UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

A Unified Framework for Image Restoration

F. Malgouyres

August 2001
CAM Report 01-23

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90095-1555

http://www.math.ucla.edu/applied/cam/index.html

A unified framework for image restoration*

F. Malgouyres†

Abstract

We present in this paper a new framework for image restoration. Despite its simplicity, both some variational and some wavelet approaches can be expressed in this framework. This permits to define a model with only one parameter which has the advantages of both approaches. We then give a mathematical analysis of this model and show some experiments.

1 Introduction

This paper is mainly concerned with image and signal restoration. More precisely, we present two classical methods (the wavelet thresholding and the Rudin-Osher-Fatemi method) under the same point of view. In this Framework, it appears that wavelet methods mainly focus on the data fidelity term while variational methods are more concerned with the regularity criterion. This leads us to propose a new method which combines both advantages.

In the sequel, we will understand restoration as methods whose aim is to recover an image (similarly a signal) $u \in L^2(\mathbb{T})$, from a data

$$v = H(u) + b\,,$$

where $\mathbb{T}$ is the torus, H is a linear and continuous operator which goes from $L^2(\mathbb{T})$ into itself and b is a Gaussian noise of standard deviation σ.

There is a large number of papers dealing with this problem. Among these papers two "families" are often opposed : the wavelet and variational approaches. Among variational approaches, those based on the minimization of the total variation, as introduced in [15], are often considered as being the most efficient (see [1, 3, 6, 10, 12, 13]). On the other hand, wavelet methods was introduced by Donoho and Johnstone and are studied and extended in several papers (see [4, 5, 8, 9, 14, 16]).

The paper is organized as follow :

- In section 2, we expose the unified framework and propose the model we want to solve.
- In section 3, we state the mathematical results which guaranty that the proposed model has a computable solution.
- In section 4, we expose a modification of the model which facilitates the computation of its solution and briefly describe the algorithm used to compute a solution.
- In section 5, we display some experiments on the model.

2 A unified framework

We are going to expose here a framework in which fits a lot of restoration methods. Among these are Maximum A Posteriori methods (such as the Rudin-Osher-Fatemi method) and wavelet approaches (such

*This work was started while I participated to the "Geometrically Based Motion" program of the "Institute for Pure and Applied Mathematics" and was finished while I was supported by the grant ONR-N00014-97-1-0027.

†UCLA Dept of Mathematics, 6363 Math science building, Box 951555, Los Angeles, CA 90095-1555, USA;
http://www.math.ucla.edu/~malgouy
malgouy@math.ucla.edu

as the wavelet soft-thresholding). This model is very simple and describes a method in terms of the minimization of an entropy $E(w)$ under a constraint. This constraint is that $(H(w) - v)$ (the part we are removing) must belong to a set $\mathcal{N}_{\mathcal{D},\tau}$ defined by

$$\mathcal{N}_{\mathcal{D},\tau} = \{w \in L^2(\mathbb{T}), \forall \Psi \in \mathcal{D}, |\langle w, \Psi\rangle| \leq \tau\}$$

for a dictionary $\mathcal{D} = \{\Psi_l\}_{l\in I}$ of functions of $L^2(\mathbb{T})$.

The definition of the entropy together with the dictionary $\mathcal{D}$ have great consequences on the result. Let us explain this through two examples : the wavelet soft-thresholding and the Rudin-Osher-Fatemi model.

The wavelet soft-thresholding :

For simplicity, we only consider the case of denoising in a wavelet basis. One can refer to [14, 9] for the deblurring case. The usual wavelet soft-thresholding $W(v)$ (see [5]) can be rewritten in the above framework by taking $\mathcal{D}$ equal to the wavelet basis and an entropy of the type[1]

$$E(w) = \sum_l c_l \langle w, \Psi_l\rangle^{i_l} \tag{1}$$

for any $c_l \geq 0$ and $i_l \geq 0$.

We see that wavelet method does not pay too much attention to the entropy and focuses on the dictionary. However, among such entropies (in the case of an appropriate wavelet basis) are a number of Besov norm. Note also that a soft-thresholding of the coordinates of an image in any basis (curvelet, wavelet packet for the deconvolution) can always be expressed in the framework of a minimization of entropies of the form (1).

Probably the main drawback of wavelet soft-thresholding methods is that they are local in the wavelet domain ($\langle W(v), \Psi_l\rangle$ only depends on $\langle v, \Psi_l\rangle$). This drawback is of importance in the case of image deblurring when some information is lost during the degradation. In this case, we could expect to restore $\langle W(v), \Psi_l\rangle$ according to some information contained in other coordinates of v. That is one of the advantage of the next method. An other drawback of the method is that the constraint $(W(v) - v) \in \mathcal{N}_{\mathcal{D},\tau}$ only constraints the movement along orthogonal directions while we could expect to do it in other direction (typically use a real dictionary instead of only a basis).

Note that the good choice for $\mathcal{D}$ permits to have a small threshold. For instance, in the case of finite dimensional images of size N, $\tau = \sigma\sqrt{2\ln N}$ is considered (see [5]) an optimal choice for the threshold.

The Rudin-Osher-Fatemi model :

The Rudin-Osher-Fatemi restoration method can be stated in several forms. One of those (see [1]) is the constrained minimization problem[2] :

$$\text{Minimize, } \int_{\mathbb{T}} |\nabla w|,$$

among functions such that

$$\int_{\mathbb{T}} |H(w) - v|^2 \leq \sigma^2 .$$

This method can be expressed in the above framework by noting that

$$\{w \in L^2(\mathbb{T}), \int_{\mathbb{T}} |H(w) - v|^2 \leq \sigma^2\} = \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\sigma}\}$$

with

$$\mathcal{D} = \{\Psi \in L^2(\mathbb{T}), \|\Psi\|_2 = 1\} .$$

[1]This has already been noticed in [2] in the case of an l^2 constraint.

[2]For simplicity, we note the total variation with an integral instead of the total mass of a measure. This notation is only valid when the function is differentiable though. One can refer to [7] for details on the total variation and the space $BV(\mathbb{T})$.

The main advantage of this method is that it allows to reconstruct some lost information (see [10, 12]).

However, when presented that way it is clear that the constraint could be improved since it restricts changes along direction which are not autocorrelated (such as the direction of the noise b). Therefore, the parameter τ of this method has to be much larger than in the previous case. Indeed, since $\frac{b}{\|b\|_2} \in \mathcal{D}$ we must have $\tau = \sigma$. For instance, in the case of the restoration of a discrete image of size N it becomes $\tau = \sigma N$ which has to be compared to $\sigma\sqrt{2\ln N}$ for wavelet shrinkage).

An hybrid model :

This leads us to a model where we minimize the total variation under a constraint defined by a dictionary smaller than for the Rudin-Osher-Fatemi method. Moreover, since for such a model we do not use a reconstruction formula such as in the case of the wavelet thresholding, we can use a dictionary which contains more than one basis. This leads us to

$$\begin{array}{rl} \text{Minimize,} & \int_{\mathbb{T}} |\nabla w| \\ \text{under the constraint} & (H(w) - v) \in \mathcal{N}_{\mathcal{D},\tau} \end{array} \tag{2}$$

For a dictionary $\mathcal{D}$ and a parameter $\tau > 0$. Note that such a model has recently and independently been introduced in [16] for the purpose of denoising. The design of the dictionary is of course now the key problem. It has to be understood as the set of the "structure" we do not want to erase. Therefore it seems important to have very different types of elements such as wavelet and curvelet. However, in the case of the deconvolution (or the inversion of an operator H), we want to control the noise in some particular direction along which the noise is enhanced during the inversion of H. For instance for the deconvolution problem it is advisable to have some wavelet packet or a Fourier basis in the dictionary.

3 Mathematical analysis of the model

In this section, we are going to show that the proposed model is interesting in the sense that it has a solution and we can compute it.

Theorem 1 *Let $v \in L^2(\mathbb{T})$ and H be a linear operator continuous from $L^2(\mathbb{T})$ into itself. Let $\mathcal{D} \subset L^2(\mathbb{T})$ and $\tau > 0$. Assume $BV(\mathbb{T}) \cap \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\} \neq \emptyset$. Then (2) admits a solution $w \in BV(\mathbb{T}) \cap L^2(\mathbb{T})$.*

We can unfortunately not guaranty the uniqueness of the solution to such a model. Indeed neither $\{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\}$ nor the total variation are strictly convex. However, we could probably state with this regard some results similar to the one given in [1] and [6].

In order to find a solution to (2), we could of course use a relaxation method, but it might converge very slowly. We could also use a projected steepest descent algorithm but, at each iteration, the projection needs a lot of calculus and can probably be compared to a matching pursuit. This leads us to define a method by penalization. We will see however that in the general case we have difficulties in actually computing our solution this way and have to approximate our model

Theorem 2 *Let $v \in L^2(\mathbb{T})$ and H be a linear operator continuous from $L^2(\mathbb{T})$ into itself. Let $\mathcal{D} \subset L^2(\mathbb{T})$ be a countable set and $\tau > 0$. Assume $BV(\mathbb{T}) \cap \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\} \neq \emptyset$. Then, for any $\epsilon > 0$,*

$$\int_{\mathbb{T}} |\nabla w| + \frac{1}{\epsilon} \sum_{\Psi \in \mathcal{D}} (|\langle H(w) - v, \Psi\rangle| - \tau)^{+2} \tag{3}$$

has a solution w_ϵ. Moreover, we can extract a sequence $(w_{\epsilon_l})_{l\in\mathbb{N}}$ (with $\lim_{l\to\infty} \epsilon_l = 0$) that converges to a function $w \in BV(\mathbb{T}) \cap L^2(\mathbb{T})$. w is solution to (2).

4 Numerical algorithm and approximation to the model

In the case of denoising, we can do a gradient descent algorithm in order to minimize (3). For classical method on the minimization of functionals involving the total variation, one can refer to [3, 13]. In our

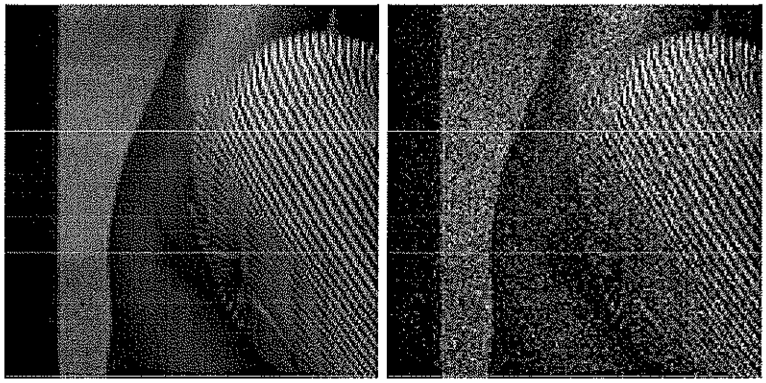

Figure 1: Left : Original image; Right : Noisy image.

particular case, for a general H, in order to compute the steepest descent for the data fidelity term we need to compute and store all the

$$\langle H(w_l), \Psi \rangle$$

for $\Psi \in \mathcal{D}$ and all the elements of a basis (w_l). Note that for some particular cases this could reasonably be achieved with the help of a wavelet-vaguelet decomposition (see [4]). However, it needs in general too much memory and this leads us to modify the model.

Basically, we modify $\{w \in L^2(\mathbb{T}), H(w)-v \in \mathcal{N}_{\mathcal{D},\tau}\}$. More precisely, we assume known a pseudo-inverse H^{-1} to H and we replace $\{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\}$ by

$$\widetilde{\mathcal{N}}_{\mathcal{D},\tau} = \{w \in L^2(\mathbb{T}), \forall \Psi \in \mathcal{D}, \tau_\Psi |\langle w - H^{-1}(v), \Psi \rangle| \leq 1\}.$$

Of course the risk of such an approximation is when $\langle H^{-1}(v), \Psi \rangle$ does depend on the definition of H^{-1}. For instance, when H is a convolution with a filter whose Fourier transform is equal to 0 at the frequency ξ we do want $\tau_{e^{i\xi}}$ to be equal to 0. This is one of the aspect we need to be careful about when doing this approximation. In practice, we take

$$\tau_\Psi = \frac{\langle H(\Psi), \Psi \rangle}{\tau}$$

or an approximation of this scalar product.

5 Experimental results

All the wavelet packet bases used in this section are based on the cubic spline wavelet. Concerning their tree, they are all of maximum depth 3. For image deblurring, we use the mirror tree which is described in [8].

5.1 Experiments on denoising

All the experiments of this section are made with the data displayed on Figure 1. On this figure, the image on the right is obtained by adding a Gaussian noise of standard deviation 20 to the image on the left.

We display the results of several denoising methods on Figure 2. Here is the description of the methods[3].

- Up : Rudin-Osher-Fatemi method with $\lambda = 0.01$.
- Middle-Left : The wavelet soft-thresholding with a parameter $\tau = 70$.

[3]Note that all these methods only involve one parameter. We could (for instance like in [9, 11]) get better results by adding other parameters.

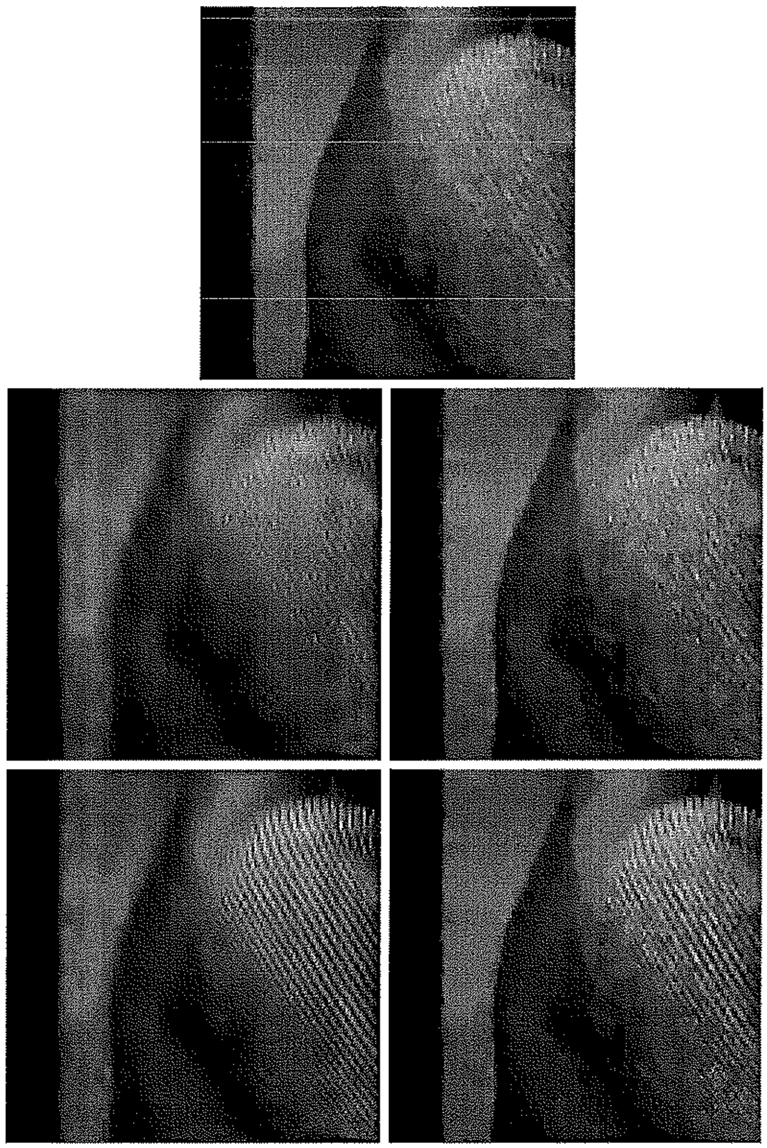

Figure 2: Restoration with: Up : Rudin-Osher-Fatemi method; Middle-Left : wavelet soft-thresholding; Middle-Right : solution of (2) with only a wavelet basis; Down-Left : noise selection with a wavelet packet dictionary; Down-Right : solution of (2) with a wavelet packet dictionary.

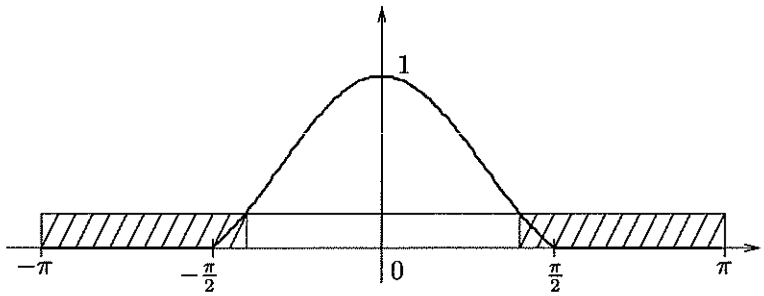

Figure 3: Profile of the Fourier transform of h (see (4)). The hatching represents the frequencies which are, in practice, lost during the degradation.

- Middle-Right : The restoration using (2) with just one wavelet basis in the dictionary and a parameter $\tau = 70$.
- Down-Left : The noise selection approach (the idea is just to compose the restoration in all the wavelet packet bases) with a wavelet packet dictionary (see [11]), with a parameter $\tau = 70$.
- Down-Right : The restoration using (2) with a dictionary made of 4 translations of a wavelet packet dictionary for a parameter $\tau = 120$.

First note that be it with a wavelet basis or a wavelet packet dictionary the introduction of the total variation compared to the thresholding (or noise selection) yields sharper edges. Note also that they do not present Gibbs phenomena in the vicinity of these edges. This is due to the fact that (2) allows the reconstruction of some small coefficient which are canceled by the thresholding.

Moreover, if we compare the images of the middle with the one at the bottom of Figure 2, we clearly see that we gain in putting a larger dictionary. Indeed, the texture of the pans can be represented by few large wavelet packet coefficients (well localized in Fourier domain) or a lot of small wavelet coefficients. This explains why they are preserved with the larger dictionary. However, we do not claim here that the wavelet packet dictionary is particularly appropriate and it is even probable that dictionaries made of very different elements and/or texture would, in general, be more efficient.

Finally, the texture is better preserved in the image on the bottom right than in the one with Rudin-Osher-Fatemi method since we restrict the evolution of the result in the direction of the texture much more in the case of (2) than in the usual Rudin-Osher-Fatemi model.

5.2 Experiments on deblurring

We present here some experiments on image deconvolution. Therefore, we want to recover u given

$$v = h * u + b$$

Since we do not want to consider the aliasing in the creation of the image and want to show to evidence the ability of the method to struggle against Gibbs phenomena, we take h such that it cancels all the frequencies outside $[-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$. More precisely, the Fourier transform of h is given by

$$\widehat{h}(\xi, \eta) = \left(\frac{sin(2\xi)}{2\xi}\right)\left(\frac{sin(2\eta)}{2\eta}\right), \tag{4}$$

for ξ and $\eta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and 0 otherwise (see Figure 3). We add a Gaussian noise of standard deviation 2. Note that this degradation model is particularly not adapted to wavelet packet methods since it cancels a wide band of frequencies (see Figure 3). It fact, we know that, because of their ability to reconstruct some lost frequencies (see [6]), variational methods are better suited to this kind of degradation model.

We display on Figure 4

- Up : the reference image.

- Middle-Left : the degraded image.
- Middle-Right : The noise selection approach with a sub-dictionary of the wavelet packet dictionary (see [11]), with a parameter $\tau = 12$. The idea is just to compose the restoration in all the wavelet packet bases which are more decomposed than the mirror basis (see [8]) and whose maximum depth is 3. Note that we also average the result of the algorithm over 4 translations.
- Down-Left : The restoration with Rudin-Osher-Fatemi method with a parameter $\lambda = 0.1$.
- Down-Right : The restoration with (2) for a parameter $\tau = 12$ and a dictionary made of all the wavelet packet basis which are more decomposed than the mirror basis and whose maximum depth is 3 (as well as 4 of its translations).

It is clear that compared to Rudin-Osher-Fatemi method, the new model better preserves the texture since it is more constrained by the data fidelity term. Moreover, compared to the wavelet packet method, we have a similar constraint but, since our model permits to extrapolate the lost frequencies, we do not have any Gibbs phenomena. Note that this is of course also the case with Rudin-Osher-Fatemi method.

References

[1] A. Chambolle and P.L. Lions. Image recovery via total variation minimisation and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.

[2] A. Chambolle, R.A. De Vore, N. Lee, and B.J. Lucier. Nonlinear wavelet image processing: Variational problems, compression and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7:319–355, 1998. Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis.

[3] T. F. Chan and P. Mulet. On the convergence of the lagged diffusivity fixed method in total variation image restoration. *SIAM Journal of Numerical Analysis*, 36(2):354–367, 1999.

[4] D. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2:101–126, 1995.

[5] D. Donoho and I.M. Johnstone. Minimax estimation via wavelet shrinkage. Technical report, Departement of Stat., Stanford University, 1992.

[6] S. Durand, F. Malgouyres, and B. Rougé. Image de-blurring, spectrum interpolation and application to satellite imaging. *Control, Optimisation and Calculus of Variation*, 5(445-475), 2000. A preliminary version is available at http://www.math.ucla.edu/~malgouy.

[7] L.C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, 1992.

[8] J. Kalifa. *Restauration minimax et déconvolution dans une base d'ondelettes miroirs*. PhD thesis, Ecole Polytechnique, 1999. Available at http://www.cmap.polytechnique.fr/~kalifa.

[9] F. Malgouyres. A framework for image deblurring using wavelet packet bases. CAM report at UCLA, available at http://www.math.ucla.edu/~malgouy, November 2000.

[10] F. Malgouyres. *Increase in the Resolution of Digital Images: Variational Theory and Applications*. PhD thesis, Ecole Normale Supérieure de Cachan, Cachan, France, 2000. Available at http://www.math.ucla.edu/~malgouy.

[11] F. Malgouyres. A noise selection approach of image restoration. In *SPIE, international Conference on Wavelets IX*, volume 4478, San Diego, August 2001.

[12] F. Malgouyres and F. Guichard. Edge direction preserving image zooming: a mathematical and numerical analysis. *SIAM, J. Num. Anal.*, 39(1):1–37, 2001. A preliminary version is available at: http://www.math.ucla.edu/~malgouy.

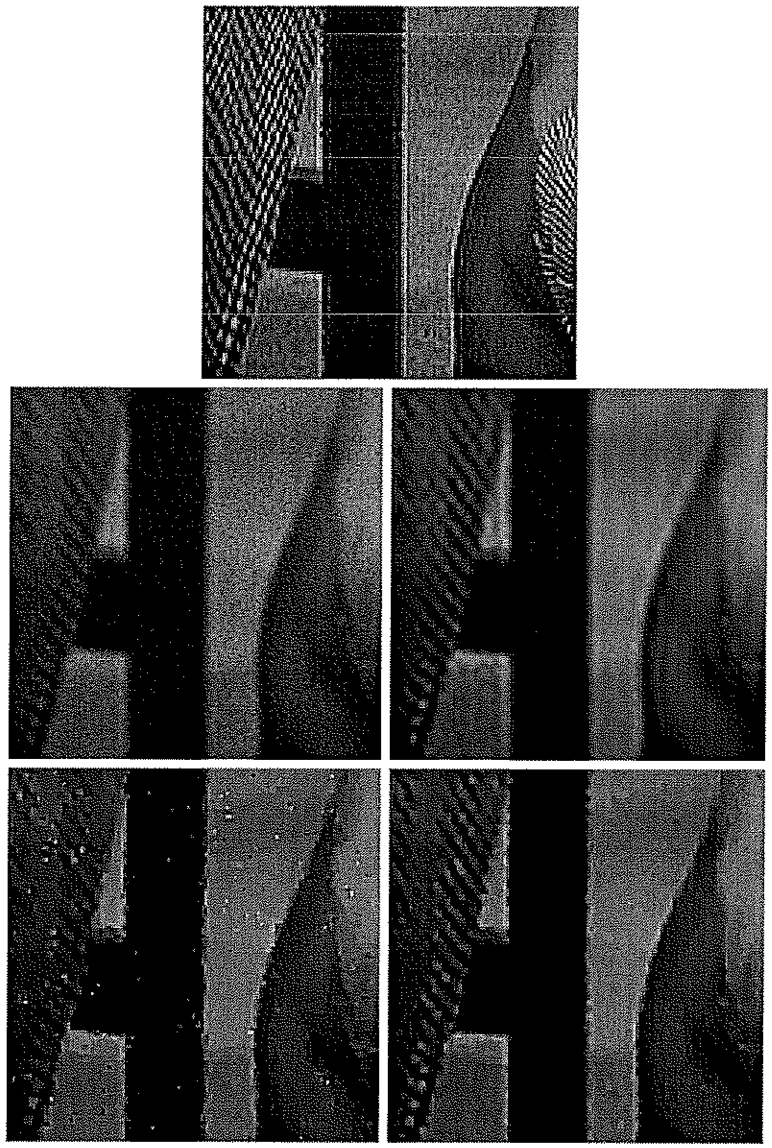

Figure 4: All the images have been **sharpened** for the display. Up : Original image; Middle-Left : Degraded image; Middle-Right : Noise selection in a sub-dictionary of the wavelet packet dictionary; Down-Left : Rudin-Osher-Fatemi method; Down-Right : solution of (2) with a sub-dictionary of the wavelet packet dictionary.

[13] A. Marquina and S. Osher. Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal. *SIAM, Journal of Sientific Computing*, 22(2):387–405, 2000.

[14] B. Rougé. Fixed chosen noise restauration (fcnr). In *IEEE 95 Philadelphia*, 1995.

[15] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[16] J.L. Starck, D. Donoho, and E. Candes. Very high quality image restoration. In *Wavelet and applications in signal and image processing IX*, volume 4478, San Diego, August 2001. SPIE.