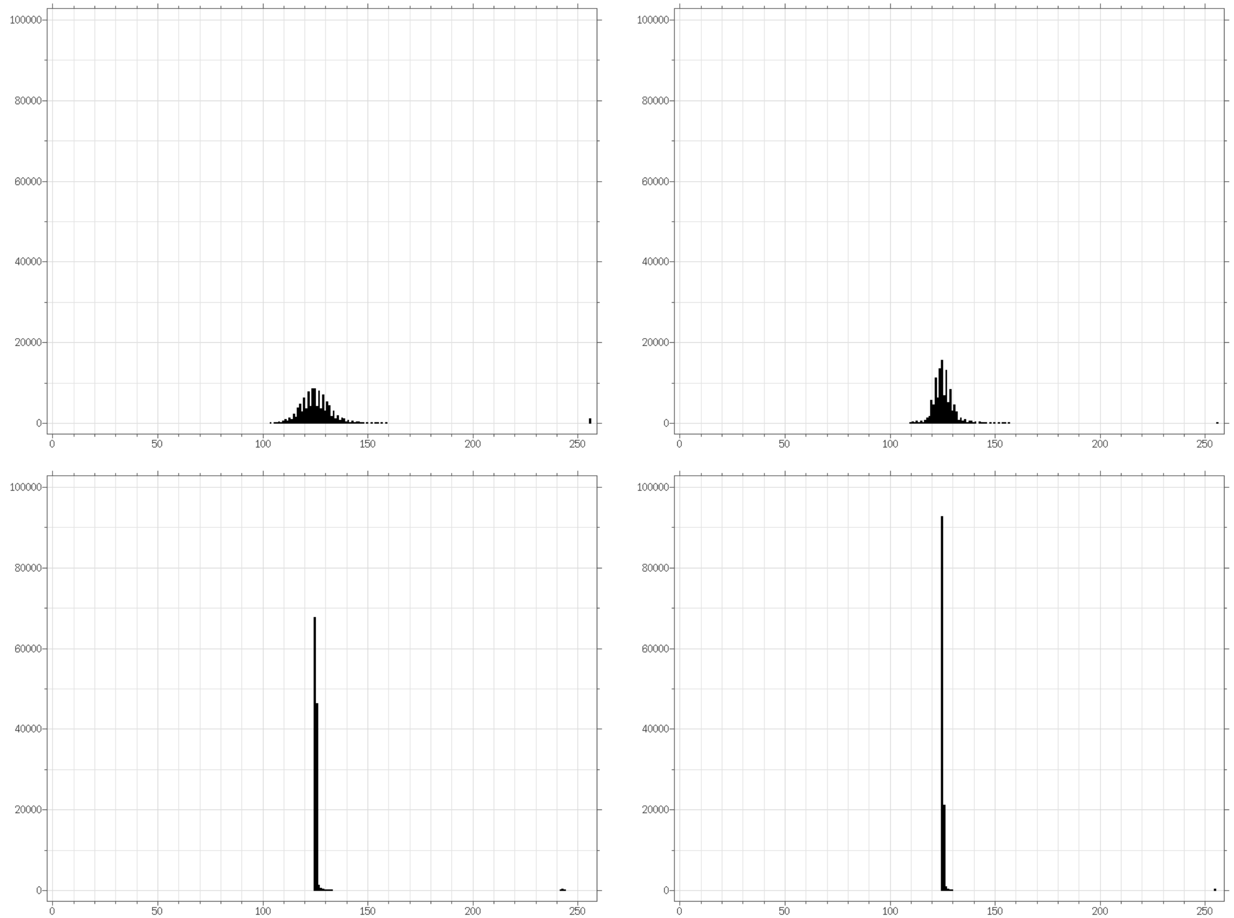


Figure 6: Grey level histogram of interest 3D areas. Top left: original 3D image before. Top right: after median filtering. Bottom left: after proposed method with neighborhood filter weights. Bootm right; proposed method with NL-means weights The background is now represented by a few grey level values when the volume is filtered by the proposed method and therefore a threshold can be more easily and automatically applied.

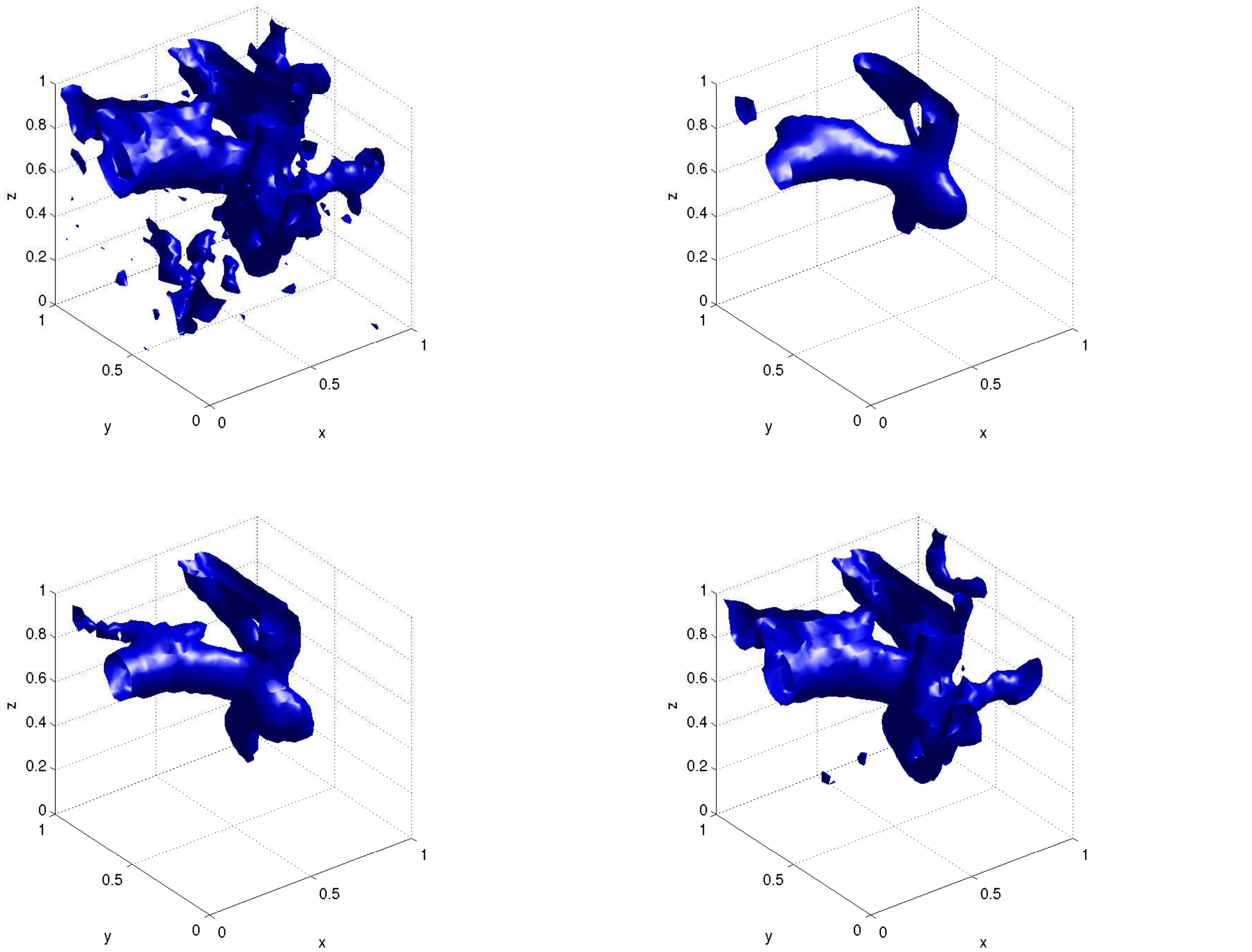


Figure 7: From top to bottom and left to right: original volume, filtered by iterative median filter, neighborhood filter and NL-means weights. The isosurface extracted from the original image presents many irregularities due to noise, while the median filter makes important parts disappear and vessels disconnect. NL-means weight keeps more vessels.

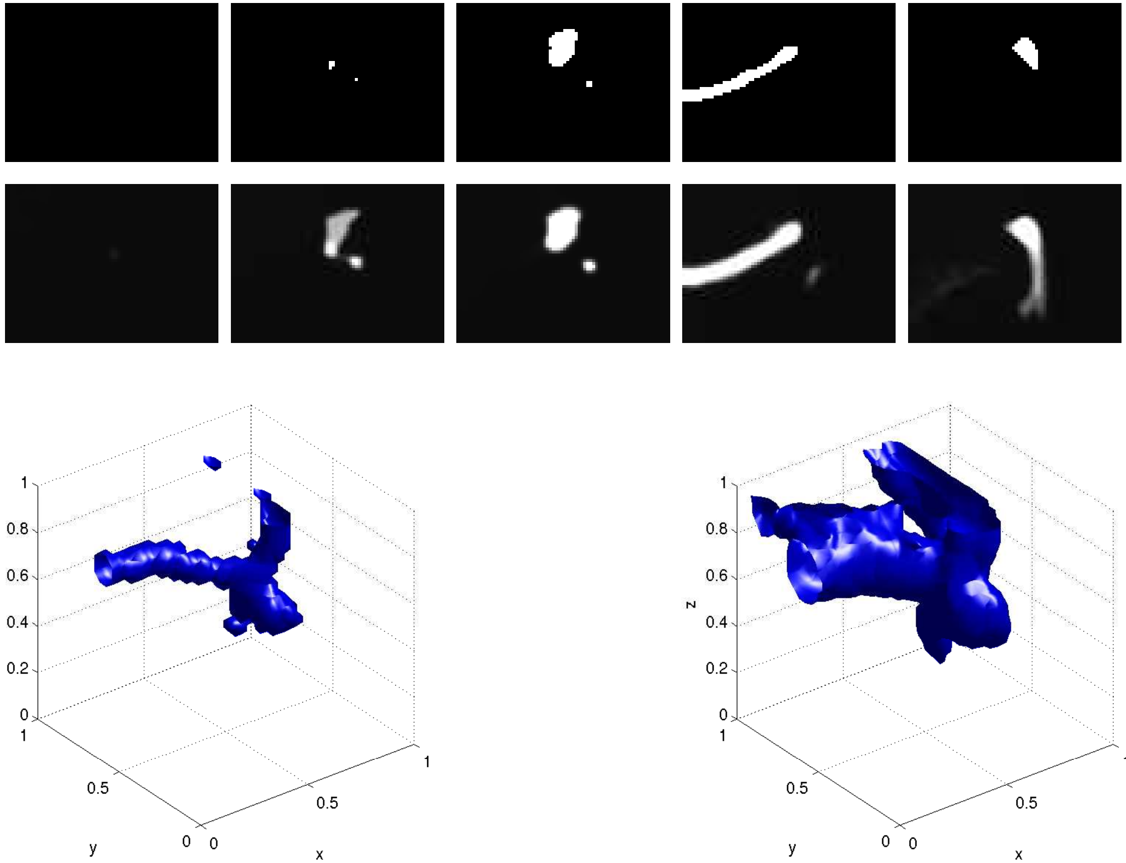


Figure 8: Result of the Grady seed neighborhood diffusion

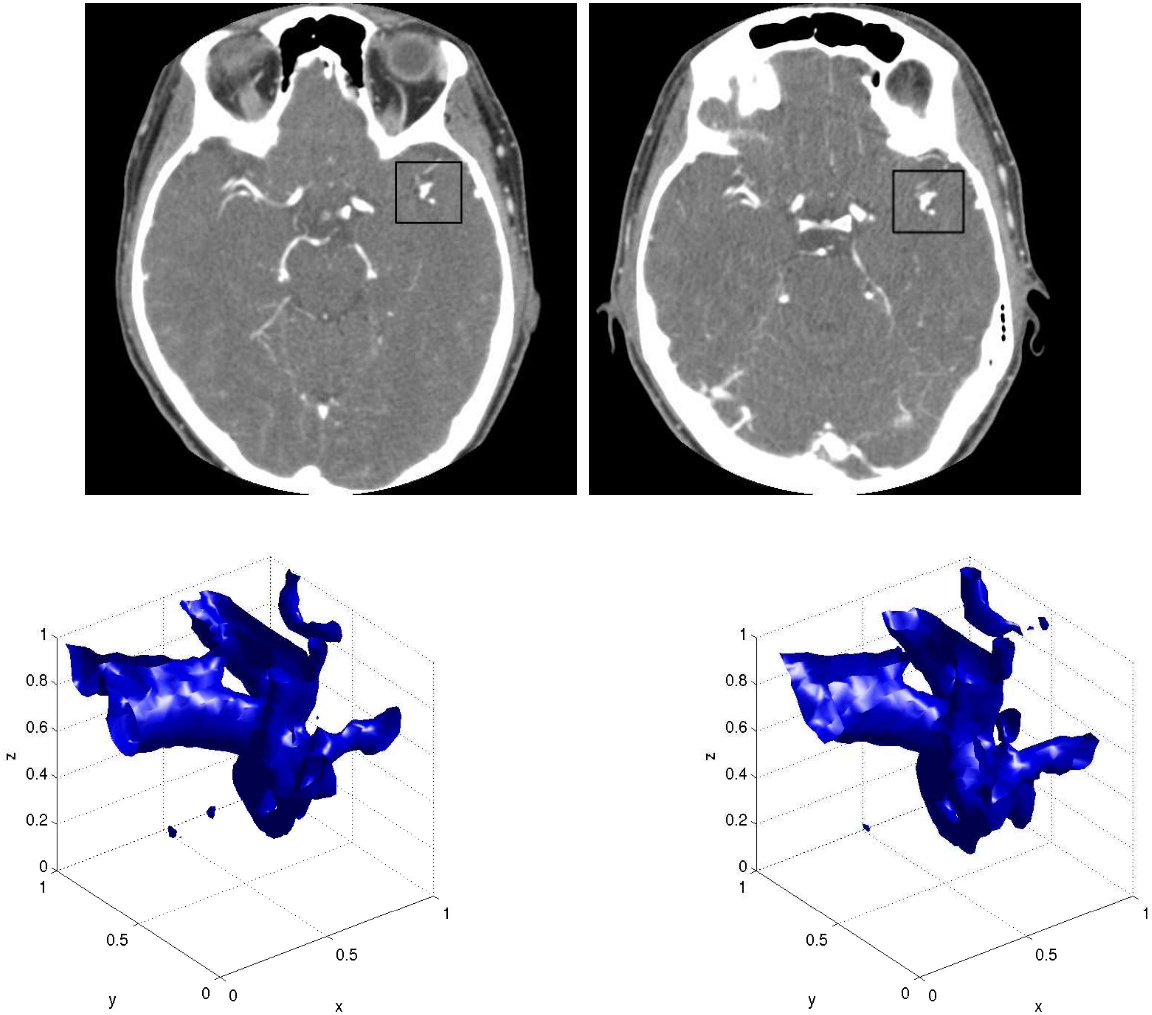
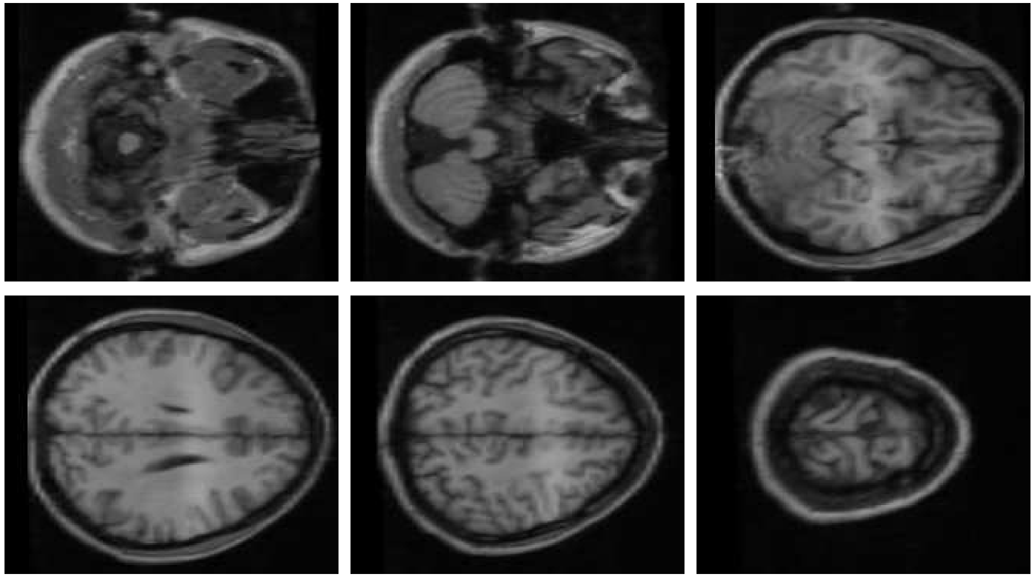


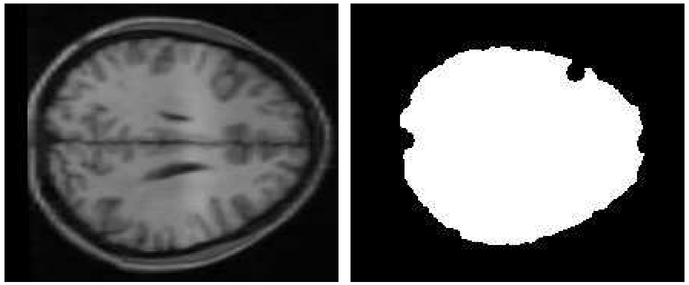
Figure 9: Same experiment with automatic threshold at the first local minimum after the background pick, same patient, one year interval.

3D Image segmentation by nonlocal heat equation

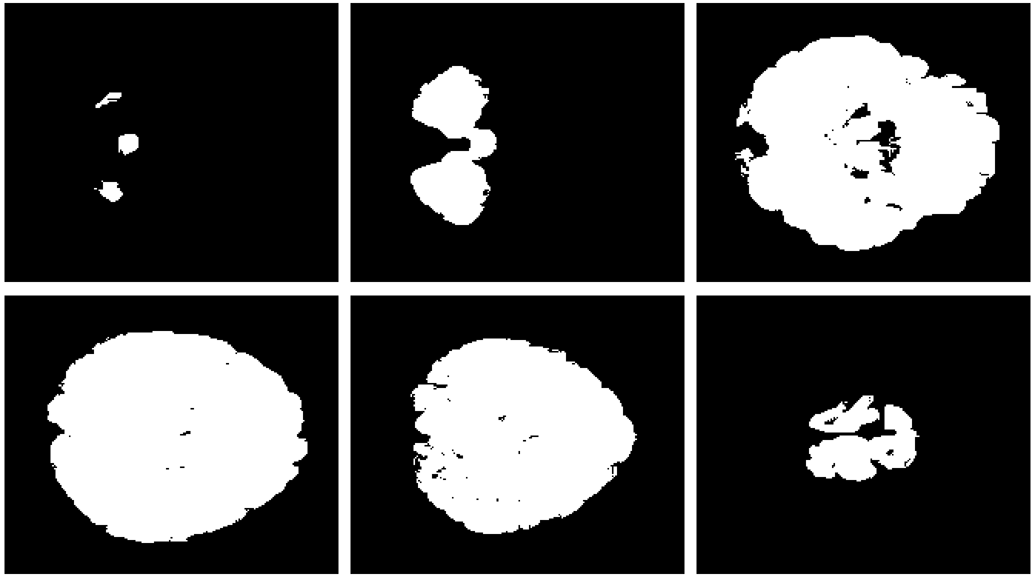
- Six of the 180 slices of a brain IRM.



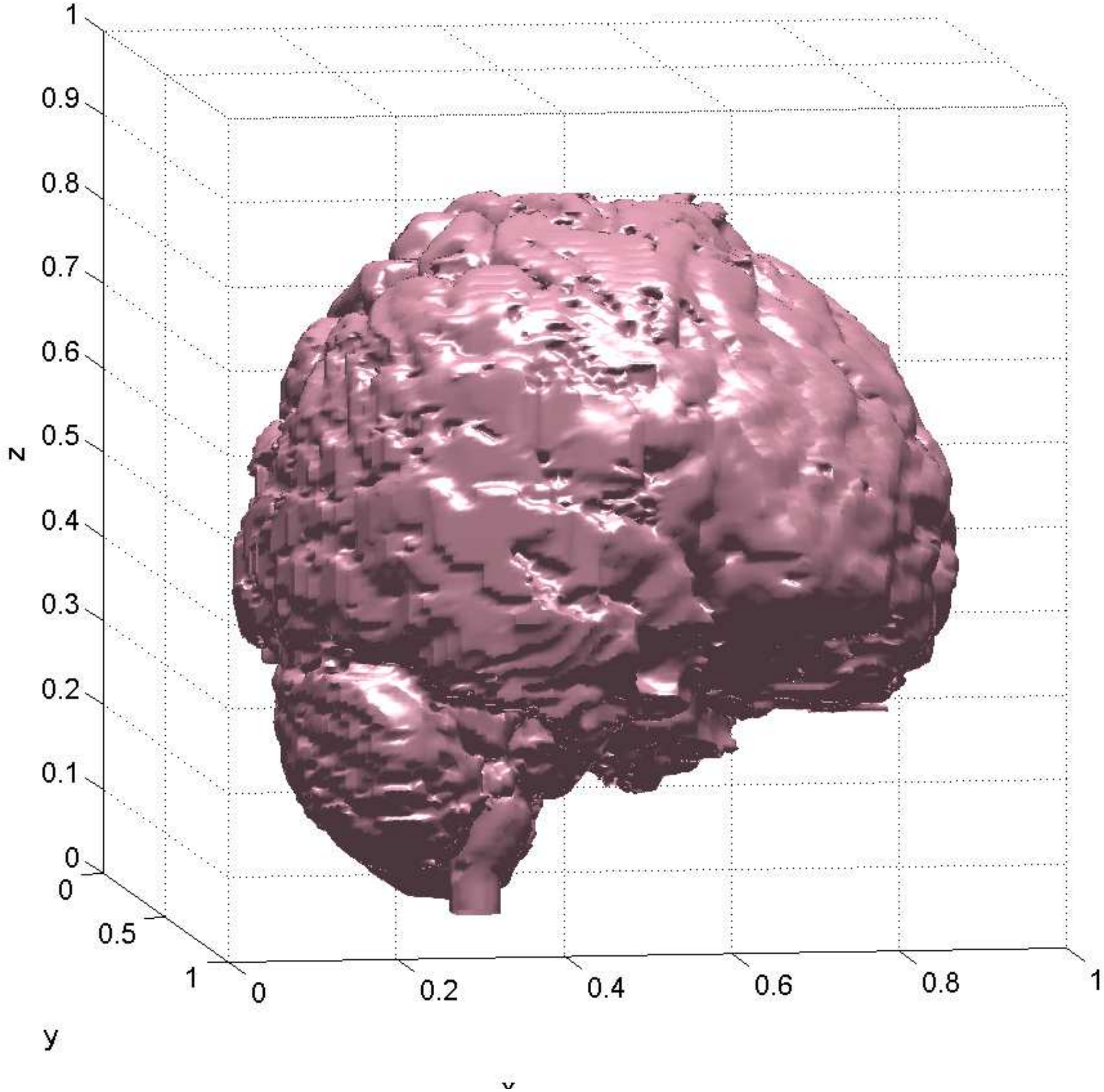
- Manual segmentation of a single slice (0 black, 1 white).



- Colorization algorithm
- Propagation to the whole 3D image from a single slice.



Superficie 3D extreta de la segmentació del cervell.



DEMOSAICING

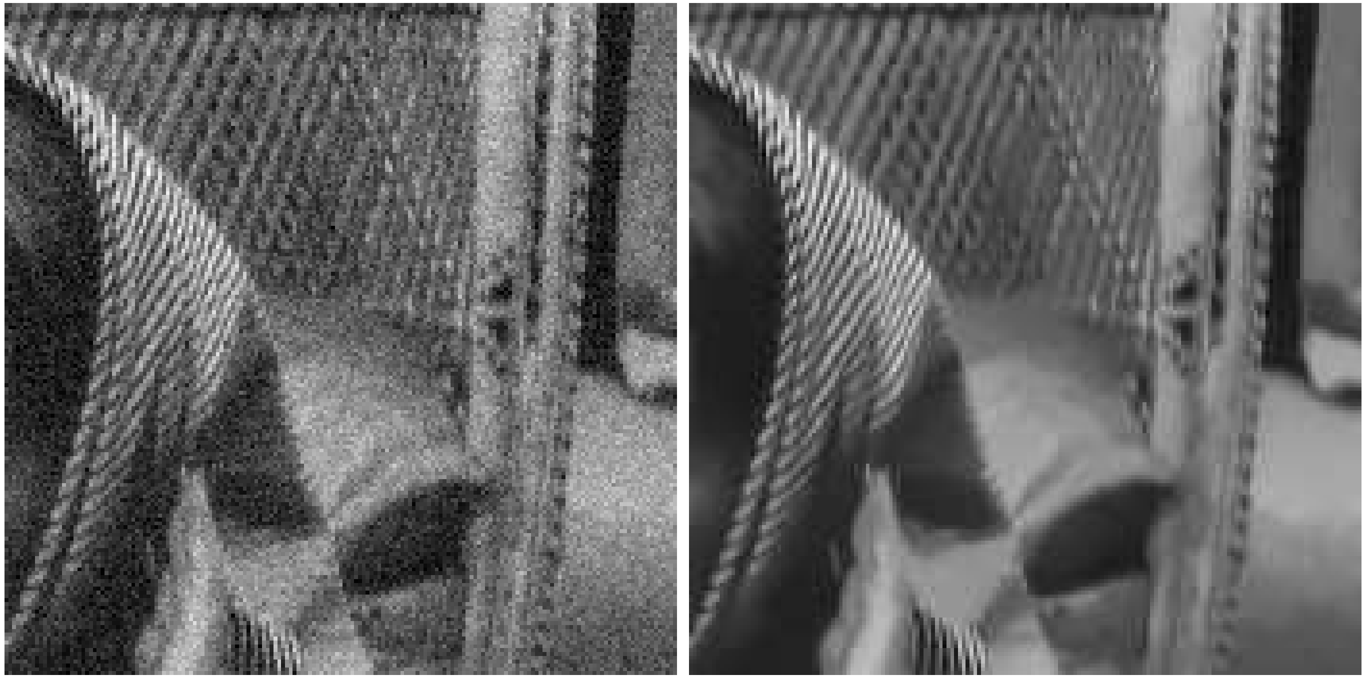
PDEs II: AVOIDING SHOCKS BY LINEAR REGRESSION

Drawbacks

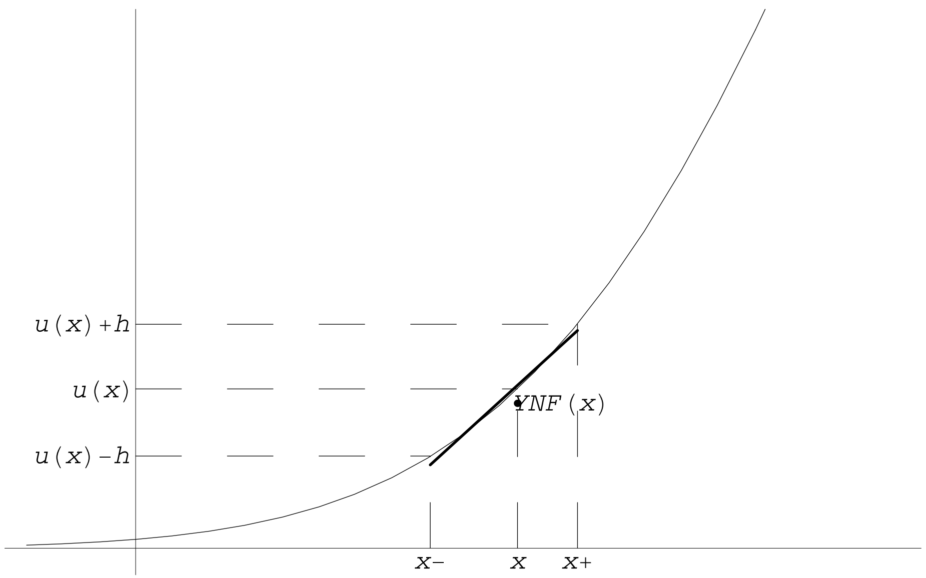
- The neighborhood filter and the NL-means share a **shock effect**.

In general, image filters can be better understood by establishing their asymptotic action when they are made more and more local. This action is then described by a PDE.

- Details and fine structures can be **excessively filtered** because of the window comparison and the exponential kernel.



Geometrical explanation



The number of points y satisfying $u(x) - h < u(y) \leq u(x)$ is larger than the number satisfying $u(x) \leq u(y) < u(x) + h$. Thus, the average value $YNF(x)$ is smaller than $u(x)$, enhancing that part of the signal. The regression line of u inside (x_-, x_+) better approximates the signal at x.

Linear regression correction

- Locally approximate the image by a plane.
- The filtered value at $\mathbf{x} = (x_1, x_2)$ is given by $ax_1 + bx_2 + c$, where a, b, c minimize

$$\min_{a,b,c} \int_{B_\rho(\mathbf{x})} w(\mathbf{x}, \mathbf{y})(u(\mathbf{y}) - ay_1 - by_2 - c)^2 d\mathbf{y}$$

and

$$w(\mathbf{x}, \mathbf{y}) = e^{-\frac{|u(\mathbf{y}) - u(\mathbf{x})|^2}{h^2}}.$$

Points with a grey level value close to $u(x)$ will have a larger influence in the minimization process than those with a further grey level value.

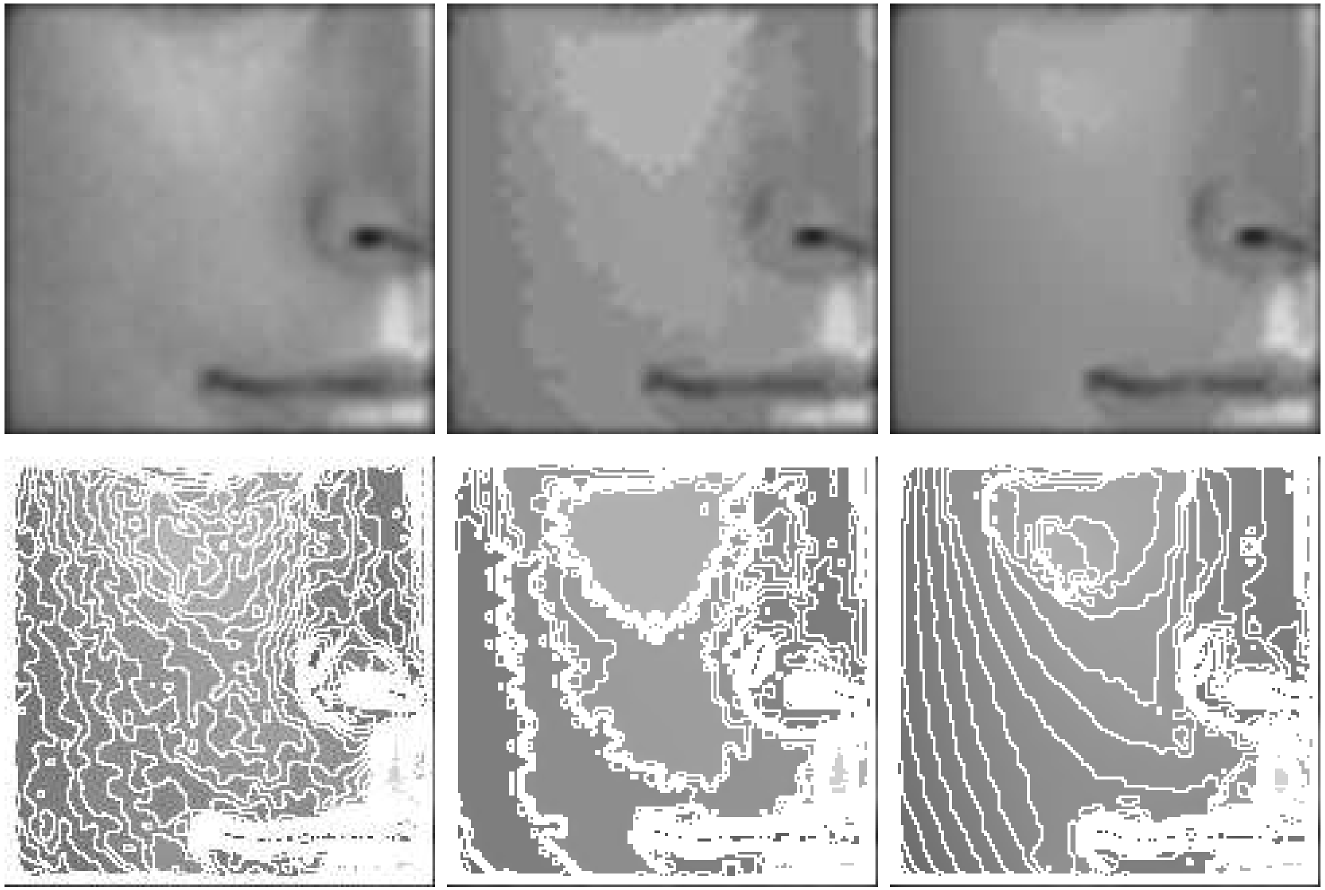
Theorem 5 *Suppose $u \in C^2(\Omega)$, and let $\rho, h > 0$ such that $\rho, h \to 0$ and $O(\rho) = O(h)$. Let $\tilde{g}$ be the continuous function defined as $\tilde{g}(0) = \frac{1}{6}$,*

$$\tilde{g}(t) = \frac{1}{4t^2} \quad 1 - \frac{2t\, e^{-t^2}}{E(t)} \Bigg),$$

for $t \neq 0$, where $E(t) = 2\int_0^t e^{-s^2}\, ds$. Then

$$LYNF_{h,\rho} u(\boldsymbol{x}) - u(\boldsymbol{x}) \simeq (u_{\xi\xi} + \tilde{g}(\frac{\rho}{h}\,|Du|)\, u_{\eta\eta})\, \frac{\rho^2}{6}.$$

where $\tilde{g}$ is a positive and decreasing to zero function satisfying $g(0) = 1$.



The regression correction filters the level lines by a curvature motion without computing any derivative.

The NL-means linear regression correction

- In order to apply the regression correction to the NL-means algorithm we restrict the search zone for a pixel $\mathbf{x} = (x_1, x_2)$ to a neighborhood $B_\rho(\mathbf{x})$. The filtered value is given by $ax_1 + bx_2 + c$, where a, b, c minimize

$$\min_{a,b,c} \int_{B_\rho(\mathbf{x})} w(\mathbf{x}, \mathbf{y})(u(\mathbf{y}) - ay_1 - by_2 - c)^2 d\mathbf{y}$$

and

$$w(\mathbf{x}, \mathbf{y}) = e^{-\frac{1}{h^2} \int_{\mathbb{R}^2} G_a(t)|u(\mathbf{x}+t)-u(\mathbf{y}+t)|^2 dt}.$$

- The Neighborhood filter

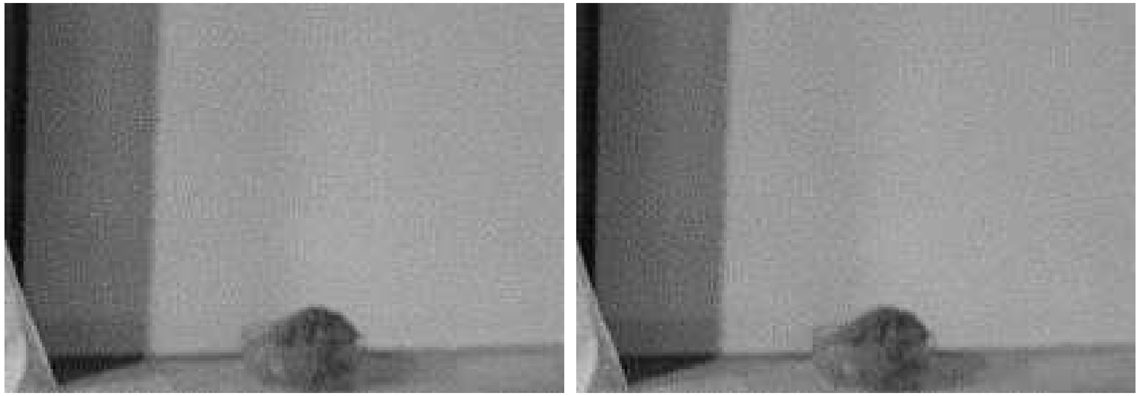
  

- The NL-means

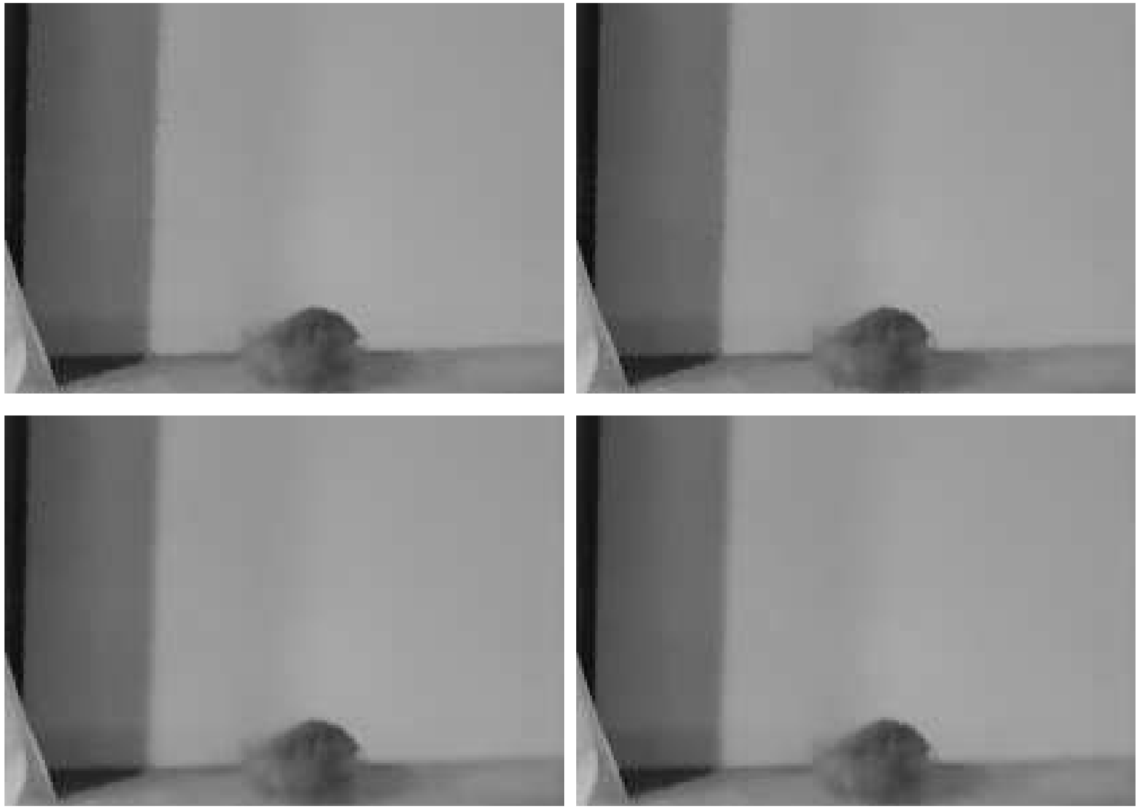
- Dr Mabuse sequence.

Linear regression correction films

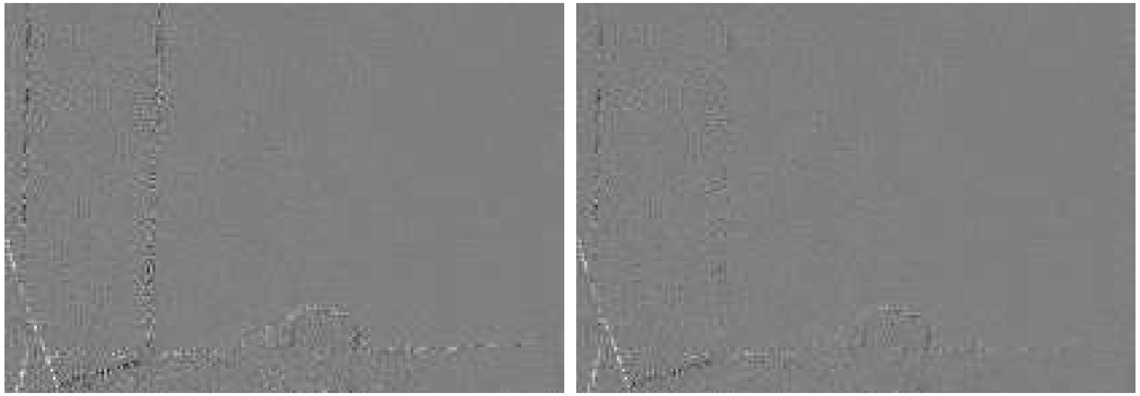
The irregularities of the edges due to the noise are enhanced and lead to very irregular and oscillatory contours. The irregularities of a certain edge are different from frame to frame leading to a false motion impression when the sequence is played.



Two consecutive frames of a noisy sequence



Frames filtered by the neighborhood filter and its correction



Difference between the two consecutive frames

The linear regression correction avoids this effect and oscillations from frame to frame are reduced.

STATISTICAL CONSISTENCY OF NL-MEANS

- Hypothesis: As the size of the image grows we can find many similar samples for all the details of the image (stationarity assumptions).
  - Let Z denote the sequence of random variables $Z_i = \{V(i), V(\mathcal{N}_i \backslash \{i\})\}$
  - Let denote $\hat{NL}_n$ the NL-means algorithm applied to the subsequence of Z, $Z_n = \{V(i), V(\mathcal{N}_i \backslash \{i\})\}_{i=1}^n$ and where $v(\mathcal{N}_i \backslash \{i\})$ is used to compute the weights instead of $v(\mathcal{N}_i)$.
  - Let $r(i)$ denote $E[V(i) \mid V(\mathcal{N}_i \backslash \{i\}) = v(\mathcal{N}_i \backslash \{i\})]$.

- **Theorem 6 (Conditional expectation theorem)** *Let* $Z = \{V(i), V(\mathcal{N}_i \backslash \{i\})\}$ *for* $i = 1, 2, \ldots$ *be a strictly stationary and mixing process. Then,*

$$|NL_n(i) - r(i)| \to 0 \quad a.s$$

for $i \in \{1, \ldots, n\}$.

- Additive white noise model

  Theorem 7 *Let* V, U, N *be random fields on* I *such that* $V = U + N$, *where* N *is a signal independent white noise. Then, the following statements are hold.*

  (i) $E[V(i) \mid V(\mathcal{N}_i \backslash \{i\}) = x] = E[U(i) \mid V(\mathcal{N}_i \backslash \{i\}) = x]$ *for all* $i \in I$ *and* $x \in \mathbb{R}^p$.

  (ii) The expected random variable $E[U(i) \mid V(\mathcal{N}_i \backslash \{i\})]$ *is the function of* $V(\mathcal{N}_i \backslash \{i\})$ *that*

minimizes the mean square error

$$\min_g E[U(i) - g(V(\mathcal{N}_i \backslash \{i\}))]^2 \tag{1}$$