Image Denoising Methods.
A New Nonlocal Principle*

A. Buades[†]
B. Coll[‡]
J. M. Morel[§]

Abstract. The search for efficient image denoising methods is still a valid challenge at the crossing of functional analysis and statistics. In spite of the sophistication of the recently proposed methods, most algorithms have not yet attained a desirable level of applicability. All show an outstanding performance when the image model corresponds to the algorithm assumptions but fail in general and create artifacts or remove fine structures in images. The main focus of this paper is, first, to define a general mathematical and experimental methodology to compare and classify classical image denoising algorithms and, second, to propose a nonlocal means (NL-means) algorithm addressing the preservation of structure in a digital image. The mathematical analysis is based on the analysis of the "method noise," defined as the difference between a digital image and its denoised version. The NL-means algorithm is proven to be asymptotically optimal under a generic statistical image model. The denoising performance of all considered methods is compared in four ways; mathematical: asymptotic order of magnitude of the method noise under regularity assumptions; perceptual-mathematical: the algorithms artifacts and their explanation as a violation of the image model; quantitative experimental: by tables of L^2 distances of the denoised version to the original image. The fourth and perhaps most powerful evaluation method is, however, the visualization of the method noise on natural images. The more this method noise looks like a real white noise, the better the method.

1. Introduction.

1.1. Digital Images and Noise. The need for efficient image restoration methods has grown with the massive production of digital images and movies of all kinds, often

taken in poor conditions. No matter how good cameras are, an image improvement is always desirable to extend their range of action.

A digital image is generally encoded as a matrix of grayscale or color values. In the case of a movie, this matrix has three dimensions, the third one corresponding to time. Each pair $(i, u(i))$, where $u(i)$ is the value at i, is called a pixel, short for "picture element." In the case of grayscale images, i is a point on a two-dimensional (2D) grid and $u(i)$ is a real value. In the case of classical color images, $u(i)$ is a triplet of values for the red, green, and blue components. All of what we shall say applies identically to movies, three-dimensional (3D) images, and color or multispectral images.

The two main limitations in image accuracy are categorized as blur and noise. Blur is intrinsic to image acquisition systems, as digital images have a finite number of samples and must satisfy the Shannon–Nyquist sampling conditions [81]. The second main image perturbation is noise.

Each one of the pixel values $u(i)$ is the result of a light intensity measurement, usually made by a charge coupled device (CCD) matrix coupled with a light focusing system. Each captor of the CCD is roughly a square in which the number of incoming photons is being counted for a fixed period corresponding to the obturation time. When the light source is constant, the number of photons received by each pixel fluctuates around its average in accordance with the central limit theorem. In other words, one can expect fluctuations of order $\sqrt{n}$ for n incoming photons. In addition, each captor, if not adequately cooled, receives heat photons. This perturbation is usually called "obscurity noise." In a first rough approximation one can write

$$v(i) = u(i) + n(i),$$

where $i \in I$, $v(i)$ is the observed value, $u(i)$ would be the "true" value at pixel i, namely, the one which would be observed by averaging the photon counting on a long period of time, and $n(i)$ is the noise perturbation. As indicated, the amount of noise is signal-dependent; that is, $n(i)$ is larger when $u(i)$ is larger. In noise models, the normalized values of $n(i)$ and $n(j)$ at different pixels are assumed to be independent random variables, and one talks about "white noise."

1.2. Signal and Noise Ratios. A good quality photograph (for visual inspection) has about 256 grayscale values, where 0 represents black and 255 represents white. Measuring the amount of noise by its standard deviation, $\sigma(n)$, one can define the signal-to-noise ratio (SNR) as $SNR = \frac{\sigma(u)}{\sigma(n)}$, where $\sigma(u)$ denotes the empirical standard deviation of u, $\sigma(u) = \left(\frac{1}{|I|} \sum_{i \in I} (u(i) - \overline{u})^2\right)^{\frac{1}{2}}$, and $\overline{u} = \frac{1}{|I|} \sum_{i \in I} u(i)$ is the average grayscale value. The standard deviation of the noise can also be obtained as an empirical measurement or formally computed when the noise model and parameters are known. A good quality image has a standard deviation of about 60.

The best way to test the effect of noise on a standard digital image is to add a Gaussian white noise, in which case $n(i)$ are independently and identically distributed (i.i.d.) Gaussian real variables. When $\sigma(n) = 3$, no visible alteration is usually observed. Thus, a $\frac{60}{3} \simeq 20$ SNR is nearly invisible. Surprisingly enough, one can add white noise up to a $\frac{2}{1}$ ratio and still *see* everything in a picture! This fact is illustrated in Figure 1 and constitutes a major enigma of human vision. It justifies the many attempts to define convincing denoising algorithms. As we shall see, the results have been rather disappointing. Denoising algorithms see no difference between small details and noise, and therefore they remove them. In many cases, they create new distortions, and researchers are so used to them that they have created a taxonomy of denoising artifacts: "ringing," "blur," "staircase effect," "checkerboard

Fig. 1 *A digital image with standard deviation* 55, *the same with noise added (standard deviation* 3*), the SNR therefore being equal to* 18, *and the same with SNR slightly larger than* 2. *In the second image, no alteration is visible. In the third, a conspicuous noise with standard deviation* 25 *has been added, but, surprisingly enough, all details of the original image still are visible.*

effect," "wavelet outliers," etc. This fact is not quite a surprise. Indeed, to the best of our knowledge, all denoising algorithms are based on

- a noise model;
- a generic image smoothness model, local or global.

In experimental settings, the noise model is perfectly precise. So the weak point of the algorithms is the inadequacy of the image model. All of the methods assume that the noise is oscillatory and that the image is smooth or piecewise smooth. So they try to separate the smooth or patchy part (the image) from the oscillatory one. Actually, many fine structures in images are as oscillatory as noise is; conversely, white noise has low frequencies and therefore smooth components. Thus a separation method based on smoothness arguments only is hazardous.

1.3. The "Method Noise." All denoising methods depend on a filtering parameter h. For most methods, the parameter h depends on an estimation of the noise variance σ^2. One can define the result of a denoising method D_h as a decomposition of any image v as $v = D_h v + n(D_h, v)$, where

1. $D_h v$ is more smooth than v;
2. $n(D_h, v)$ is the noise guessed by the method.

It is not enough to smooth v to ensure that $n(D_h, v)$ will look like a noise. The more recent methods are actually not content with a smoothing but try to recover lost information in $n(D_h, v)$ [59, 71]. So the focus is on $n(D_h, v)$.

DEFINITION 1.1 (method noise). *Let u be a (not necessarily noisy) image and D_h a denoising operator depending on h. Then we define the* method noise *of u as the image difference*

$$(1.1) \qquad\qquad n(D_h, u) = u - D_h(u).$$

This method noise should be as similar to a white noise as possible. In addition, since we would like the original image u not to be altered by denoising methods, the method noise should be as small as possible for the functions with the right regularity.

According to the preceding discussion, five criteria can and will be taken into account in the comparison of denoising methods:

- A display of typical artifacts in denoised images.
- A formal computation of the method noise on smooth images, evaluating how small it is in accordance with image local smoothness.

- A comparative display of the *method noise* of each method on real images with $\sigma = 2.5$. We mentioned that a noise standard deviation smaller than 3 is subliminal, and it is expected that most digitization methods allow themselves this amount of noise.
- A classical comparison receipt based on noise simulation: it consists of taking a good quality image, adding Gaussian white noise with known σ, and then computing the best image recovered from the noisy one by each method. A table of L^2 distances from the restored to the original can be established. The L^2 distance does not provide a good quality assessment. However, it reflects well the relative performances of algorithms.
- A comparison following the "noise to noise" criterion, which requires the residual noise to be as white as possible, and therefore artifact free.

On top of this, in two cases, a proof of asymptotic recovery of the image can be obtained by statistical arguments.

1.4. Which Methods to Compare. We had to make a selection of the denoising methods we wished to compare. Here a difficulty arises, as most original methods have caused an abundant literature proposing many improvements. So we tried to get the best available version, while keeping the simple and genuine character of the original method: no hybrid method. We shall analyze the following:

1. Gaussian smoothing (Gabor quoted in Lindenbaum, Fischer, and Bruckstein [54]);
2. anisotropic filtering (Perona and Malik [72], Alvarez, Lions, and Morel [3]);
3. the Rudin–Osher–Fatemi total variation [79];
4. the Yaroslavsky neighborhood filters [95, 97];
5. the Wiener local empirical filter as implemented by Yaroslavsky [97];
6. the translation invariant wavelet thresholding [24], a simple and performing variant of the wavelet thresholding [33];
7. the nonlocal means (NL-means) algorithm, which we introduce here.

This last algorithm is given by a simple closed formula. Let u be defined in a bounded domain $\Omega \subset \mathbb{R}^2$; then

$$NL(u)(\mathbf{x}) = \frac{1}{C(\mathbf{x})} \int e^{-\frac{(G_a * |u(\mathbf{x}+.) - u(\mathbf{y}+.)|^2)(0)}{h^2}} \ u(\mathbf{y}) \ d\mathbf{y},$$

where $\mathbf{x} \in \Omega$, G_a is a Gaussian kernel of standard deviation a, h acts as a filtering parameter, and $C(\mathbf{x}) = \int e^{-\frac{(G_a * |u(\mathbf{x}+.) - u(\mathbf{z}+.)|^2)(0)}{h^2}} d\mathbf{z}$ is the normalizing factor. In order to make clear the previous definition, we recall that

$$(G_a * |u(\mathbf{x} + .) - u(\mathbf{y} + .)|^2)(0) = \int_{\mathbb{R}^2} G_a(\mathbf{t})|u(\mathbf{x} + \mathbf{t}) - u(\mathbf{y} + \mathbf{t})|^2 d\mathbf{t}.$$

This amounts to saying that $NL(u)(\mathbf{x})$, the denoised value at $\mathbf{x}$, is a mean of the values of all pixels whose Gaussian neighborhood looks like the neighborhood of $\mathbf{x}$.

1.5. Plan of the Paper. Section 2 computes formally the *method noise* for the best elementary local smoothing methods, namely, Gaussian smoothing, anisotropic smoothing (mean curvature motion), total variation minimization, and the neighborhood filters. For all of them we prove or recall the asymptotic expansion of the filter at smooth points of the image and therefore obtain a formal expression of the method noise. This expression permits us to characterize places where the filter performs well and where it fails. In section 3, we treat the Wiener-like methods, which proceed by

a soft or hard threshold on frequency or space-frequency coefficients. We examine in turn the Wiener–Fourier filter, the Yaroslavsky local adaptive discrete cosine transform (DCT)-based filters, and the wavelet threshold method. Of course, the Gaussian smoothing belongs to both classes of filters. In section 4, we introduce the NL-means filter. This method is not easily classified in the preceding terminology, since it can work adaptively in a local or nonlocal way. We first give a proof that this algorithm is asymptotically consistent (it gives back the conditional expectation of each pixel value given an observed neighborhood) under the assumption that the image is a fairly general stationary random process. The works of Efros and Leung [38] and Levina [52] have shown that this assumption is sound for images having enough samples in each texture patch. In section 5, we compare all algorithms from several points of view, do a performance classification, and explain why the NL-means algorithm shares the consistency properties of most of the aforementioned algorithms. Section 6 is new. It reviews the numerous improvements and generalizations of NL-means discovered since 2005.

1.6. Note to the Reader. The present paper is an updated version of "A Review of Image Denoising Algorithms, with a New One" [20]. The text and structure of the original paper have been preserved. However, several spurious comparisons, technical proofs, and appendices have been adapted or removed. On the recommendation of the editorial board, the controversial benchmark image Lena has been replaced. The new section 6 reviews the abundant literature on "nonlocal image processing" stemming from the original paper. The denoising algorithm NL-means can be tested online: *http://mw.cmla.ens-cachan.fr/megawave/demo/*.

2. Local Smoothing Filters. The original image u is defined in a bounded domain $\Omega \subset \mathbb{R}^2$ and denoted by $u(\mathbf{x})$ for $\mathbf{x} = (x, y) \in \mathbb{R}^2$. This continuous image is usually interpreted as the Shannon interpolation of a discrete grid of samples [81] and is therefore analytic. The distance between two consecutive samples is denoted by ε. The noise itself is a discrete phenomenon on the sampling grid. According to the usual screen and printing visualization practice, we do not interpolate the noise samples n_i as a band-limited function but rather as a piecewise constant function, constant on each pixel i and equal to n_i. We write $|\mathbf{x}| = (x^2 + y^2)^{\frac{1}{2}}$ and $\mathbf{x}_1.\mathbf{x}_2 = x_1 x_2 + y_1 y_2$ as the norm and scalar product and denote the derivatives of u by $u_x = \frac{\partial u}{\partial x}$, $u_y = \frac{\partial u}{\partial y}$, and $u_{xy} = \frac{\partial^2 u}{\partial x \partial y}$. The gradient of u is written as $Du = (u_x, u_y)$ and the Laplacian of u as $\Delta u = u_{xx} + u_{yy}$.

2.1. Gaussian Smoothing. By Riesz's theorem, image isotropic linear filtering boils down to a convolution of the image by a linear radial kernel. The smoothing requirement is usually expressed by the positivity of the kernel. The paradigm of such kernels is, of course, the Gaussian $\mathbf{x} \rightarrow G_h(\mathbf{x}) = \frac{1}{(4\pi h^2)} e^{-\frac{|\mathbf{x}|^2}{4h^2}}$. In that case, G_h has standard deviation h, and the following theorem is easily seen.

THEOREM 2.1 (Gabor [42]). *The image method noise of the convolution with a Gaussian kernel G_h is*

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

A similar result is actually valid for any positive radial kernel with bounded variance, so one can keep the Gaussian example without loss of generality. The preceding estimate is valid if h is small enough. On the other hand, the noise reduction properties depend upon the fact that the neighborhood involved in the smoothing is large enough, so that the noise gets reduced by averaging. So in the following we

assume that $h = k\varepsilon$, where k stands for the number of samples of the function u and noise n in an interval of length h. The spatial ratio k must be much larger than 1 to ensure a noise reduction. The effect of a Gaussian smoothing on the noise can be evaluated at a reference pixel $i = 0$. At this pixel,

$$G_h * n(0) = \sum_{i \in I} \int_{P_i} G_h(\mathbf{x})n(\mathbf{x})d\mathbf{x} = \sum_{i \in I} \varepsilon^2 G_h(i)n_i,$$

where we recall that $n(x)$ is being interpolated as a piecewise constant function, the P_i square pixels centered in i have size ε^2, and $G_h(i)$ denotes the mean value of the function G_h on the pixel i. Denoting by $\mathrm{Var}\,(X)$ the variance of a random variable X, the additivity of variances of independent centered random variables yields

$$\mathrm{Var}\,(G_h * n(0)) = \sum_i \varepsilon^4 G_h(i)^2 \sigma^2 \simeq \sigma^2 \varepsilon^2 \int G_h(\mathbf{x})^2 d\mathbf{x} = \frac{\varepsilon^2 \sigma^2}{8\pi h^2}.$$

So we have proved the following theorem.

THEOREM 2.2. *Let $n(x)$ be a piecewise constant white noise, with $n(\boldsymbol{x}) = n_i$ on each square pixel i. Assume that the n_i are i.i.d. with zero mean and variance σ^2. Then the "noise residue" after a Gaussian convolution of n by G_h satisfies*

$$\mathrm{Var}\,(G_h * n(0)) \simeq \frac{\varepsilon^2 \sigma^2}{8\pi h^2}.$$

In other terms, the standard deviation of the noise, which can be interpreted as the noise amplitude, is multiplied by $\frac{\varepsilon}{h\sqrt{8\pi}}$.

Theorems 2.1 and 2.2 traduce the delicate equilibrium between noise reduction and image destruction by any linear smoothing. Denoising does not alter the image at points where it is smooth at a scale h much larger than the sampling scale ε. The first theorem tells us that the *method noise* of the Gaussian denoising method is zero in harmonic parts of the image. A Gaussian convolution is optimal on harmonic functions and instead performs poorly on singular parts of u, namely, edges or texture, where the Laplacian of the image is large. See Figure 2.

2.2. Anisotropic Filters and Curvature Motion. The anisotropic filter (AF) attempts to avoid the blurring effect of the Gaussian by *convolving the image u at $\mathbf{x}$ only in the direction orthogonal to $Du(\mathbf{x})$.* The idea of such a filter goes back to Perona and Malik [72] and actually again to Gabor (quoted in Lindenbaum, Fischer, and Bruckstein [54]). Set

$$AF_h u(\mathbf{x}) = \int G_h(t)u\left(\mathbf{x} + t\frac{Du(\mathbf{x})^\perp}{|Du(\mathbf{x})|}\right) dt$$

for $\mathbf{x}$ such that $Du(\mathbf{x}) \neq 0$ and where $(x,y)^\perp = (-y,x)$ and $G_h(t) = \frac{1}{\sqrt{2\pi}h}e^{-\frac{t^2}{2h^2}}$ is the one-dimensional (1D) Gauss function with variance h^2. At points where $Du(\mathbf{x}) = 0$ an isotropic Gaussian mean is usually applied, and the result of Theorem 2.1 holds at those points. If one assumes that the original image u is twice continuously differentiable (C^2) at $\mathbf{x}$, the following theorem is easily shown by a second-order Taylor expansion.

THEOREM 2.3. *The image method noise of an anisotropic filter AF_h is*

$$u(\boldsymbol{x}) - AF_h u(\boldsymbol{x}) \simeq -\frac{1}{2}h^2 D^2 u\left(\frac{Du^\perp}{|Du|}, \frac{Du^\perp}{|Du|}\right) = -\frac{1}{2}h^2 |Du| curv(u)(\mathbf{x}),$$

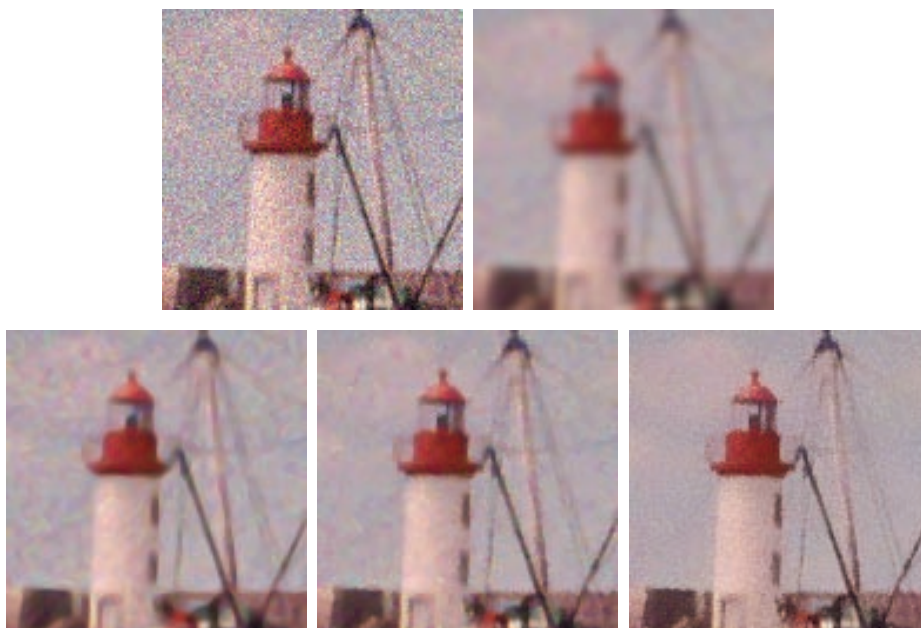*where the relation holds when $Du(\mathbf{x}) \neq 0$.*

Fig. 2 *Denoising experience on a natural image. From left to right and from top to bottom: noisy image (standard deviation $\sigma = 20$), Gaussian convolution ($h = 1.8$), anisotropic filter ($h = 2.4$), total variation ($\lambda = 0.04$), and the Yaroslavsky neighborhood filter ($\rho = 7$, $h = 28$). Parameters have been set for each algorithm so that the removed energy is equal to the energy of the added noise.*

By $curv(u)(\mathbf{x})$, we denote the curvature, i.e., the signed inverse of the radius of curvature of the level line passing by $\mathbf{x}$. When $Du(\mathbf{x}) \neq 0$, this means that

$$curv(u) = \frac{u_{xx}u_y^2 - 2u_{xy}u_xu_y + u_{yy}u_x^2}{(u_x^2 + u_y^2)^{\frac{3}{2}}}.$$

This method noise is zero wherever u behaves locally like a one-variable function, $u(x, y) = f(ax + by + c)$. In such a case, the level line of u is locally the straight line with equation $ax + by + c = 0$, and the gradient of f may instead be very large. In other terms, *with anisotropic filtering, an edge can be maintained*. On the other hand, we have to evaluate the Gaussian noise reduction. This is easily done by a 1D adaptation of Theorem 2.2. Notice that the noise on a grid is not isotropic; so the Gaussian average when Du is parallel to one coordinate axis is made roughly on $\sqrt{2}$ more samples than the Gaussian average in the diagonal direction.

THEOREM 2.4. *By anisotropic Gaussian smoothing, when ε is small enough with respect to h, the noise residue satisfies*

$$\mathrm{Var}\,(AF_h(n)) \leq \frac{\varepsilon}{\sqrt{2\pi}h}\sigma^2.$$

In other terms, the standard deviation of the noise n is multiplied by a factor at most equal to $(\frac{\varepsilon}{\sqrt{2\pi}h})^{1/2}$, this maximal value being attained in the diagonals.

Proof. Let L be the line $\mathbf{x} + t\frac{Du^{\perp}(\mathbf{x})}{|Du(\mathbf{x})|}$ passing by $\mathbf{x}$, parameterized by $t \in \mathbb{R}$, and denote by P_i, $i \in I$, the pixels which meet L, $n(i)$ the noise value, constant on

pixel P_i, and ε_i the length of the intersection of $L \cap P_i$. Denote by $g(i)$ the average of $G_h(\mathbf{x} + t\frac{Du^\perp(\mathbf{x})}{|Du(\mathbf{x})|})$ on $L \cap P_i$. Then one has $AF_h n(\mathbf{x}) \simeq \sum_i \varepsilon_i n(i) g(i)$. The $n(i)$ are i.i.d. with standard variation σ, and therefore

$$\mathrm{Var}\,(AF_h(n)) = \sum_i \varepsilon_i^2 \sigma^2 g(i)^2 \leq \sigma^2 \max(\varepsilon_i) \sum_i \varepsilon_i g(i)^2, \text{ yielding}$$

$$\mathrm{Var}\,(AF_h(n)) \leq \sqrt{2}\varepsilon\sigma^2 \int G_h(t)^2 dt = \frac{\varepsilon}{\sqrt{2\pi}h}\sigma^2. \quad \square$$

There are many versions of AF_h, all yielding an asymptotic estimate equivalent to the one in Theorem 2.3: the famous median filter [46], an inf-sup filter on segments centered at $\mathbf{x}$ [21], and the clever numerical implementation of the mean curvature equation in [64]. So all of those filters have in common the good preservation of edges, but they perform poorly on flat regions and are worse there than a Gaussian blur. This fact derives from the comparison of the noise reduction estimates of Theorems 2.1 and 2.4 and is experimentally patent in Figure 2.

2.3. Total Variation. The total variation minimization was introduced by Rudin and Osher [78]. The original image u is supposed to have a simple geometric description, namely, a set of connected sets, the objects, along with their smooth contours, or edges. The image is smooth inside the objects but with jumps across the boundaries. The functional space modeling these properties is $BV(\Omega)$, the space of integrable functions with finite total variation $TV_\Omega(u) = \int |Du|$, where Du is assumed to be a Radon measure. Given a noisy image $v(\mathbf{x})$, the above-mentioned authors proposed to recover the original image $u(\mathbf{x})$ as the solution of the constrained minimization problem

$$(2.1) \qquad\qquad\qquad \arg\min_u TV_\Omega(u),$$

subject to the noise constraints

$$\int_\Omega (u(\mathbf{x}) - v(\mathbf{x}))d\mathbf{x} = 0 \quad \text{and} \quad \int_\Omega |u(\mathbf{x}) - v(\mathbf{x})|^2 d\mathbf{x} = \sigma^2.$$

The solution u must be as regular as possible in the sense of the total variation, while the difference $v(\mathbf{x}) - u(\mathbf{x})$ is treated as an error, with a prescribed energy. The constraints prescribe the right mean and variance to $u - v$ but do not ensure that it is similar to a noise (see a thorough discussion in [65]). The preceding problem is naturally linked to the unconstrained problem

$$(2.2) \qquad\qquad \arg\min_u TV_\Omega(u) + \lambda \int_\Omega |v(\mathbf{x}) - u(\mathbf{x})|^2 d\mathbf{x}$$

for a given Lagrange multiplier λ. The above functional is strictly convex and lower semicontinuous with respect to the weak-star topology of BV. Therefore the minimum exists, is unique, and is computable (see, e.g., [22]). The parameter λ controls the tradeoff between the regularity and fidelity terms. As λ gets smaller the weight of the regularity term increases. Therefore λ is related to the degree of filtering of the solution of the minimization problem. Let us denote by $TVF_\lambda(v)$ the solution of

problem (2.2) for a given value of λ. The Euler–Lagrange equation associated with the minimization problem is given by

$$(u(\mathbf{x}) - v(\mathbf{x})) - \frac{1}{2\lambda} curv(u)(\mathbf{x}) = 0$$

(see [78]). Thus, we have the following theorem.

THEOREM 2.5. *The method noise of the total variation minimization* (2.2) *is*

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

As in the anisotropic case, straight edges are maintained because of their small curvature. However, details and texture can be oversmoothed if λ is too small, as is shown in Figure 2.

2.4. Neighborhood Filters. The previous filters are based on a notion of spatial neighborhood or proximity. *Neighborhood filters* instead take into account grayscale values to define neighboring pixels. In the simplest and more extreme case, the denoised value at pixel i is an average of values at pixels which have a grayscale value close to $u(i)$. The grayscale neighborhood is therefore

$$B(i, h) = \{j \in I \mid u(i) - h < u(j) < u(i) + h\}.$$

This is a fully nonlocal algorithm, since pixels belonging to the whole image are used for the estimation at pixel i. This algorithm can be written in a more continuous form,

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

where $\Omega \subset \mathbb{R}^2$ is an open and bounded set, and $C(\mathbf{x}) = \int_\Omega e^{-\frac{|u(\mathbf{y}) - u(\mathbf{x})|^2}{h^2}} d\mathbf{y}$ is the normalization factor.

The Yaroslavsky neighborhood filters [97, 95] consider mixed neighborhoods $B(i, h)$ $\cap B_\rho(i)$, where $B_\rho(i)$ is a ball of center i and radius ρ. So the method takes an average of the values of pixels which are both close in grayscale and spatial distance. This filter can be easily written in a continuous form as

$$YNF_{h,\rho}(\mathbf{x}) = \frac{1}{C(\mathbf{x})} \int_{B_\rho(\mathbf{x})} u(\mathbf{y}) e^{-\frac{|u(\mathbf{y}) - u(\mathbf{x})|^2}{h^2}} d\mathbf{y},$$

where $C(\mathbf{x}) = \int_{B_\rho(\mathbf{x})} e^{-\frac{|u(\mathbf{y}) - u(\mathbf{x})|^2}{h^2}} d\mathbf{y}$ is the normalization factor. More recent versions, namely, the *SUSAN filter* [83] and the *bilateral filter* [86], weigh the distance to the reference pixel $\mathbf{x}$ instead of considering a fixed spatial neighborhood,

In the next theorem we compute the asymptotic expansion of the Yaroslavky neighborhood filter when $\rho, h \to 0$.

THEOREM 2.6. *Suppose* $u \in C^2(\Omega)$, *and let* $\rho, h, \alpha > 0$ *such that* $\rho, h \to 0$ *and* $h = O(\rho^\alpha)$. *Let us consider the continuous function* $\tilde{g}$ *defined by* $\tilde{g}(t) = \frac{1}{3} \frac{te^{-t^2}}{E(t)}$, *for* $t \neq 0$, $\tilde{g}(0) = \frac{1}{6}$, *where* $E(t) = 2 \int_0^t e^{-s^2} ds$. *Let* $\tilde{f}$ *be the continuous function defined by* $\tilde{f}(t) = 3\tilde{g}(t) + \frac{3\tilde{g}(t)}{t^2} - \frac{1}{2t^2}$, $\tilde{f}(0) = \frac{1}{6}$. *Then, for* $\boldsymbol{x} \in \Omega$,
   1. *if* $\alpha < 1$, $YNF_{h,\rho} u(\boldsymbol{x}) - u(\boldsymbol{x}) \simeq \frac{\triangle u(\boldsymbol{x})}{6} \rho^2$;
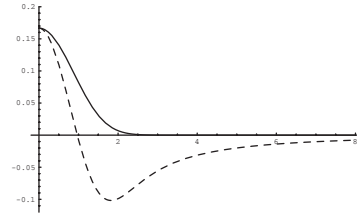
Fig. 3 *Magnitude of the tangent diffusion (continuous line) and normal diffusion (dashed line) of Theorem 2.6 in the case that $\rho = h$.*

2. *if $\alpha = 1$,*
   (2.3)
   $$YNF_{h,\rho}u(\boldsymbol{x}) - u(\boldsymbol{x}) \simeq \left[\tilde{g}\left(\frac{\rho}{h} |Du(\boldsymbol{x})| \right) \ u_{\xi\xi}(\boldsymbol{x}) + \tilde{f}\left(\frac{\rho}{h} |Du(\boldsymbol{x})| \right) \ u_{\eta\eta}(\boldsymbol{x}) \right] \rho^2;$$

3. *if $1 < \alpha < \frac{3}{2}$,*

   $$YNF_{h,\rho}u(\boldsymbol{x}) - u(\boldsymbol{x}) \simeq \tilde{g}\left(\rho^{1-\alpha}|Du(\boldsymbol{x})| \right) [\ u_{\xi\xi}(\boldsymbol{x}) + 3u_{\eta\eta}(\boldsymbol{x})] \ \rho^2.$$

According to Theorem 2.6 the Yaroslavsky neighborhood filter acts as an evolution PDE with two terms. The first term is proportional to the second derivative of u in the direction ξ, which is tangent to the level line passing through **x**. The second term is proportional to the second derivative of u in the direction η which is orthogonal to the level line passing through **x**. The evolution equations $u_t = c_1 u_{\xi\xi}$ and $u_t = c_2 u_{\eta\eta}$ act as filtering or enhancing models depending on the signs of c_1 and c_2. Following the previous theorem, we can distinguish three cases, depending on the values of h and ρ.

First, if h is much larger than ρ, both second derivatives are weighted by the same positive constant. In that case, the addition of both terms is equivalent to the Laplacian of u, Δu, and we get back to Gaussian filtering.

Second, if h and ρ have the same order of magnitude, the neighborhood filter behaves as a filtering/enhancing algorithm. The weighting coefficient of the tangent diffusion, $u_{\xi\xi}$, is given by $\tilde{g}(\frac{\rho}{h}|Du|)$. The function $\tilde{g}$ is positive and decreasing. Thus, there is always diffusion in that direction. The weight of the normal diffusion, $u_{\eta\eta}$, is given by $\tilde{f}(\frac{\rho}{h}|Du|)$. As the function $\tilde{f}$ takes positive and negative values (see Figure 3), the filter behaves as a filtering/enhancing algorithm in the normal direction and depending on $|Du|$. If $\tilde{B}$ denotes the zero of $\tilde{f}$, then a filtering model is applied wherever $|Du| < \tilde{B}\frac{h}{\rho}$ and an enhancing strategy wherever $|Du| > \tilde{B}\frac{h}{\rho}$. The intensity of the filtering in the tangent diffusion and the enhancing in the normal diffusion tend to zero when the gradient tends to infinity. Thus, points with a very large gradient are not altered. In this case the neighborhood filter asymptotically behaves as the Perona–Malik equation [72] (see [16] for more details on this comparison). Finally, if ρ is much larger than h, the value $\frac{\rho}{h}$ tends to infinity, and if the gradient of the image is bounded, then the filtering magnitude $\tilde{g}(\frac{\rho}{h}|Du|)$ tends to zero. Thus, the original image is hardly altered.

3. Frequency Domain Filters. Let u be the original image defined on the grid I. The image is supposed to be modified by the addition of a signal-independent white noise N. N is a random process where $N(i)$ are i.i.d. with zero mean and have constant

variance σ^2. The resulting noisy process depends on the random noise component, and therefore it is modeled as a random field V,

$$(3.1) \qquad\qquad V(i) = u(i) + N(i).$$

Given a noise observation $n(i)$, $v(i)$ denotes the observed noisy image,

$$(3.2) \qquad\qquad v(i) = u(i) + n(i).$$

Let $\mathcal{B} = \{g_\alpha\}_{\alpha \in A}$ be an orthonormal basis of $\mathbb{R}^{|I|}$. The noisy process is transformed as

$$(3.3) \qquad\qquad V_\mathcal{B}(\alpha) = u_\mathcal{B}(\alpha) + N_\mathcal{B}(\alpha),$$

where $V_\mathcal{B}(\alpha) = \langle V, g_\alpha \rangle$, $u_\mathcal{B}(\alpha) = \langle u, g_\alpha \rangle$, $N_\mathcal{B}(\alpha) = \langle N, g_\alpha \rangle$ are the scalar products of V, u, and N with $g_\alpha \in \mathcal{B}$. The noise coefficients $N_\mathcal{B}(\alpha)$ remain uncorrelated and with zero mean, but the variances are multiplied by $\|g_\alpha\|^2$:

$$\begin{aligned} E[N_\mathcal{B}(\alpha)N_\mathcal{B}(\beta)] &= \sum_{m,n \in I} g_\alpha(m)g_\beta(n)E[N(m)N(n)] \\ &= \langle g_\alpha, g_\beta \rangle \sigma^2 = \sigma^2 \|g_\alpha\|^2 \delta[\alpha - \beta]. \end{aligned}$$

Frequency domain filters are applied independently to every transform coefficient $V_\mathcal{B}(\alpha)$, and then the solution is estimated by the inverse transform of the new coefficients. Noisy coefficients $V_\mathcal{B}(\alpha)$ are modified to $a(\alpha)V_\mathcal{B}(\alpha)$. This is a nonlinear algorithm because $a(\alpha)$ depends on the value $V_\mathcal{B}(\alpha)$. The inverse transform yields the estimate

$$(3.4) \qquad\qquad \hat{U} = DV = \sum_{\alpha \in A} a(\alpha)\ V_\mathcal{B}(\alpha)\ g_\alpha.$$

D is also called a *diagonal operator*. Let us look for the frequency domain filter D which minimizes a certain estimation error. This error is based on the squared Euclidean distance, and it is averaged over the noise distribution.

DEFINITION 3.1. *Let u be the original image, N be a white noise, and $V = u + N$. Let D be a frequency domain filter. Define the* risk of D *as*

$$(3.5) \qquad\qquad r(D, u) = E\{\|u - DV\|^2\},$$

where the expectation is taken over the noise distribution.

The following easy theorem gives the diagonal operator D_{inf} that minimizes the risk,

$$D_{inf} = \arg\min_D r(D, u).$$

THEOREM 3.2. *The operator D_{inf} which minimizes the risk is given by the family $\{a(\alpha)\}_\alpha$, where*

$$(3.6) \qquad\qquad a(\alpha) = \frac{|u_\mathcal{B}(\alpha)|^2}{|u_\mathcal{B}(\alpha)|^2 + \|g_\alpha\|^2 \sigma^2},$$

and the corresponding risk is

$$(3.7) \qquad\qquad r_{inf}\,(u) = \sum_{s \in S} \|g_\alpha\|^4 \frac{|u_\mathcal{B}(\alpha)|^2 \sigma^2}{|u_\mathcal{B}(\alpha)|^2 + \|g_\alpha\|^2 \sigma^2}.$$

The previous optimal operator attenuates all noisy coefficients in order to minimize the risk. If one restricts $a(\alpha)$ to be 0 or 1, one gets a projection operator. In that case, a subset of coefficients is kept, and the rest gets canceled. The projection operator that minimizes the risk $r(D, u)$ is obtained by the family $\{a(\alpha)\}_\alpha$, where

$$a(\alpha) = \begin{cases} 1 & |u_\mathcal{B}(\alpha)|^2 \geq \|g_\alpha\|^2 \sigma^2, \\ 0 & \text{otherwise,} \end{cases}$$

and the corresponding risk is

$$r_p(u) = \sum \|g_\alpha\|^2 \min(|u_\mathcal{B}(\alpha)|^2, \|g_\alpha\|^2 \sigma^2).$$

Note that both filters are ideal operators because they depend on the coefficients $u_\mathcal{B}(\alpha)$ of the original image, which are not known. We call, as classical, *Fourier–Wiener filter* the optimal operator (3.6) where $\mathcal{B}$ is a Fourier basis. This is an ideal filter, since it uses the (unknown) Fourier transform of the original image. By the use of the Fourier basis, global image characteristics may prevail over local ones and create spurious periodic patterns. To avoid this effect, the basis must take into account more local features, as the wavelet and local DCT transforms do. The search for the ideal basis associated with each image is still open. One way to adapt the basis to the image is to use a dictionary of bases, and to adapt the basis to the image by minimizing a variational criterion [59].

3.1. Local Adaptive Filters in Transform Domain. The local adaptive filters have been introduced by Yaroslavsky and Eden [95] and Yaroslavsky [94]. In this case, the noisy image is analyzed in a moving window, and in each position of the window its spectrum is computed and modified. Finally, an inverse transform is used to estimate only the signal value in the central pixel of the window.

Let $i \in I$ be a pixel and $W = W(i)$ a window centered in i. Then the DCT transform of W is computed and modified. The original image coefficients of W, $u_{\mathcal{B},W}(\alpha)$, are estimated, and the optimal attenuation of Theorem 3.2 is applied. Finally, only the center pixel of the restored window is used. This method is called the *empirical Wiener filter*. In order to approximate $u_{\mathcal{B},W}(\alpha)$, one can take averages on the additive noise model, that is,

$$E|V_{\mathcal{B},W}(\alpha)|^2 = |u_{\mathcal{B},W}(\alpha)|^2 + \sigma^2 \|g_\alpha\|^2.$$

Denoting by $\mu = \sigma \|g_\alpha\|$, the unknown original coefficients can be written as

$$|u_{\mathcal{B},W}(\alpha)|^2 = E|V_{\mathcal{B},W}(\alpha)|^2 - \mu^2.$$

The observed coefficients $|v_{\mathcal{B},W}(\alpha)|^2$ are used to approximate $E|V_{\mathcal{B},W}(\alpha)|^2$, and the estimated original coefficients are replaced in the optimal attenuation, leading to the family $\{a(\alpha)\}_\alpha$, where

$$a(\alpha) = \max\left\{0, \frac{|v_{\mathcal{B},W}(\alpha)|^2 - \mu^2}{|v_{\mathcal{B},W}(\alpha)|^2}\right\}.$$

Denote by $EWF_\mu(i)$ the filter given by the previous family of coefficients. The method noise of the $EWF_\mu(i)$ is easily computed, as proved in the following theorem.
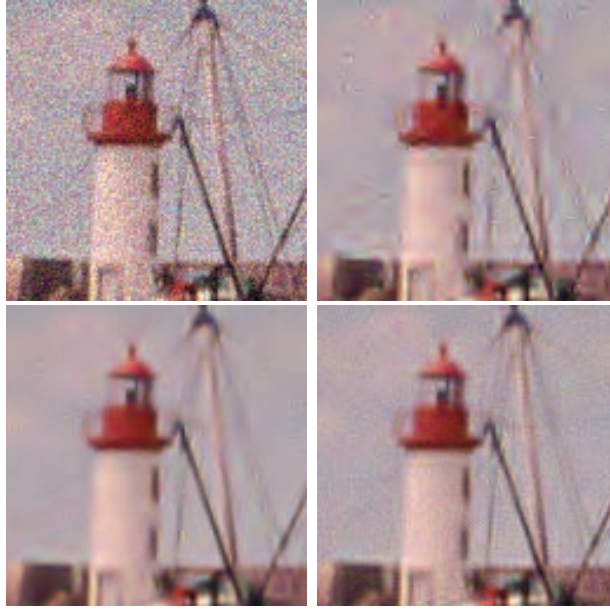
Fig. 4 *Denoising experiment on a natural image. From left to right and from top to bottom: noisy image (standard deviation $\sigma = 20$), the wavelet hard thresholding (HWT, $\mu = 3\sigma$), translation invariant hard wavelet thresholding (TIHWT, $\mu = 3\sigma$), and the DCT empirical Wiener filter ($W = 15 \times 15$ pixels, $\mu = \sigma/2$). Parameters have been set for each algorithm in such a way the removed energy is similar to the added noise energy, σ^2, getting the best result in terms of Euclidean distance to the original image.*

THEOREM 3.3. *Let u be an image defined in a grid I, and let $i \in I$ be a pixel. Let $W = W(i)$ be a window centered in the pixel i. Then the method noise of the $EWF_\mu(i)$ is given by*

$$u(i) - EWF_\mu(i) = \sum_{\alpha \in \Lambda} v_{\mathcal{B},\mathcal{W}}(\alpha)\ g_\alpha(i) + \sum_{\alpha \notin \Lambda} \frac{\mu^2}{|v_{\mathcal{B},\mathcal{W}}(\alpha)|^2} v_{\mathcal{B},\mathcal{W}}(\alpha)\ g_\alpha(i),$$

where $\Lambda = \{\alpha \mid |v_{\mathcal{B},\mathcal{W}}(\alpha)| < \mu\}$.

The presence of an edge in the window W will produce a large number of large coefficients, and, as a consequence, the cancelation of these coefficients will produce oscillations. Then spurious cosines will also appear in the image under the form of checkerboard patterns; see Figure 4.

3.2. Wavelet Thresholding. Let $\mathcal{B} = \{g_\alpha\}_{\alpha \in A}$ be an orthonormal basis of wavelets [60]. Let us discuss two procedures modifying the noisy coefficients, called *wavelet thresholding methods* (Donoho and Johnstone [33]). The first procedure is a projection operator which approximates the ideal projection (3.6). It is called a *hard thresholding* and cancels coefficients smaller than a certain threshold μ,

$$a(\alpha) = \begin{cases} 1 & |v_{\mathcal{B}}(\alpha)| > \mu, \\ 0 & \text{otherwise.} \end{cases}$$

Let us denote this operator by $HWT_\mu(v)$. This procedure is based on the idea that the image is represented with large wavelet coefficients, which are kept, whereas the noise

is distributed across small coefficients, which are canceled. The performance of the method depends on the capacity of approximating u by a small set of large coefficients. Wavelets are, for example, an adapted representation for smooth functions.

THEOREM 3.4. *Let u be an image defined in a grid I. The method noise of a hard thresholding $HWT_\mu(u)$ is*

$$u - HWT_\mu(u) = \sum_{\{\alpha \,||\, u_\mathcal{B}(\alpha)| < \mu\}} u_\mathcal{B}(\alpha) g_\alpha.$$

Unfortunately, edges cause many small wavelet coefficients, which are lower than the threshold. The cancelation of these wavelet coefficients causes small oscillations near the edges, i.e., a Gibbs-like phenomenon. Spurious wavelets can also be seen in the restored image due to the cancelation of small coefficients; see Figure 4. This artifact will be called *wavelet outliers*, as was introduced in [34]. Donoho [32] showed that these effects can be partially avoided with the use of a soft thresholding,

$$a(\alpha) = \begin{cases} \dfrac{v_\mathcal{B}(\alpha) - sgn(v_\mathcal{B}(\alpha))\mu}{v_\mathcal{B}(\alpha)}, & |v_\mathcal{B}(\alpha)| \geq \mu, \\ 0 & \text{otherwise,} \end{cases}$$

which will be denoted by $SWT\mu(v)$. The continuity of the soft thresholding operator better preserves the structure of the wavelet coefficients, reducing the oscillations near discontinuities. Note that a soft thresholding attenuates all coefficients in order to reduce the noise, as an ideal operator does. As we shall see at the end of this paper, the L^2 norm of the method noise is lessened when replacing the hard threshold by a soft threshold. See Figure 11 for a comparison of both method noises.

THEOREM 3.5. *Let u be an image defined in a grid I. The method noise of a soft thresholding $SWT_\mu(u)$ is*

$$u - SWT_\mu(u) = \sum_{\{\alpha \,||\, u_\mathcal{B}(\alpha)| < \mu\}} u_\mathcal{B}(\alpha) g_\alpha + \mu \sum_{\{\alpha \,||\, u_\mathcal{B}(\alpha)| > \mu\}} sgn(u_\mathcal{B}(\alpha))\ g_\alpha.$$

A simple example can show how to fix the threshold μ. Suppose the original image u is zero; then $v_\mathcal{B}(\alpha) = n_\mathcal{B}(\alpha)$, and therefore the threshold μ must be taken over the maximum of noise coefficients to ensure their suppression and the recovery of the original image. It can be shown that the maximum amplitude of a white noise has a high probability of being smaller than $\sigma\sqrt{2\log|I|}$. It can be proved that the risk of a wavelet thresholding with the threshold $\mu = \sigma\sqrt{2\log|I|}$ is near the risk r_p of the optimal projection; see [33, 60].

THEOREM 3.6. *The risk $r_t(u)$ of a hard or soft thresholding with the threshold $\mu = \sigma\sqrt{2\log|I|}$ is such that for all $|I| \geq 4$*

$$(3.8) \qquad\qquad r_t(u) \leq (2\log|I| + 1)(\sigma^2 + r_p(u)).$$

The factor $2\log|I|$ is optimal among all the diagonal operators in $\mathcal{B}$, that is,

$$(3.9) \qquad\qquad \lim_{|I| \to \infty} \inf_{D \in D_\mathcal{B}} \sup_{u \in \mathbb{R}^{|I|}} \frac{E\{\|u - DV\|^2\}}{\sigma^2 + r_p(u)} \frac{1}{2\log|I|} = 1.$$

In practice the optimal threshold μ is very high and cancels too many coefficients not produced by the noise. A threshold lower than the optimal is used in the experiments and produces much better results; see Figure 4. For a hard thresholding the threshold is fixed to $3 * \sigma$. For a soft thresholding this threshold still is too high; it is better fixed at $\frac{3}{2}\sigma$.

3.3. Translation Invariant Wavelet Thresholding. Coifman and Donoho [24] improved the wavelet thresholding methods by averaging the estimation of all translations of the degraded signal. Calling $v^p(i)$ the translated signal $v(i-p)$, the wavelet coefficients of the original and translated signals can be very different, and they are not related by a simple translation or permutation,

$$v_{\mathcal{B}}^p(\alpha) = \langle v(n-p), g_\alpha(n)\rangle = \langle v(n), g_\alpha(n+p)\rangle.$$

The vectors $g_\alpha(n+p)$ are not in general in the basis $\mathcal{B} = \{g_\alpha\}_{\alpha \in A}$, and therefore the estimation of the translated signal is not related to the estimation of v. This new algorithm yields an estimate $\hat{u}^p$ for every translated v^p of the original image,

$$(3.10) \qquad \hat{u}^p = Dv^p = \sum_{\alpha \in A} a(\alpha)v_{\mathcal{B}}^p(\alpha)g_\alpha.$$

The translation invariant thresholding is obtained by averaging all these estimators after a translation in the inverse sense,

$$(3.11) \qquad \frac{1}{|I|}\sum_{p \in I}\hat{u}^p(i+p),$$

and will be denoted by $TIHWT$ and $TISWT$, respectively, for the hard and soft thresholding. The Gibbs effect is considerably reduced by the translation invariant wavelet thresholding (see Figure 4), because the average of different estimations of the image reduces the oscillations. This is therefore the version we shall use in the comparison section.

4. NL-Means Algorithm. The local smoothing methods and the frequency domain filters aim at a noise reduction and at a reconstruction of the main geometrical configurations but not at the preservation of the fine structure, details, and texture. Due to the regularity assumptions on the original image of previous methods, details and fine structures are smoothed out because they behave in all functional aspects as noise. The NL-means algorithm we shall now discuss tries to take advantage of the high degree of redundancy of any natural image. By this, we simply mean that every small window in a natural image has many similar windows in the same image. This fact is patent for windows close by, at one pixel distance, and in that case we go back to a local regularity assumption. Now in a very general sense inspired by the neighborhood filters, one can define as "neighborhood of a pixel i" any set of pixels j in the image such that a window around j looks like a window around i. All pixels in that neighborhood can be used for predicting the value at i, as was first shown in [38] for 2D images. This first work has inspired many variants for the restoration of various digital objects, in particular 3D surfaces [82]. The fact that such a self-similarity exists is a regularity assumption, actually more general and more accurate than all regularity assumptions we have considered in section 2. It also generalizes a periodicity assumption of the image.

Let v be the noisy image observation defined on a bounded domain $\Omega \subset \mathbb{R}^2$, and let $\mathbf{x} \in \Omega$. The NL-means algorithm estimates the value of $\mathbf{x}$ as an average of the values of all the pixels whose Gaussian neighborhood looks like the neighborhood of $\mathbf{x}$,

$$NL(v)(\mathbf{x}) = \frac{1}{C(\mathbf{x})}\int_\Omega e^{-\frac{(G_a*|v(\mathbf{x}+.)-v(\mathbf{y}+.)|^2)(0)}{h^2}}\, v(\mathbf{y})\, d\mathbf{y},$$

where G_a is a Gaussian kernel with standard deviation a, h acts as a filtering parameter, and $C(\mathbf{x}) = \int_\Omega e^{-\frac{(G_a * |v(\mathbf{x}+.) - v(\mathbf{z}+.)|^2)(0)}{h^2}} d\mathbf{z}$ is the normalizing factor. We recall that

$$(G_a * |v(\mathbf{x}+.) - v(\mathbf{y}+.)|^2)(0) = \int_{\mathbb{R}^2} G_a(\mathbf{t})|v(\mathbf{x}+\mathbf{t}) - v(\mathbf{y}+\mathbf{t})|^2 d\mathbf{t}.$$

Since we are considering images defined on a discrete grid I, we shall give a discrete description of the NL-means algorithm and some consistency results. This simple and generic algorithm and its application to the improvement of the performance of digital cameras are the object of a European patent application [13].

4.1. Description. Given a discrete noisy image $v = \{v(i) \mid i \in I\}$, the estimated value $NL(v)(i)$ is computed as a weighted average of all the pixels in the image,

$$NL(v)(i) = \sum_{j \in I} w(i,j)v(j),$$

where the weights $\{w(i,j)\}_j$ depend on the similarity between the pixels i and j and satisfy the usual conditions $0 \le w(i,j) \le 1$ and $\sum_j w(i,j) = 1$.

In order to compute the similarity between the image pixels, we define a *neighborhood system* on I.

DEFINITION 4.1 (neighborhoods). *A neighborhood system on I is a family $\mathcal{N} = \{\mathcal{N}_i\}_{i \in I}$ of subsets of I such that for all $i \in I$,*
   (i) $i \in \mathcal{N}_i$,
   (ii) $j \in \mathcal{N}_i \Rightarrow i \in \mathcal{N}_j$.
The subset $\mathcal{N}_i$ is called the neighborhood or the similarity window of i. We set $\tilde{\mathcal{N}}_i = \mathcal{N}_i \setminus \{i\}$.

The similarity windows can have different sizes and shapes to better adapt to the image. For simplicity we will use square windows of fixed size. The restriction of v to a neighborhood $\mathcal{N}_i$ will be denoted by $v(\mathcal{N}_i)$:

$$v(\mathcal{N}_i) = (v(j), j \in \mathcal{N}_i).$$

The similarity between two pixels i and j will depend on the similarity of the intensity grayscale vectors $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$. The pixels with a similar grayscale neighborhood to $v(\mathcal{N}_i)$ will have larger weights on average; see Figure 5.

In order to compute the similarity of the intensity grayscale vectors $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$, one can compute a Gaussian weighted Euclidean distance, $\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_{2,a}^2$. Efros and Leung [38] showed that the L^2 distance is a reliable measure for the comparison of image windows in a texture patch. This measure is quite adapted to an additive white noise, which alters uniformly the distance between windows. Indeed,

$$E\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_{2,a}^2 = \|u(\mathcal{N}_i) - u(\mathcal{N}_j)\|_{2,a}^2 + 2\sigma^2,$$

where u and v are, respectively, the original and noisy images and σ^2 is the noise variance. This equality shows that, in expectation, the Euclidean distance preserves the order of similarity between pixels. So the most similar pixels to i in v also are expected to be the most similar pixels to i in u. The weights associated with the quadratic distances are defined by

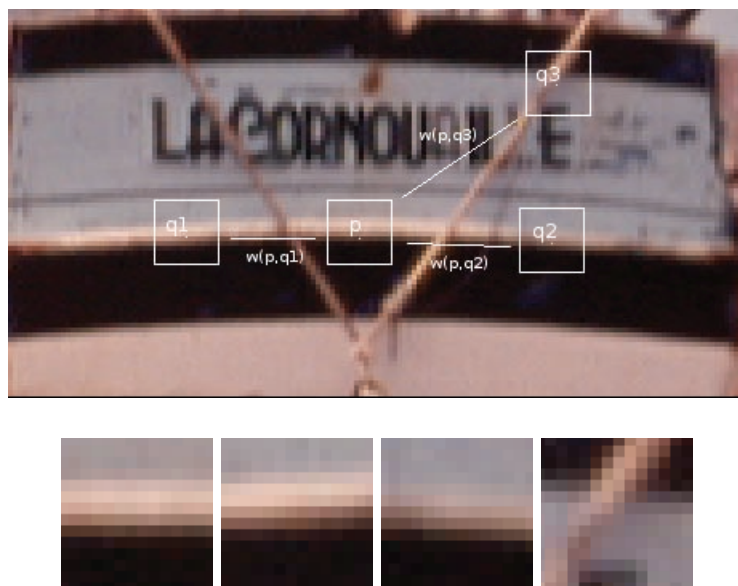$$w(i,j) = \frac{1}{Z(i)} e^{-\frac{\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_{2,a}^2}{h^2}},$$

Fig. 5 *q1 and q2 have a large weight because their similarity windows are similar to that of p. On the other hand the weight $w(p, q3)$ is much smaller because the intensity gray values in the similarity windows are very different.*

where $Z(i)$ is the normalizing factor $Z(i) = \sum_j e^{-\frac{\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_{2,a}^2}{h^2}}$ and the parameter h controls the decay of the exponential function, and therefore the decay of the weights, as a function of the Euclidean distances.

4.2. Closely Related Attempts. Two methods have independently attempted to take advantage of an image model learned from the image itself. The work by Weissman et al. [89] has led to the proposition of a "universal denoiser" for digital images. The authors prove that this denoiser is universal in the sense "of asymptotically achieving, without access to any information on the statistics of the clean signal, the same performance as the best denoiser that does have access to this information." In [70] the authors present an implementation valid for binary images with an impulse noise, with excellent results.

Awate and Whitaker [6] have also proposed a method whose principles stand close to the NL-means algorithm, since, the method involves comparison between subwindows to estimate a restored value. The objective of the algorithm is to denoise the image by decreasing the randomness of the image.

4.3. A Consistency Theorem for NL-Means. The NL-means algorithm is intuitively consistent under stationarity conditions, saying that one can find many samples of every image detail. In fact, we shall be assuming that the image is a fairly general stationary random process. Under these assumptions, for every pixel i, the NL-means algorithm converges to the conditional expectation of i knowing its neighborhood. In the case of an additive or multiplicative white noise model, this expectation is in fact the solution to a minimization problem.

Let X and Y denote two random vectors with values on $\mathbb{R}^p$ and $\mathbb{R}$, respectively. Let f_X, f_Y denote the probability distribution functions of X, Y, and let f_{XY} denote

the joint probability distribution function of X and Y. Let us recall briefly the definition of the conditional expectation.

DEFINITION 4.2.

(i) *Define the probability distribution function of Y conditioned to X as*

$$f(y \mid x) = \begin{cases} \frac{f_{XY}(x,y)}{f_X(x)} & \text{if } f_X(x) > 0, \\ 0 & \text{otherwise} \end{cases}$$

for all $x \in \mathbb{R}^p$ and $y \in \mathbb{R}$.

(ii) *Define the* conditional expectation *of Y given $\{X = x\}$ as the expectation with respect to the conditional distribution $f(y \mid x)$,*

$$E[Y \mid X = x] = \int y\, f(y \mid x)\, dy$$

for all $x \in \mathbb{R}^p$.

The conditional expectation is a function of X, and therefore a new random variable $g(X)$, which is denoted by $E[Y \mid X]$. Now let V be a random field and $\mathcal{N}$ a neighborhood system on I. Let Z denote the sequence of random variables $Z_i = \{Y_i, X_i\}_{i \in I}$, where $Y_i = V(i)$ is real valued and $X_i = V(\tilde{\mathcal{N}}_i)$ is $\mathbb{R}^p$ valued. Recall that $\tilde{\mathcal{N}}_i = \mathcal{N}_i \setminus \{i\}$. Let us restrict Z to the n first elements $\{Y_i, X_i\}_{i=1}^n$. Let us define the function $r_n(x)$,

$$(4.1) \qquad\qquad r_n(x) = R_n(x)/\hat{f}_n(x),$$

where

$$(4.2) \quad \hat{f}_n(x) = \frac{1}{nh^p} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right), \qquad R_n(x) = \frac{1}{nh^p} \sum_{i=1}^n \phi(Y_i) K\left(\frac{X_i - x}{h}\right),$$

ϕ is an integrable real-valued function, K is a nonnegative kernel, and $x \in \mathbb{R}^p$.

Let X and Y be distributed as X_1 and Y_1. Under this form the NL-means algorithm can be seen as an instance for the exponential operator of the Nadaraya–Watson estimator [67, 88]. This is an estimator of the conditional expectation $r(x) = E[\phi(Y) \mid X = x]$. Some definitions are needed for the statement of the main result.

DEFINITION 4.3. *A stochastic process $\{Z_t \mid t = 1, 2, \ldots\}$, with Z_t defined on some probability space $(\Omega, \mathcal{A}, \mathcal{P})$, is said to be (strict-sense) stationary if for any finite partition $\{t_1, t_2, \ldots, t_n\}$ the joint distributions $F_{t_1, t_2, \ldots, t_n}(x_1, x_2, \ldots, x_n)$ are the same as the joint distributions $F_{t_1 + \tau, t_2 + \tau, \ldots, t_n + \tau}(x_1, x_2, \ldots, x_n)$ for any $\tau \in \mathbb{N}$.*

In the case of images, this stationary condition amounts to saying that as the size of the image grows, we are able to find in the image many similar patches for all the details of the image. This is a crucial point in understanding the performance of the NL-means algorithm. The following mixing definition is a rather technical condition. In the case of images, it amounts to saying that regions become more independent as their distance increases. This is intuitively true for natural images.

DEFINITION 4.4. *Let Z be a stochastic and stationary process $\{Z_t \mid t = 1, 2, \ldots, n\}$, and, for $m < n$, let $\mathbb{F}_m^n$ be the σ-field induced in Ω by the random variables Z_j, $m \le j \le n$. Then the sequence Z is said to be β-mixing if for every $A \in \mathbb{F}_1^k$ and every $B \in \mathbb{F}_{k+n}^\infty$*

$$|P(A \cap B) - P(A)P(B)| \le \beta(n), \quad \text{with } \beta(n) \to 0, \ as\ n \to \infty.$$

The following theorem establishes the convergence of r_n to r; see Roussas [77]. The theorem is established under the stationary and mixing hypothesis of $\{Y_i, X_i\}_{i=1}^{\infty}$ and asymptotic conditions on the decay of ϕ, $\beta(n)$, and K. This set of conditions will be denoted by H.

THEOREM 4.5 (conditional expectation theorem). *Let* $Z_j = \{X_j, Y_j\}$ *for* $j = 1, 2, \ldots$ *be a strictly stationary and mixing process. For* $i \in I$, *let* X *and* Y *be distributed as* X_i *and* Y_i. *Let* J *be a compact subset* $J \subset \mathbb{R}^p$ *such that*

$$\inf\{f_X(x); x \in J\} > 0.$$

Then, under hypothesis H,

$$\sup[\psi_n|r_n(x) - r(x)|; x \in J] \to 0 \quad \text{almost surely},$$

where ψ_n *are positive norming factors.*

Let v be the observed noisy image, and let i be a pixel. Taking for ϕ the identity, we see that $r_n(v(\tilde{\mathcal{N}}_i))$ converges to $E[V(i) \mid V(\tilde{\mathcal{N}}_i) = v(\tilde{\mathcal{N}}_i)]$ under stationary and mixing conditions of the sequence $\{V(i), V(\tilde{\mathcal{N}}_i)\}_{i=1}^{\infty}$.

In the case where an additive or multiplicative white noise model is assumed, the next result shows that this conditional expectation is in fact the function of $V(\tilde{\mathcal{N}}_i)$ that minimizes the mean square error with the original field U.

THEOREM 4.6. *Let* V, U, N_1, *and* N_2 *be random fields on* I *such that* $V = U + N_1 + g(U)N_2$, *where* N_1 *and* N_2 *are independent white noises. Let* $\mathcal{N}$ *be a neighborhood system on* I. *Then we have the following:*

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

(ii) *The real value* $E[U(i) \mid V(\tilde{\mathcal{N}}_i) = x]$ *minimizes the mean square error*

$$(4.3) \qquad \min_{g^* \in \mathbb{R}} E[(U(i) - g^*)^2 \mid V(\tilde{\mathcal{N}}_i) = x]$$

for all $i \in I$ *and* $x \in \mathbb{R}^p$.

(iii) *The expected random variable* $E[U(i) \mid V(\tilde{\mathcal{N}}_i)]$ *is the function of* $V(\tilde{\mathcal{N}}_i)$ *that minimizes the mean square error*

$$(4.4) \qquad \min_{g} E[U(i) - g(V(\tilde{\mathcal{N}}_i))]^2.$$

Given a noisy image observation $v(i) = u(i) + n_1(i) + g(u(i))n_2(i)$, $i \in I$, where g is a real function and n_1 and n_2 are white noise realizations, the NL-means algorithm is the function of $v(\tilde{\mathcal{N}}_i)$ that minimizes the mean square error with the original image $u(i)$.

4.4. Experiments with NL-Means. The NL-means algorithm chooses for each pixel a different average configuration adapted to the image. As we explained in the previous sections, for a given pixel i, we take into account the similarity between the neighborhood configuration of i and all the pixels of the image. The similarity between pixels is measured as a decreasing function of the Euclidean distance of the similarity windows. Due to the fast decay of the exponential kernel, large Euclidean distances lead to nearly zero weights, acting as an automatic threshold. The decay of the exponential function, and therefore the decay of the weights, is controlled by the parameter h. Empirical experimentation shows that one can take a similarity window of size 7×7 or 9×9 for grayscale images and 5×5 or even 3×3 in color images with little noise. These window sizes have shown to be large enough to be
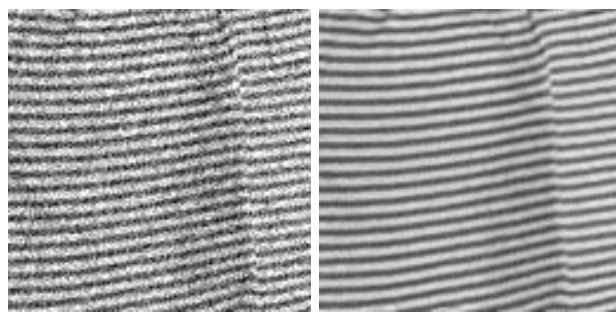
Fig. 6 *NL-means denoising experiment with a nearly periodic image. Left: Noisy image with standard deviation* 30. *Right: NL-means restored image.*
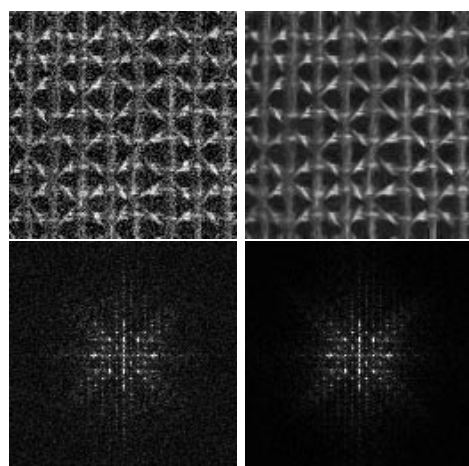

Fig. 7 *NL-means denoising experiment with a Brodatz texture image. Left: Noisy image with standard deviation* 30. *Right: NL-means restored image. The Fourier transform of the noisy and restored images shows how main features are preserved even at high frequencies.*

robust to noise and at the same time to be able to take care of the details and fine structure. Smaller windows are not robust enough to noise. Notice that in the limit case, one can take the window reduced to a single pixel i and therefore get back to the Yaroslavsky neighborhood filter. For computational aspects, in the following experiments the average is not performed in all the images. In practice, for each pixel p, we consider only a squared window centered in p and size 21×21 pixels.

Due to the nature of the algorithm, the most favorable case for the NL-means algorithm is the periodic case. In this situation, for every pixel i of the image one can find a large set of samples with a very similar configuration, leading to a noise reduction and a preservation of the original image; see Figure 6 for an example.

Another case which is ideally suitable for the application of the NL-means algorithm is the textural case. Texture images have a large redundancy. For a fixed configuration many similar samples can be found in the image. In Figure 7 one can see an example with a Brodatz texture. The Fourier transform of the noisy and restored images shows the ability of the algorithm to preserve the main features even in the case of high frequencies.

Fig. 8 *NL-means denoising experiment with a color image. Left: Noisy image with standard deviation* 20 *in every color component. Right: Restored image. NL-means algorithm has been applied with a* 3×3 *color comparison window, with a* 21×21 *pixels search window, and* $h = 16$.

The NL-means algorithm is not only able to restore periodic or texture images. Natural images also have enough redundancy to be restored. For example, in a flat zone, one can find many pixels lying in the same region and with similar configurations. In a straight or curved edge a complete line of pixels with a similar configuration is found. In addition, the redundancy of natural images allows us to find many similar configurations in far away pixels. See Figure 8.

5. Discussion and Comparison.

5.1. NL-Means as an Extension of Previous Methods. As was said before, the Gaussian convolution preserves only flat zones, while contours and fine structure are removed or blurred. Anisotropic filters instead preserve straight edges, but flat zones present many artifacts. One could think of combining these two methods to improve both results. A Gaussian convolution could be applied in flat zones, while an anisotropic filter could be applied on straight edges. Still, other types of filters should be designed to specifically restore corners or curved edges and texture. The NL-means algorithm seems to provide a feasible and rational method to automatically take the best of each mentioned algorithm, reducing for every possible geometric configuration the image method noise. Although we have not computed explicitly the image method noise, Figure 9 illustrates how the NL-means algorithm chooses in each case a weight configuration corresponding to one of the previously analyzed filters. In particular, according to this set of experiments, we can consider that the consistency results given in Theorems 2.1, 2.3, and 2.5 are all valid for this algorithm.

The NL-means algorithm is easily extended to the denoising of image sequences and video. The denoising algorithm involves indiscriminately pixels not belonging only to the same frame but also to all frames in the image. The algorithm favors pixels with a similar local configuration, as the similar configurations move, so do the weights. Thus, the algorithm is able to follow the similar configurations when they move without any explicit motion computation (see Figure 10).
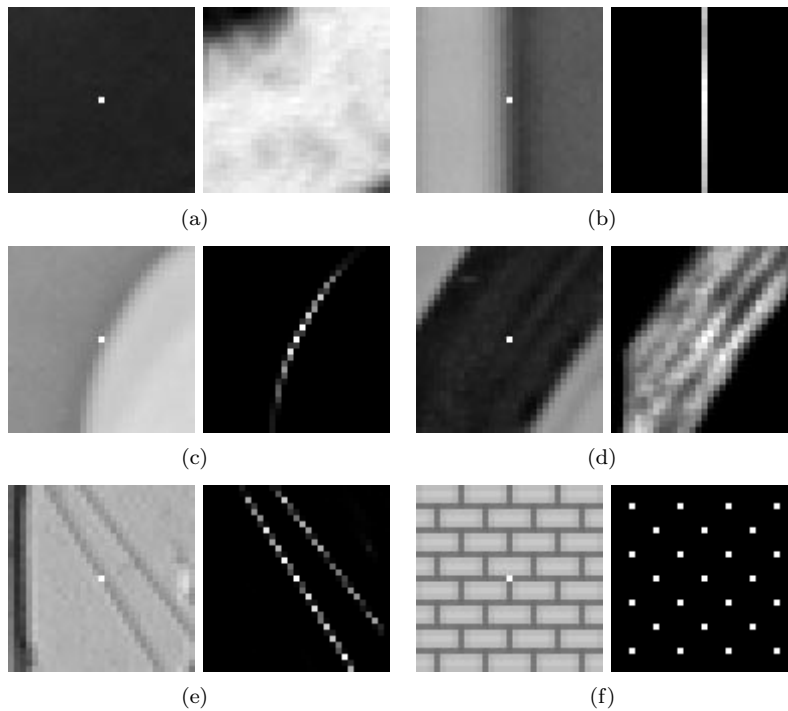
Fig. 9 *On the right-hand side of each pair, we display the weight distribution used to estimate the central pixel of the left image by the NL-means algorithm.* (a) *In flat zones, the weights are distributed as a convolution filter (as a Gaussian convolution).* (b) *In straight edges, the weights are distributed in the direction of the level line (as the mean curvature motion).* (c) *On curved edges, the weights favor pixels belonging to the same contour or level line, which is a strong improvement with respect to the mean curvature motion.* (d) *In a flat neighborhood, the weights are distributed in a grayscale neighborhood (as with a neighborhood filter).* *In the cases of* (e) *and* (f), *the weights are distributed across the more similar configurations, even though they are far away from the observed pixel. This shows a behavior similar to a nonlocal neighborhood filter or to an ideal Wiener filter.*

Classical movie denoising algorithms are motion compensated. The underlying idea is the existence of a "ground true" physical motion, which motion estimation algorithms should be able to estimate. Legitimate information should exist only along these physical trajectories. One of the major difficulties in motion estimation is the ambiguity of trajectories, the so-called *aperture problem*. The aperture problem, viewed as a general phenomenon in movies, can be positively interpreted in the following way: There are many pixels in the next or previous frames which can match the current pixel. Thus, it seems sound to use not just one trajectory, but rather *all pixels similar* to the current pixel across time and space (see [18] for more details on this discussion).

5.2. Comparison. In this section we shall compare the different algorithms based on four well-defined criteria: the *method noise*, the *noise-to-noise*, the *mean square error*, and the *visual quality* of the restored images. Note that every criterion measures a different aspect of the denoising method. It is easy to show that only one criterion is not enough to judge the restored image, and so one expects a good solution to have a high performance under the three criteria.
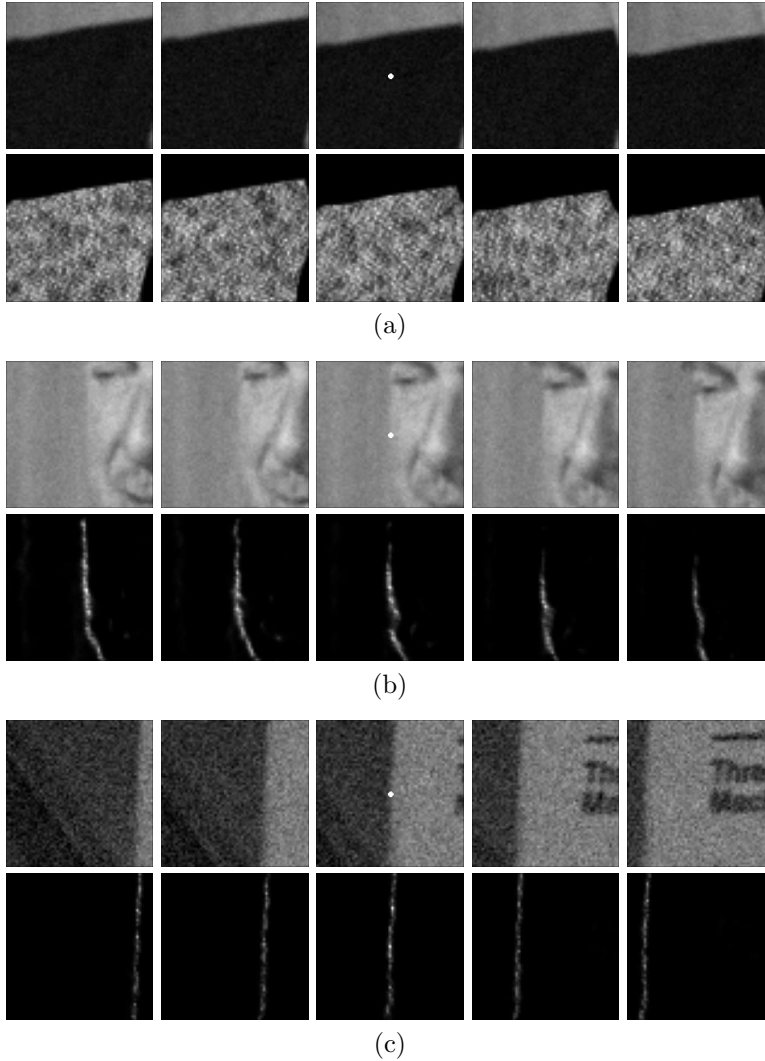
(a)

(b)

(c)

Fig. 10 *Weight distribution of NL-means applied to a movie. In (a), (b), and (c) the first row shows a five-frame image sequence. In the second row, the weight distribution used to estimate the central pixel (in white) of the middle frame is shown. The weights are equally distributed over the successive frames, including the current one. They actually involve all the candidates for the motion estimation instead of picking just one per frame. The aperture problem can be taken advantage of for a better denoising performance by involving more pixels in the average.*

5.2.1. Method Noise Comparison. In previous sections we have defined the method noise and computed it for the different algorithms. Remember that the denoising algorithm is applied on the original (slightly noisy) image. A filtering parameter, depending mainly on the standard deviation of the noise, must be fixed for most algorithms. Let us fix $\sigma = 2.5$: we can suppose that most digital images are affected by this amount of noise, since it is not visually noticeable.

The method noise tells us which geometrical features or details are preserved by the denoising process and which are eliminated. In order to preserve as many features
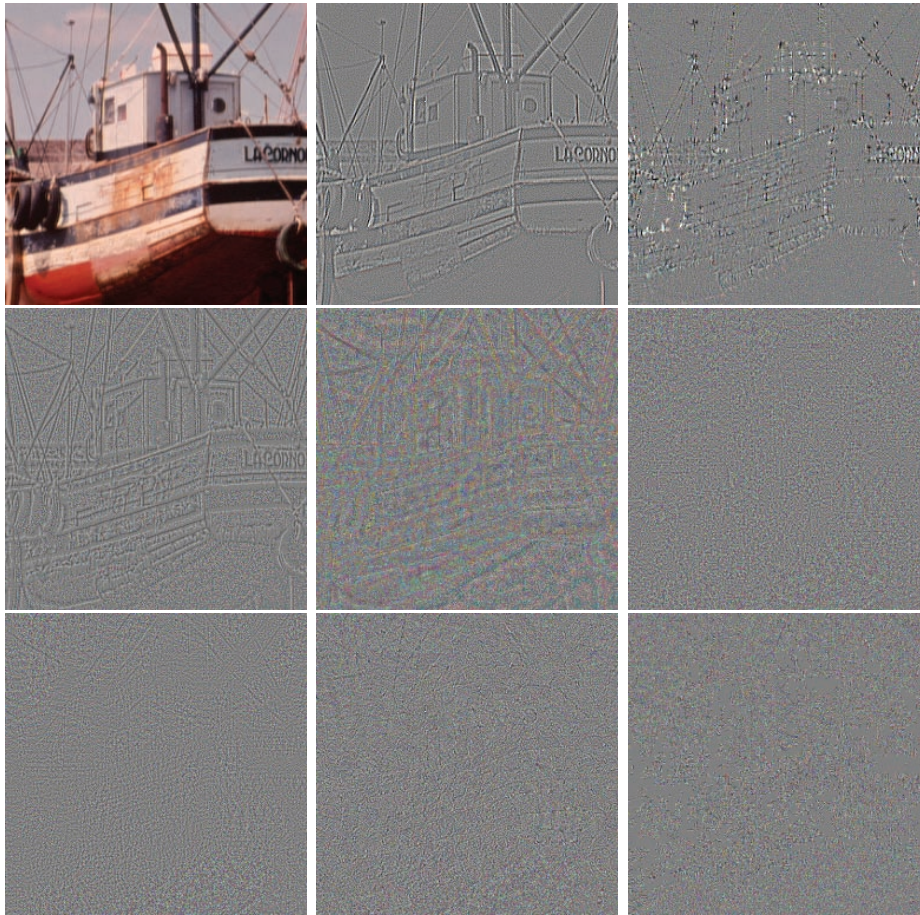
Fig. 11 *Image method noise. From left to right and from top to bottom: original image, Gaussian convolution, anisotropic filtering, total variation, neighborhood filter, HWT, TIHWT, DCT empirical Wiener filter, and the NL-means algorithm. The parameters have been set for each method to remove a method noise with variance $\sigma^2 = 2.5^2$.*

as possible of the original image, the method noise should look as much as possible like white noise. Figure 11 displays the method noise of the different methods for a set of standard natural images. Let us comment on them briefly.

- The Gaussian filter method noise highlights all important features of the image. These features have a large Laplacian and are therefore altered by the algorithm; see Theorem 2.1.
- As announced in Theorem 2.3, the anisotropic filter method noise displays the corners and high frequency features. The straight edges are instead not to be seen: they have a low curvature.
- The total variation method modifies most structures and details of the image. Even straight edges are not well preserved.
- The neighborhood filter preserves flat objects and contrasted edges, while edges with a low contrast are not kept. In any case, the contours, texture, and details seem to be well preserved.

- The *TIHWT* method noise is concentrated on the edges and high frequency features. These structures cause coefficients with large enough value, but lower than the threshold. They are removed by the algorithm. The average of the application to all translated versions reduces the method noise, and structures are hardly noticeable.
- It is difficult to find noticeable structure in the DCT empirical Wiener filter method noise. Only some contours are noticeable. In general, this filter seems to perform much better than all local smoothing filters and other frequency domain filters. Its results are similar to those of a hard stationary wavelet thresholding.
- The NL-means method noise seems closest to white noise.

5.2.2. The "Noise-to-Noise" Criterion. The *noise-to-noise* principle requires that a denoising algorithm transforms a white noise into white noise. This paradoxical requirement seems to be the best way to characterize artifact-free algorithms. The transformation of a white noise into any correlated signal creates structure and artifacts. Only white noise is perceptually devoid of structure, as was pointed out by Attneave [5].

Figure 12 shows how denoising methods transform a white noise. The convolution with a Gauss kernel is equivalent to the product in the Fourier domain with a Gauss kernel of inverse standard deviation. Therefore, convolving the noise with a kernel reinforces the low frequencies and cancels the high ones. Thus, the filtered noise actually shows big grains due to its prominent low frequencies.

Noise filtered by a wavelet thresholding is no more a white noise. The few coefficients with a magnitude larger than the threshold are spread all over the image. The pixels which do not belong to the support of one of these coefficients are set to zero. The visual result is a constant image with superposed wavelets as displayed in Figure 12. It is easy to prove that the denoised noise is spatially highly correlated.

Given a noise realization, the filtered value by the neighborhood filter at a pixel i depends only on its value $n(i)$ and the parameters h and ρ. The neighborhood filter averages noise values at a distance from $n(i)$ less than or equal to h. Thus, when the size ρ of the neighborhood increases, by the law of large numbers the filtered value tends to the expectation of the Gauss distribution restricted to the interval $(n(i) - h, n(i) + h)$. This filtered value is therefore a deterministic function of $n(i)$ and h. Independent random variables are mapped by a deterministic function on independent variables. Thus the noise-to-noise requirement is asymptotically satisfied by the neighborhood filter. The NL-means satisfies the noise-to-noise principle to the same extent as a neighborhood filter. However, a mathematical statement and proof of this property are more intricate and we shall skip them.

5.2.3. Visual Quality Comparison. The visual quality of the restored image is another important criterion to judge the performance of a denoising algorithm. On classic benchmark images, Figures 13–15 control the absence of artifacts and the correct reconstruction of edges, texture, and fine structure.

Figure 13 illustrates the fact that a nonlocal algorithm is needed for the correct reconstruction of periodic images. Local smoothing filters and local frequency filters are not able to reconstruct the wall pattern. Only NL-means and the global Fourier–Wiener filter reconstruct the original texture. The Fourier–Wiener filter is based on a global Fourier transform which is able to capture the periodic structure of the image in a few coefficients. But this only is an ideal filter: the Fourier transform of the
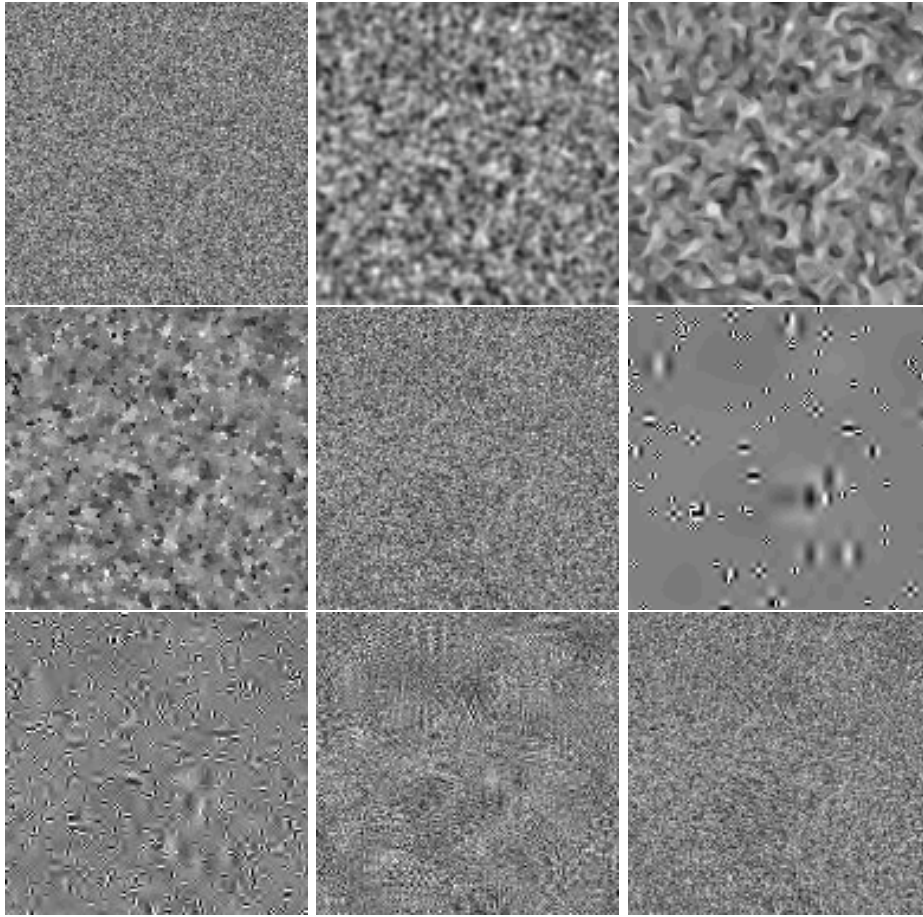
Fig. 12 *The noise to noise criterion. From left to right and from top to bottom: original noise image of standard deviation* 20, *Gaussian convolution, anisotropic filtering, total variation, neighborhood filter, HWT, TIHWT, DCT empirical Wiener filter, and NL-means. Parameters have been fixed for each method so that the noise standard deviation is reduced by a factor of* 4.

original image is being used. Figure 9(e) shows how NL-means chooses the correct weight configuration and explains the correct reconstruction of the wall pattern.

Figure 14 shows that the frequency domain filters are well adapted to the recovery of oscillatory patterns. Although some artifacts are noticeable in both solutions, the stripes are well reconstructed. The DCT transform seems to be more adapted to this type of texture, and stripes are a little better reconstructed. For a much more detailed comparison between sliding window transform domain filtering methods and wavelet threshold methods, we refer the reader to [96]. NL-means also performs well on this type of texture, due to its high degree of redundancy.

5.2.4. Mean Square Error Comparison. The mean square error is the square of the Euclidean distance between the original image and its estimate. This numerical quality measurement is the more objective one, since it does not rely on any visual interpretation. Table 1 shows the mean square error of the different denoising methods
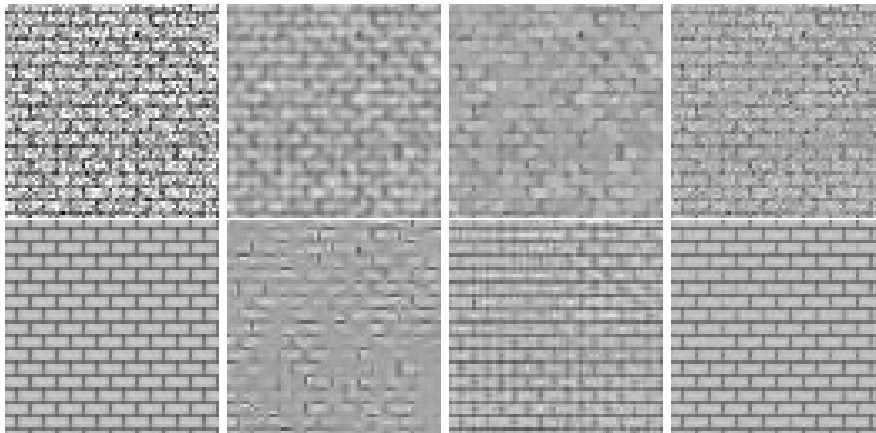
Fig. 13 *Denoising experience on a periodic image. From left to right and from top to bottom: noisy image (standard deviation 35), Gauss filtering, total variation, neighborhood filter, Wiener filter (ideal filter), TIHWT, DCT empirical Wiener filtering, and NL-means.*
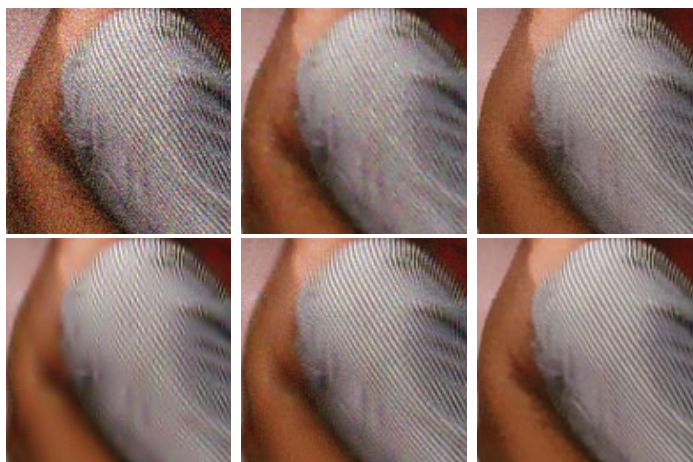

Fig. 14 *Denoising experience on a natural image. From left to right and from top to bottom: noisy image (standard deviation 35), total variation, neighborhood filter, translation invariant hard thresholding (TIHWT), empirical Wiener, and NL-means.*

with the images presented in this paper. This error table seems to corroborate the observations made for the other criteria. One sees, for example, how the frequency domain filters have a lower mean square error than the local smoothing filters. One also sees that in the presence of periodic or textural structures the empirical Wiener filter based on a DCT transform performs better than the wavelet thresholding. Note that, in the presence of periodic or stochastic patterns, NL-means is significantly more precise than the other algorithms. Of course, the errors presented in this table cannot be computed in a real denoising problem. Let us remark, however, that a small error does not guarantee a good visual quality of the restored image.
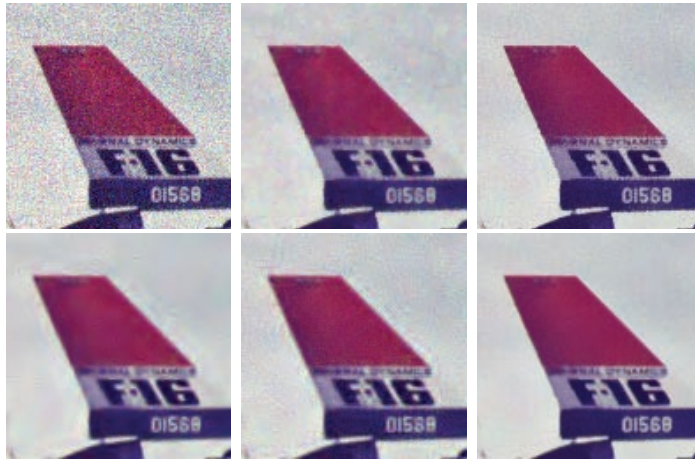
Fig. 15 *Denoising experience on a natural image. From left to right and from top to bottom: noisy image (standard deviation 35), total variation, neighborhood filter, translation invariant hard thresholding (TIHWT), empirical Wiener, and NL-means.*

Table 1 *Mean square error table. A smaller mean square error indicates that the estimate is closer to the original image. The numbers have to be compared on each row. The square of the number on the left-hand column gives the real variance of the noise. By comparing this square to the values on the same row, it is quickly checked that all studied algorithms indeed perform some denoising. This is a sanity check! In general, the comparison performance corroborates the previously mentioned quality criteria.*

Image	σ	GF	AF	TV	YNF	EWF	TIHWT	NL-means
Boat	8	53	38	39	39	33	28	23
Airplane	20	159	138.06	57.15	68.72	78.3	50.97	38.31
Barbara	25	220	216	186	176	111	135	72
Wall	35	580	660	721	598	325	712	59

6. NL-Means Five Years Later.

6.1. The Nonlocal Paradigm. Since the publication of the 2005 paper [20], several equivalent terms have been devised: "exemplar-based" processing, "self-similarity," "patch-based" processing, "nonlocal" processing, "block-matching," and "sparse dictionaries of blocks." They all refer to generalizations or alternatives of the nonlocal means algorithm.

Our conclusions on the better denoising performance of a nonlocal method with respect to state-of-the-art algorithms such as the total variation or the wavelet thresholding have been widely accepted [80] and [44]. The method noise methodology to compare the denoising performance has been adopted ever since. The noise-to-noise criterion introduced in [18] tackles the risk that, starting from pure noise, a denoising algorithm creates structured features. This may well explain the interest for NL-means in scientific or medical imaging, where the neutrality of algorithms is crucial. Improvements or adaptations of NL-means have been proposed in cryon microscopy [29], fluorescence microscopy [9], magnetic resonance imaging (MRI) [61], [26], [91], [68], multispectral MRI [62], [12], and diffusion tensor MRI (DT-MRI) [90].

The notion of image self-similarity itself has been explored and quantified in [2], where it is noted that "blocks are well approximated by a number of other blocks at the same or different scales when affine grey scale transformations are employed." See a similar conclusion in [73].

Likewise, several papers have explored which degree of invariance could be applied to image blocks. [100] explores a rotationally invariant block-matching strategy improving NL-means, and [37] uses cross-scale (i.e., downsampled) neighborhoods in the NL-means filter. Self-similarity has also been explored in the Fourier domain for MRI in [63].

NL-means is a computationally demanding algorithm. Several papers have proposed fast and extremely fast (linear) implementations, by block preselection [56], [7], by Gaussian KD-trees to classify image blocks [1], by SVD [69], by using the FFT to compute correlation between blocks [87], and by statistical arguments [25]. The statistical validity of the NL-means algorithm is wide open. See [85], [50], and [36] (where a Bayesian interpretation is proposed) or [93], where a bias of NL-means is corrected. [45] gives "a probabilistic interpretation and analysis of the method viewed as a random walk on the patch space."

6.2. Generalization to Other Image Processing Tasks or Other Data. The nonlocal denoising principle also works for 3D data set points [98], [31], [48]. But it has also be expanded to most image processing tasks: *Demosaicking*, the operation which transforms the "R or G or B" raw image in each camera into an "R and G and B" image [19], [57]; *movie colorization* [41], [53]; *image inpainting* by proposing a nonlocal image inpainting variational framework with a unified treatment of geometry and texture [4] (see also [92]); *zooming* by a fractal-like technique where examples are taken from the image itself at different scales [36]; *movie flicker stabilization* [30], compensating for spurious oscillations in the colors of successive frames; and *superresolution*, an image zooming method by which several frames from a video, or several low resolution photographs, can be fused into a larger image [76].

The main point of this superresolution technique is that it gives up an explicit estimate of the motion, allowing actually for a multiple motion, since a block can look like several other blocks in the same frame. The very same observation is made in [35] for devising a superresolution algorithm; see also [39], [28].

Other classic image nonlocal applications include *image contrast enhancement* by applying a reverse nonlocal heat equation [15], and *stereo vision* by performing simultaneous nonlocal depth reconstruction and restoration of noisy stereo images [47]. Of course, nonlocal techniques apply still better to movies, the main discovery being that, contrary to the previous state-of-the-art algorithms, *denoising image sequences does not require motion estimation* [14]. This claim is extensively explored in [9], [8], [18].

6.3. Hybrid Approaches: Combining NL-Means and Linear Transforms. Most improvements of NL-means combine the nonlocal principle with former classic algorithms and have indeed shown an improved denoising performance. Probably the best performing method so far is the hybrid method BM3D proposed in [27], which combines not less than block-matching, linear transform thresholding, and Wiener filtering! In the words of the authors:

> The enhancement of the sparsity is achieved by grouping similar 2-D image fragments (e.g., blocks) into 3-D data arrays which we call groups. Collaborative filtering is a special procedure developed to deal with these 3-D groups. We realize it using the three successive steps: 3-D transfor-

mation of a group, shrinkage of the transform spectrum, and inverse 3-D
transformation. The result is a 3-D estimate that consists of the jointly
filtered grouped image blocks. By attenuating the noise, the collaborative
filtering reveals even the finest details shared by grouped blocks and, at
the same time, it preserves the essential unique features of each individual
block. The filtered blocks are then returned to their original positions. Be-
cause these blocks are overlapping, for each pixel, we obtain many different
estimates which need to be combined. Aggregation is a particular averag-
ing procedure which is exploited to take advantage of this redundancy. A
significant improvement is obtained by a specially developed collaborative
Wiener filtering.

Another iterative and variational version of NL-means which improves the patch
distance at each iteration is given in [11]. A related approach combining NL-means
and wavelet subband mixing is proposed in [74].

6.4. Nonlocal PDEs, Nonlocal Variational Principles. The relationship of neigh-
borhood filters to classic local PDEs has been discussed in [16] and [17], leading to
an adaptation of NL-means which avoids the staircasing effect. Yet the main interest
has shifted to defining *nonlocal PDEs*. The extension of the NL-means method to de-
fine nonlocal image-adapted differential operators and nonlocal variational methods
starts with [51], which proposes to perform denoising and deblurring by a nonlocal
functionals. The general goal of this development is actually to give a variational to
all neighborhood filters, and to give a nonlocal form to the total variation as well.
More precisely, the neighborhood filters derive from the functional

$$J(u) = \int_{\Omega \times \Omega} g\left(\frac{|u(x) - u(y)|^2}{h^2}\right) w(|x - y|) dxdy,$$

where g and w have a Gaussian decay. The *nonlocal total variation* is

$$NLBV(u) := \int_{\Omega \times \Omega} |\nabla u(x) - \nabla(y)| dxdy.$$

In the same line, a functional yields a (variational) interpretation to NL-means:

$$JNL(u) = \int_{\Omega \times \Omega} \left(1 - e^{-\frac{G_\sigma * |u(x - .) - u(y - .)|^2(0)}{h^2}}\right) w(|x - y|) dxdy.$$

Nonlocal regularization seems adapted to image deblurring, a notoriously ill-posed
problem. The nonlocal differential operators permit one to define a total variation
or a Dirichlet integral. Several articles on deblurring have followed this variational
line [49], [66], [43] (for image segmentation), [9] (in fluorescence microscopy), [99],
(again for nonlocal deconvolution), and [55] (for deconvolution and tomographic re-
construction). In [39], a paper dedicated to another notoriously ill-posed problem, the
superresolution, the nonlocal variational principle is viewed as "an emerging powerful
family of regularization techniques"; the paper "proposes to use the example-based
approach as a new regularizing principle in ill-posed image processing problems such
as image superresolution from several low resolution photographs." A particular no-
tion of nonlocal PDE has emerged, whose coefficients are actually image-dependent.
For instance, in [41] the image colorization is viewed as the minimization of a discrete
partial differential functional on the weighted block graph. Thus, it can be seen either
as a nonlocal heat equation on the image or as a local heat equation on the space of
image patches.

In an almost equivalent framework, in [84] the set of blocks is viewed as a weighted graph, and the weights of the edge between two blocks centered at x and y, respectively, are decreasing functions of the block distances. Then a graph Laplacian can be calculated on this graph, seen as the sampling of a manifold, and NL-means can be interpreted as the heat equation on the set of blocks endowed with these weights. In the same way, the neighborhood filter is associated with a heat equation on the image graph. This approach is further extended to a variational formulation on patch graphs in [40]. In this same framework [15] proposed to perform image contrast enhancement by applying a nonlocal reverse heat equation. Finally, always in this nonlocal partial differential framework, [10] extends the Mumford–Shah image segmentation energy to contain a nonlocal self-similarity term replacing the usual Dirichlet term. The square of the gradient is replaced by the square of the nonlocal gradient.

6.5. Sparse Representations on Blocks. The exploration of image redundancy and its application to image restoration has led to new attempts at sparse image representations by block dictionaries [50]. The paper [57] contains a careful bibliographical analysis of sparse image representations. It also proposes an algorithm building a redundant block basis. Using the K-SVD algorithm, an optimal redundant basis is looked for in the space of blocks, with the criterion that the decomposition of all other blocks in the image must be as sparse as possible. Such sparse bases can be efficiently used for such tasks as image denoising and image demosaicking. This approach and variants applied to video have been considered in [58], [75], and [23].

REFERENCES

[1] A. ADAMS, N. GELFAND, J. DOLSON, AND M. LEVOY, *Gaussian KD-trees for fast high-dimensional filtering*, ACM Trans. Graphics, 28 (2009), article 21.

[2] S. ALEXANDER, E. VRSCAY, AND S. TSURUMI, *A simple, general model for the affine self-similarity of images*, in Proceedings of the 5th International Conference on Image Analysis and Recognition, Lecture Notes in Comput. Sci. 5112, Springer-Verlag, Berlin, 2008, pp. 192–203.

[3] L. ALVAREZ, P.-L. LIONS, AND J.-M. MOREL, *Image selective smoothing and edge detection by nonlinear diffusion.* II, SIAM J. Numer. Anal., 29 (1992), pp. 845–866.

[4] P. ARIAS, V. CASELLES, AND G. SAPPIRO, *A variational framework for non-local image inpainting*, in Proceedings of the 7th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), Lecture Notes in Comput. Sci. 5681, Springer-Verlag, Berlin, 2009, pp. 345–358.

[5] F. ATTNEAVE, *Some informational aspects of visual perception*, Psychol. Rev., 61 (1954), pp. 183–193.

[6] S. AWATE AND R. WHITAKER, *Unsupervised, information-theoretic, adaptive image filtering for image restoration*, IEEE Trans. Pattern Anal. Mach. Intell., 28 (2006), pp. 364–376.

[7] R. BILCU AND M. VEHVILAINEN, *Combined non-local averaging and intersection of confidence intervals for image denoising*, in Proceedings of the 15th IEEE International Conference on Image Processing, 2008, pp. 1736–1739.

[8] J. BOULANGER, C. KERVRANN, AND P. BOUTHEMY, *Space-time adaptation for patch-based image sequence restoration*, IEEE Trans. Pattern Anal. Mach. Intell., 29 (2007), pp. 1096–1102.

[9] J. BOULANGER, J. SIBARITA, C. KERVRANN, AND P. BOUTHEMY, *Non-parametric regression for patch-based fluorescence microscopy image sequence denoising*, in Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008, pp. 748–751.

[10] X. BRESSON AND T. CHAN, *Non-local Unsupervised Variational Image Segmentation Models*, UCLA CAM Report 08-67, University of California, Los Angeles, 2008.

[11] T. BROX AND D. CREMERS, *Iterated nonlocal means for texture restoration*, Scale Space and Variational Methods in Computer Vision, Lecture Notes in Comput. Sci. 4485, Springer-Verlag, Berlin, 2007, pp. 13–24.

[12] A. BUADES, A. CHIEN, J. M. MOREL, AND S. OSHER, *Topology preserving linear filtering applied to medical imaging*, SIAM J. Imaging Sci., 1 (2008), pp. 26–50.

[13] A. BUADES, B. COLL, AND J. MOREL, *Image Data Process by Image Noise Reduction and Camera Integrating the Means for Implementing This Process*, French Patent 0404837.

[14] A. BUADES, B. COLL, AND J. MOREL, *Denoising image sequences does not require motion estimation*, in Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2005, pp. 70–74.

[15] A. BUADES, B. COLL, AND J. MOREL, *Image Enhancement by Non-local Reverse Heat Equation*, CMLA Preprint 22, Centre de Mathématiques et Leurs Applications, Cachan, France, 2006.

[16] A. BUADES, B. COLL, AND J. MOREL, *Neighborhood filters and PDE's*, Numer. Math., 105 (2006), pp. 1–34.

[17] A. BUADES, B. COLL, AND J. MOREL, *The staircasing effect in neighborhood filters and its solution*, IEEE Trans. Image Process., 15 (2006), pp. 1499–1505.

[18] A. BUADES, B. COLL, AND J. MOREL, *Nonlocal image and movie denoising*, Internat. J. Computer Vision, 76 (2008), pp. 123–139.

[19] A. BUADES, B. COLL, J. MOREL, AND C. SBERT, *Self-similarity driven color demosaicking*, IEEE Trans. Image Process., 18 (2009), pp. 1192–1202.

[20] A. BUADES, B. COLL, AND J. M. MOREL, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul., 4 (2005), pp. 490–530.

[21] F. CATTÉ, F. DIBOS, AND G. KOEPFLER, *A morphological scheme for mean curvature motion and applications to anisotropic diffusion and motion of level sets*, SIAM J. Numer. Anal., 32 (1995), pp. 1895–1909.

[22] A. CHAMBOLLE AND P. L. LIONS, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.

[23] P. CHATTERJEE AND P. MILANFAR, *Image denoising using locally learned dictionaries*, in Computational Imaging VII, P. Chatterjee and P. Milanfar, eds., Proc. SPIE 7246, SPIE, Bellingham, WA, 2009, article 72460V.

[24] R. R. COIFMAN AND D. DONOHO, *Translation-invariant de-noising*, in Wavelets and Statistics, Lecture Notes in Statist., Springer-Verlag, New York, 1995, pp. 125–150.

[25] P. COUPE, P. YGER, AND C. BARILLOT, *Fast non local means denoising for 3D MR images*, in Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer-Verlag, Berlin, 2006, pp. 33–40.

[26] P. COUPÉ, P. YGER, S. PRIMA, P. HELLIER, C. KERVRANN, AND C. BARILLOT, *An optimized blockwise nonlocal means denoising filter for 3D magnetic resonance images*, Trans. Med. Imaging, 27 (2008), pp. 425–441.

[27] K. DABOV, A. FOI, V. KATKOVNIK, AND K. EGIAZARIAN, *Image denoising by sparse 3-D transform-domain collaborative filtering*, IEEE Trans. Image Process., 16 (2007), pp. 2080–2095.

[28] A. DANIELYAN, A. FOI, V. KATKOVNIK, AND K. EGIAZARIAN, *Image and Video Super-Resolution via Spatially Adaptive Block-Matching Filtering*, in Proceedings of the International Workshop on Local and Non-local Approximation in Image Processing, 2008.

[29] J. DARBON, A. CUNHA, T. CHAN, S. OSHER, AND G. JENSEN, *Fast nonlocal filtering applied to electron cryomicroscopy*, in Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008, pp. 1331–1334.

[30] J. DELON AND A. DESOLNEUX, *Flicker Stabilization in Image Sequences*, preprint; available online from http://hal.archives-ouvertes.fr/docs/00/40/77/96/PDF/flicker_hal_juillet2009.pdf.

[31] B. DONG, J. YE, S. OSHER, AND I. DINOV, *Level set based nonlocal surface restoration*, Multiscale Model. Simul., 7 (2008), pp. 589–598.

[32] D. DONOHO, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.

[33] D. DONOHO AND I. JOHNSTONE, *Ideal spatial adaptation via wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.

[34] S. DURAND AND M. NIKOLOVA, *Restoration of wavelet coefficients by minimizing a specially designed objective function*, in Proceedings of the IEEE Workshop on Variational and Level Set Methods in Computer Vision, 2003, pp. 145–152.

[35] M. Ebrahimi and E. Vrscay, *Multi-frame super-resolution with no explicit motion estimation*, in Proceedings of the 2008 International Conference on Image Processing, Computer Vision, and Pattern Recognition, 2008, pp. 455–459.

[36] M. Ebrahimi and E. Vrscay, *Solving the Inverse Problem of Image Zooming Using "Self-Examples,"* in Image Analysis and Recognition, Lecture Notes in Comput. Sci. 4633, Springer-Verlag, Berlin, 2007, pp. 117–130.

[37] M. Ebrahimi and E. Vrscay, *Examining the role of scale in the context of the non-local-means filter*, in Image Analysis and Recognition, Lecture Notes in Comput. Sci. 5112, Springer-Verlag, Berlin, 2008, pp. 170–181.

[38] A. Efros and T. Leung, *Texture synthesis by non-parametric sampling*, in Proceedings of the IEEE International Conference on Computer Vision, Vol. 2, Corfu, Greece, 1999, pp. 1033–1038.

[39] M. Elad and D. Datsenko, *Example-based regularization deployed to super-resolution reconstruction of a single image*, The Computer J., 50 (2007), pp. 1–16.

[40] A. Elmoataz, O. Lezoray, and S. Bougleux, *Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing*, IEEE Trans. Image Process., 17 (2008), pp. 1047–1060.

[41] A. Elmoataz, O. Lezoray, S. Bougleux, and V. Ta, *Unifying local and nonlocal processing with partial difference operators on weighted graphs*, in International Workshop on Local and Non-Local Approximation in Image Processing, 2008; available online from http://www.eurasip.org/Proceedings/Ext/LNLA2008/papers/cr1028.pdf.

[42] D. Gabor, *Information theory in electron microscopy*, Laboratory Investigation, 14 (1965), pp. 801–807.

[43] G. Gilboa and S. Osher, *Nonlocal linear image regularization and supervised segmentation*, Multiscale Model. Simul., 6 (2007), pp. 595–630.

[44] B. Goossens, H. Luong, A. Pizurica, and W. Philips, *An improved non-local denoising algorithm*, in Proceedings of the 2008 International Workshop on Local and Non-local Approximation in Image Processing, Lausanne, Switzerland, 2008; available online from http://130.230.88.154/images/8/8b/Cr1026-lausanne.pdf.

[45] S. Gornale, R. Manza, V. Humbe, and K. Kale, *Performance analysis of bi-orthogonal wavelet filters for lossy fingerprint image compression*, Internat. J. Image Sci. Engrg., 1 (2007), pp. 16–20.

[46] F. Guichard, J. M. Morel, and R. Ryan, *Contrast Invariant Image Analysis and PDE's*, book in preparation.

[47] Y. Heo, K. Lee, and S. Lee, *Simultaneous depth reconstruction and restoration of noisy stereo images using non-local pixel distribution*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.

[48] B. Huhle, T. Schairer, P. Jenke, and W. Strasser, *Robust non-local denoising of colored depth data*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1–7.

[49] M. Jung and L. Vese, *Nonlocal variational image deblurring models in the presence of Gaussian or impulse noise*, in Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision, Lecture Notes in Comput. Sci. 5567, Springer-Verlag, Berlin, 2009, pp. 401–412.

[50] C. Kervrann, J. Boulanger, and P. Coupe, *Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal*, in Scale Space and Variational Methods in Computer Vision, Lecture Notes In Comput. Sci. 4485, Springer-Verlag, Berlin, 2007, pp. 520–532.

[51] S. Kindermann, S. Osher, and P. W. Jones, *Deblurring and denoising of images by nonlocal functionals*, Multiscale Model. Simul., 4 (2005), pp. 1091–1115.

[52] E. Levina, *Statistical Issues in Texture Analysis*, Ph.D. thesis, University of California, Berkeley, CA, 2002.

[53] O. Lezoray, V. Ta, and A. Elmoataz, *Nonlocal graph regularization for image colorization*, in Proceedings of the 19th International Conference on Pattern Recognition, 2008, pp. 1–4.

[54] M. Lindenbaum, M. Fischer, and A. M. Bruckstein, *On Gabor's contribution to image enhancement*, Pattern Recognition, 27 (1994), pp. 1–8.

[55] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, *Image Recovery via Nonlocal Operators*, UCLA CAM Report 08-35, University of California, Los Angeles, 2008.

[56] M. Mahmoudi and G. Sapiro, *Fast image and video denoising via nonlocal means of similar neighborhoods*, IEEE Signal Process. Lett., 12 (2005), pp. 839–842.

[57] J. MAIRAL, M. ELAD, AND G. SAPIRO, *Sparse representation for color image restoration*, IEEE Trans. Image Process., 17 (2008), pp. 53–69.

[58] J. MAIRAL, G. SAPIRO, AND M. ELAD, *Learning multiscale sparse representations for image and video restoration*, Multiscale Model. Simul., 7 (2008), pp. 214–241.

[59] F. MALGOUYRES, *Noise selection approach to image restoration*, in Wavelets: Applications in Signal and Image Processing IX, Proc. SPIE 4478, SPIE, Bellingham, WA, 2001, pp. 34–41.

[60] S. MALLAT, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1999.

[61] J. MANJÓN, J. CARBONELL-CABALLERO, J. LULL, G. GARCÍA-MARTÍ, L. MARTÍ-BONMATÍ, AND M. ROBLES, *MRI denoising using non-local means*, Med. Image Anal., 12 (2008), pp. 514–523.

[62] J. MANJÓN, M. ROBLES, AND N. THACKER, *Multispectral MRI denoising using non-local means*, in Proceedings of Medical Image Understanding and Analysis 2007 (MIUA 2007), Vol. 7, Wales, 2007, pp. 41–45.

[63] G. MAYER, E. VRSCAY, M. LAUZON, B. GOODYEAR, AND J. MITCHELL, *Self-similarity of Images in the Fourier Domain, with Applications to MRI*, in Image Analysis and Recognition, Lecture Notes in Comput. Sci. 5112, Springer-Verlag, Berlin, 2008, pp. 43–52.

[64] B. MERRIMAN, J. BENCE, AND S. OSHER, *Diffusion generated motion by mean curvature*, in Computational Crystal Growers Workshop, AMS, Providence, RI, 1992.

[65] Y. MEYER, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, Univ. Lecture Ser. 22, AMS, Providence, RI, 2002.

[66] M. MIGNOTTE, *A non-local regularization strategy for image deconvolution*, Pattern Recognition Lett., 29 (2008), pp. 2206–2212.

[67] É. A. NADARAYA, *On non-parametric estimates of density functions and regression curves*, Theory Probab. Appl., 10 (1964), pp. 186–190.

[68] B. NAEGEL, A. CERNICANU, J. HYACINTHE, M. TOGNOLINI, AND J. VALLÉE, *SNR enhancement of highly-accelerated real-time cardiac MRI acquisitions based on non-local means algorithm*, Med. Image Anal., 13 (2009), pp. 598–608.

[69] J. ORCHARD, M. EBRAHIMI, AND A. WONG, *Efficient non-local-means denoising using the SVD*, in Proceedings of the 15th IEEE International Conference on Image Processing, 2008, pp. 1732–1735.

[70] E. ORDENTLICH, G. SEROUSSI, S. VERDÚ, M. WEINBERGER, AND T. WEISSMAN, *A discrete universal denoiser and its application to binary images*, in Proceedings of the 2003 IEEE International Conference on Image Processing, Vol. 1, Barcelona, Spain, 2003, pp. 117–120.

[71] S. OSHER, M. BURGER, D. GOLDFARB, J. XU, AND W. YIN, *Using Geometry and Iterated Refinement for Inverse Problems* 1: *Total Variation Based Image Restoration*, Department of Mathematics, University of California, Los Angeles, 2004, pp. 4–13.

[72] P. PERONA AND J. MALIK, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intell., 12 (1990), pp. 629–639.

[73] G. PEYRÉ, *Sparse modeling of textures*, J. Math. Imaging Vision, 34 (2009), pp. 17–31.

[74] C. PIERRICK ET AL., *3D wavelet subbands mixing for image denoising*, Internat. J. Biomed. Imaging, 2008 (2008), pp. 1–11.

[75] M. PROTTER AND M. ELAD, *Image sequence denoising via sparse and redundant representations*, IEEE Trans. Image Process., 18 (2009), pp. 27–35.

[76] M. PROTTER, M. ELAD, H. TAKEDA, AND P. MILANFAR, *Generalizing the non-local-means to super-resolution reconstruction*, IEEE Trans. Image Process., 18 (2009), pp. 36–51.

[77] G. G. ROUSSAS, *Nonparametric regression estimation under mixing conditions*, Stochastic Process. Appl., 36 (1990), pp. 107–116.

[78] L. RUDIN AND S. OSHER, *Total variation based image restoration with free local constraints*, in Proceedings of the IEEE International Conference on Image Processing, Vol. 1, Austin, TX, 1994, pp. 31–35.

[79] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[80] H. SEO, P. CHATTERJEE, H. TAKEDA, AND P. MILANFAR, *A comparison of some state of the art image denoising methods*, in Proceedings of the 41st Asilomar Conference on Signals, Systems, and Computers, 2007.

[81] C. SHANNON, *A mathematical theory of communication*, ACM SIGMOBILE Mobile Comput. Commun. Rev., 5 (2001), pp. 3–55.

[82] A. SHARF, M. ALEXA, AND D. COHEN-OR, *Context-based surface completion*, ACM Trans. Graphics, 23 (2004), pp. 878–887.

[83] S. M. SMITH AND J. M. BRADY, *SUSAN—a new approach to low level image processing*, Internat. J. Computer Vision, 23 (1997), pp. 45–78.

[84] A. Szlam, M. Maggioni, and R. Coifman, *A General Framework for Adaptive Regularization Based on Diffusion Processes on Graphs*, Technical Report, Yale University, New Haven, CT, 2006.

[85] N. Thacker, J. Manjon, and P. Bromiley, *A statistical interpretation of non-local means*, in Proceedings of the 5th International Conference on Visual Information Engineering, 2008, pp. 250–255.

[86] C. Tomasi and R. Manduchi, *Bilateral filtering for gray and color images*, in Proceedings of the 6th International Conference on Computer Vision, 1998, pp. 839–846.

[87] J. Wang, Y. Guo, Y. Ying, Y. Liu, and Q. Peng, *Fast non-local algorithm for image denoising*, in Proceedings of the 2006 IEEE International Conference on Image Processing, 2006, pp. 1429–1432.

[88] G. S. Watson, *Smooth regression analysis*, Sankhyā Ser. A, 26 (1964), pp. 359–372.

[89] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdu, and M. Weinberger, *Universal discrete denoising: Known channel*, IEEE Trans. Inform. Theory, 51 (2005), pp. 5–28.

[90] N. Wiest-Daesslé, S. Prima, P. Coupé, S. Morrissey, and C. Barillot, *Non-local means variants for denoising of diffusion-weighted and diffusion tensor MRI*, in Medical Image Computing and Computer-Assisted Intervention (MICCAI), Lecture Notes in Comput. Sci. 4792, Springer-Verlag, Berlin, 2007, pp. 344–351.

[91] N. Wiest-Daessle, S. Prima, P. Coupé, S. Morrissey, and C. Barillot, *Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI*, in Medical Image Computing and Computer-Assisted Intervention (MICCAI), Lecture Notes in Comput. Sci. 5242, Springer-Verlag, Berlin, 2008, pp. 171–179.

[92] A. Wong and J. Orchard, *A nonlocal-means approach to exemplar-based inpainting*, in Proceedings of the 15th IEEE International Conference on Image Processing, 2008, pp. 2600–2603.

[93] H. Xu, J. Xu, and F. Wu, *On the biased estimation of nonlocal means filter*, in Proceedings of the 2008 IEEE International Conference on Multimedia and Expo, 2008, pp. 1149–1152.

[94] L. Yaroslavsky, *Local adaptive image restoration and enhancement with the use of DFT and DCT in a running window*, in Mathematical Imaging: Wavelet Applications in Signal and Image Processing IV, Proc. SPIE 2825, SPIE, Bellingham, WA, 1996, pp. 2–13.

[95] L. Yaroslavsky and M. Eden, *Fundamentals of Digital Optics*, Birkhäuser, Boston, 2003.

[96] L. Yaroslavsky, K. Egiazarian, and J. Astola, *Transform domain image restoration methods: Review, comparison and interpretation*, in Nonlinear Image Processing and Pattern Analysis, SPIE, Bellingham, WA, 2001, pp. 155–169.

[97] L. P. Yaroslavsky, *Digital Picture Processing. An Introduction*, Springer-Verlag, Berlin, 1985.

[98] S. Yoshizawa, A. Belyaev, and H. Seidel, *Smoothing by example: Mesh denoising by averaging with similarity-based weights*, in Proceedings of the IEEE International Conference on Shape Modeling and Applications, 2006, p. 9.

[99] X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction*, UCLA CAM Report 09-03, University of California, Los Angeles, 2009.

[100] S. Zimmer, S. Didas, and J. Weickert, *A rotationally invariant block matching strategy improving image denoising with non-local means*, in Proceedings of the 2008 International Workshop on Local and Non-local Approximation in Image Processing, 2008, pp. 135–142.