

Optimal Spatial Adaptation for Patch-Based Image Denoising

Charles Kervrann and Jérôme Boulanger

Abstract—A novel adaptive and patch-based approach is proposed for image denoising and representation. The method is based on a pointwise selection of small image patches of fixed size in the variable *neighborhood* of each pixel. Our contribution is to associate with each pixel the weighted sum of data points within an adaptive neighborhood, in a manner that it balances the accuracy of approximation and the stochastic error, at each spatial position. This method is general and can be applied under the assumption that there exists repetitive patterns in a local neighborhood of a point. By introducing spatial adaptivity, we extend the work earlier described by Buades *et al.* which can be considered as an extension of bilateral filtering to image patches. Finally, we propose a nearly parameter-free algorithm for image denoising. The method is applied to both artificially corrupted (white Gaussian noise) and real images and the performance is very close to, and in some cases even surpasses, that of the already published denoising methods.

I. INTRODUCTION

TRADITIONALLY, the problem with image restoration is to reduce undesirable distortions and noise while preserving important features, such as homogeneous regions, discontinuities, edges, and textures. Most of the more efficient regularization methods are based on energy functionals minimization since they are designed to explicitly account for the image geometry, involving the adjustment of global weights that balance the contribution of prior smoothness terms and a fidelity term [41], [48]. Thus, related partial differential equations (PDEs) and variational methods have shown impressive results to tackle the problem of edge-preserving smoothing [9], [10], [11], [43], [48]. For reasons of performance and robustness in image processing, other smoothing algorithms aggregate information over a neighborhood of fixed size, based on two basic criteria: a spatial criterion so that filtering must be local and a brightness criterion in order to select only points which are similar in some sense. In view of this generic approach, a typical filter is the sigma filter [32]. A continuous version of this filter gives the well-known nonlinear Gaussian filter [25] defined as

$$\hat{u}(\mathbf{x}_i) = \frac{\sum_{\mathbf{x}_j \in \Delta_i} L_g(Y_i - Y_j) Y_j}{\sum_{\mathbf{x}_j \in \Delta_i} L_g(Y_i - Y_j)} \quad (1)$$

Manuscript received July 13, 2005; revised January 4, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Giovanni Ramponi.

The authors are with the IRISA—INRIA Rennes/INRA MIA, 35 042 Rennes cedex, France (e-mail: ckervran@irisa.fr; jboulang@irisa.fr).

Digital Object Identifier 10.1109/TIP.2006.877529

where Δ_i is the window centered at pixel \mathbf{x}_i , Y_i is the observation at \mathbf{x}_i , $L_g(z) = \exp(-z^2/2g^2)$, and g is a smoothing parameter to be determined and depending on the signal-to-noise ratio (SNR). Finally, if we substitute a Gaussian window to the hard disk-shaped window around the current position \mathbf{x}_i , we get variants of the *bilateral filtering* [2], [55] (e.g., Lee's [32], Susan's [51], and Saint-Marc's [49] filters) of the form

$$\hat{u}(\mathbf{x}_i) = \frac{\sum_{\mathbf{x}_j \in \Delta_i} L_g(Y_i - Y_j) K_h(\mathbf{x}_i - \mathbf{x}_j) Y_j}{\sum_{\mathbf{x}_j \in \Delta_i} L_g(Y_i - Y_j) K_h(\mathbf{x}_i - \mathbf{x}_j)} \quad (2)$$

where $K_h(\cdot) = (1/h)K(\cdot/h)$ and $L_g(\cdot) = (1/g)L(\cdot/g)$ are rescaled versions of nonnegative kernel functions K and L . By definition, the continuous kernel functions K and L fulfill $\int K(z)dz = 1$, $\int L(z)dz = 1$ and have M vanishing moments $\int z^m K(z)dz = 0$, $\int z^m L(z)dz = 0$ for $1 \leq m \leq M$. The weight functions $L_g(\cdot)$ and $K_h(\cdot)$ are typically Gaussian kernels where g and h are, respectively, the standard deviations of the intensity and spatial components. We can control the spatial support of the filter and, thus, the level of blurring by varying h . By varying g , we can adapt the sensitivity of the filter to abrupt changes. As effective as nonlinear Gaussian filters, they lacked a theoretical basis and some of connections to better understood methods have been investigated. Here, we report some recent results. First, emphasizing the importance of extended neighborhoods, Barash and Comaniciu [3] have showed that *bilateral filtering* represents a weighted averaging algorithm which turns out to be an implementation of anisotropic diffusion [43], controlled by a global scale parameter. Elad [19] established further how the *bilateral filter* is algorithmically related to anisotropic diffusion [43] and robust estimation [4] in terms of minimizing functionals. The *bilateral filter* can also be viewed as an Euclidean approximation of the Beltrami flow and originates from image manifold area minimization [52]. Furthermore, Barash and Comaniciu showed that kernel density estimation applied into the joint spatial-range domain yields a powerful processing paradigm—the *mean-shift* procedure [12], [14]—also related to *bilateral filtering* but having additional flexibility [3]. The link between iterative *mean-shift* algorithm, local mode filtering, clustering, local M-estimators, nonlinear diffusion, regularization approaches were already analyzed in [3], [19], [40], [56]. Also, all these methods have been cast into a unified framework for functional minimization combining non-local data and nonlocal smoothness terms in [40]. In particular,

Mrazek *et al.* brought to the fore the large amount of structural similarities between the iterated bilateral filter

$$\hat{u}_{n+1}(\mathbf{x}_i) = \frac{\sum_{\mathbf{x}_j \in \Delta_i} L_g(\hat{u}_n(\mathbf{x}_i) - \hat{u}_n(\mathbf{x}_j)) K_h(\mathbf{x}_i - \mathbf{x}_j) \hat{u}_n(\mathbf{x}_j)}{\sum_{\mathbf{x}_j \in \Delta_i} L_g(\hat{u}_n(\mathbf{x}_i) - \hat{u}_n(\mathbf{x}_j)) K_h(\mathbf{x}_i - \mathbf{x}_j)} \quad (3)$$

and the local M-smoother

$$\hat{u}_{n+1}(\mathbf{x}_i) = \frac{\sum_{\mathbf{x}_j \in \Delta_i} w_g(Y_j - \hat{u}_n(\mathbf{x}_i)) K_h(\mathbf{x}_i - \mathbf{x}_j) Y_j}{\sum_{\mathbf{x}_j \in \Delta_i} w_g(Y_j - \hat{u}_n(\mathbf{x}_i)) K_h(\mathbf{x}_i - \mathbf{x}_j)} \quad (4)$$

where $w_g(z) = \rho'_g(z)$ and $\rho(z)$ is the error norm for M-estimators. It is confirmed that local M-smoothing uses the initial image in the averaging procedure and searches for the minimum of a local criterion whereas iterated bilateral filtering uses the evolving image and has to stop after a certain number of iterations in order to avoid a flat image.

The major drawback of all these smoothing methods, including the TV minimizing process [10], [48], is that there is no satisfying way to retrieve the smoothing or regularization parameters from data. Note that they are usually chosen to give a good visual impression and are heuristically chosen [2], [55]. There have been some recent attempts to automatically estimate the regularization parameter for smoothing (e.g., see [24] and [39]). However, when local characteristics of the data differ significantly across the domain, selecting global smoothing parameters is probably not satisfying. There is seldom a single scale that is appropriate for a complete image. Recently, these difficulties motivated the development of more adaptive methods to cope with inhomogeneities in images. First, in [27], the local amount of Gaussian smoothing is computed in terms of variance in a space-scale framework, through the minimal description length criterion (MDL). An alternative way to select the local scale is to maximize a measure of edge strength over scale space as proposed in [36], but the resulting scale computed from image derivatives, is sensitive to signal-to-noise ratio. Also, the local variance is actually useful for localization of significant image features [5], [23], [27], [30]. More recently, *Total Variation* flow has been suggested since it includes a nonexplicit scale parameter useful for detecting the scale of image features [6]. Another competitive approach consists in decomposing the image into its primary noise, texture and bounded variation (BV) components [1], [38], [42], which actually can be hard to compute, in practice. The use of variable bandwidth mean-shift procedures may be also considered for image simplification [13], but this approach is problematic since it is known to be highly sensitive to noise in images and to the choice of a global initial bandwidth.

What makes image denoising very hard, is that natural images often contain many irrelevant objects. This type of “noise” is sometimes referred to as “clutter.” To develop better image enhancement algorithms that can deal with structured noise, we need explicit models for the many regularities and geometries seen in local patterns. In contrast to the above-cited methods, another line of work consists then in modeling

nonlocal pairwise interactions from training data [59] or a library of natural image patches [21], [47]. The idea is to improve the traditional Markov random field (MRF) models by learning potential functions from examples and extended neighborhoods for computer vision applications (e.g., image modeling [59], image denoising [47], image reconstruction and super-resolution [21], and image rendering [20]). Also, it has been experimentally confirmed that more intuitive patch-based approaches are fearsome for two-dimensional texture synthesis [18] and image inpainting [15], [58].

In our framework, we will also assume that small image patches in the neighborhood of an estimation point contain the essential process required for local denoising. Thus, the proposed patch-based denoising approach is conceptually very simple being based on the key idea of iteratively growing a window at each pixel and adaptively weighting input data. The data points with a similar patch to the central patch will have larger weights in the average as recently proposed by Buades *et al.* [7], [8] who defined the so-called *nonlocal means* filter as

$$\hat{u}(\mathbf{x}_i) = \frac{\sum_{\mathbf{x}_j \in \Delta_i} L_g(\mathbf{Y}_i - \mathbf{Y}_j) K_h(\mathbf{x}_i - \mathbf{x}_j) Y_j}{\sum_{\mathbf{x}_j \in \Delta_i} L_g(\mathbf{Y}_i - \mathbf{Y}_j) K_h(\mathbf{x}_i - \mathbf{x}_j)} \quad (5)$$

where \mathbf{Y}_i denotes a vector of pixel values taken in the neighborhood of a point \mathbf{x}_i . In (5), the similarity between two points \mathbf{x}_i and \mathbf{x}_j is based on the Euclidean distance $\|\mathbf{Y}_i - \mathbf{Y}_j\|^2$ between two vectorized image *patches* (see also [58]). It is worth noting that, if the size of the patch is reduced to one pixel, the *nonlocal means* filter, also controlled by a small number of smoothing parameters g and h , is strictly equivalent to (2). As in [7], [8], [58], we also use small image patches (e.g., 7×7 or 9×9 patches) to compute these weights since they are able to capture local geometric patterns and texels seen in images. Moreover, we adaptively choose a window (neighborhood) which could be large, to balance the accuracy of approximation and the stochastic error, at each spatial position [34]. This adaptation method is a kind of change-point detection procedure, initiated by Lepskii [34].

In fact, our approach exploits this idea of *pointwise adaptive estimation* [34], [35] combined with patch-based techniques for denoising. The proposed approach shares some common points with the recent *nonlocal means* algorithm [8], other patch-based methods [15], [18], [58] and the DUDE algorithm [57]. In [57], the authors propose a two-pass approach and substitute the most frequent patch/symbol seen in a local window to the current corrupted patch/symbol. If theoretically well grounded, this framework cannot be easily applied to denoise one-byte images (256 values) because of infeasible complexity and robustness, and has been only tested for denoising binary images. Other related works to our approach are nonlinear Gaussian filters [2], [40], [55], [56] and statistical smoothing schemes [29], [45], but are enhanced via incorporating either a variable window scheme or patch-based weights.

- In [30] and [45], the weights are not based on similarities between small image patches but similarities between

pointwise estimates in a local neighborhood; this corresponds to the special case when the patch is made up of one pixel.

- In [29], the authors propose to use weights calculated from spatial distances between nearby points and to decompose the neighborhood into sectors to construct an “aggregated” estimator.
- Unlike [7] and [8], the set of parameters and the window sizes vary spatially in our approach.

The remainder of the paper is organized as follows. In Section II, we introduce the image modeling and some notations. In Section III, we formulate the problem of the selection of the best possible window and present a practical algorithm with no hidden parameter for image denoising. In Section IV, we demonstrate the ability of the method to restore artificially corrupted images with additive white Gaussian noise (WGN). We just point out that, to our knowledge, the more competitive methods are recent wavelet-based methods [16], [17], [33], [44], [46], [54]. In Section IV, we have then reported the experimental results when these wavelet-based methods are applied to a commonly used image dataset in image denoising [46] for comparison purposes. Conclusions and perspectives are presented in Section V.

II. PROBLEM STATEMENT

In our framework, we propose to use image patches to take into account complex spatial interactions in images. In contrast to exemplar-based approaches for image modeling [47], [59], we propose an unsupervised method that uses no library of image patches and no computational intensive training algorithms [21], [47]. Our adaptive smoothing works in the joint spatial-range domain as the *nonlocal means* filter [8] but has a more powerful adaptation to the local structure of the data since the size of windows and control parameters are estimated from local image statistics as follows.

Consider the following basic image model:

$$Y_i = u(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, |G| \quad (6)$$

where $\mathbf{x}_i \in G \subset \mathbb{R}^d$, $d \geq 2$ represents the spatial coordinates of the discrete image domain of $|G|$ pixels, and $Y_i \in \mathbb{R}_+$ is the observed intensity at location \mathbf{x}_i . We suppose the errors ε_i to be independent, distributed Gaussian zero-mean random variables with unknown variances σ^2 . In order to recover $u : \mathbb{R}^d \rightarrow \mathbb{R}$ from noisy observations, we need minimal prior assumptions on the structure of the image. In particular, we assume that the unknown image $u(\mathbf{x}_i)$ can be calculated as the weighted average of input data over a variable neighborhood Δ_i around that pixel \mathbf{x}_i . The points $\mathbf{x}_j \in \Delta_i$ with a similar regularized patch \mathbf{u}_j to the reference regularized image patch \mathbf{u}_i will have larger weights in the average. It amounts to the assumption that there exists repetitive patterns in the local neighborhood of a point \mathbf{x}_i which can help to recover $u(\mathbf{x}_i)$. However, our ambition is not to learn generic image priors from a database of image patches as proposed in [21], [47], [59]. We only focus on image patches as nonlocal image features, and adapt kernel regression techniques for image denoising.

For simplicity, an image patch \mathbf{u}_i is modeled as a fixed size square window of $p \times p$ pixels centered at \mathbf{x}_i . In what follows, \mathbf{u}_i will denote indifferently a patch or a vector of p^2 elements where the pixels are concatenated along a fixed lexicographic ordering. As with all patch-based techniques, the size of image patches must be specified in advance [15], [18], [58]. Traditionally, the size of the image patch is a parameter-free that specifies how stochastic the user believes the image to be. However, we shall see that a patch size of 7×7 or 9×9 pixels is able to take care of the local geometries and textures in the image while removing undesirable distortions. Finally, the proposed approach requires no training step and may be then considered as *unsupervised*. This makes the method somewhat more attractive for many image processing applications.

Another important question under such an estimation approach is how to determine the size and shape of the variable neighborhood Δ_i at each pixel, from image data. The selected window must be different at each pixel to take into account the inhomogeneous smoothness of the image. The choice of the set \mathcal{N}_Δ of candidate neighborhoods will play the key role. For the sake of simplicity and computational efficiency, the set of admissible neighborhoods will be arbitrarily chosen as a geometric grid of nested square windows

$$\mathcal{N}_\Delta = \left\{ \Delta_{i,n} : |\Delta_{i,n}| = (2^n + 1) \times (2^n + 1), \right. \\ \left. n = 1, \dots, N_\Delta \right\}$$

where $|\Delta_{i,n}| = \#\{\mathbf{x}_j \in \Delta_{i,n}\}$ is the cardinality of $\Delta_{i,n}$ and N_Δ is the number of elements of \mathcal{N}_Δ . For technical reasons, we will require the following conditions: $\Delta_{i,n}$ is centered at \mathbf{x}_i and $\Delta_{i,n} \subset \Delta_{i,n+1}$. In Section III, we will describe a local window selector which achieves two objectives: spatial adaptivity and computational efficiency. We will introduce the notion of *local* L_2 risk as an objective criterion to guide the optimal selection of the smoothing window for constructing the “best” possible estimator. This optimization will be mainly accomplished by starting, at each pixel, with a small window as a *pilot* estimate, and growing $\Delta_{i,n}$ with n .

III. ADAPTIVE ESTIMATION PROCEDURE

The proposed procedure is iterative and works as follows [30], [45]. At the initialization, we choose a local window $\Delta_{i,0}$ containing only the point of estimation \mathbf{x}_i ($|\Delta_{i,0}| = 1$). A first estimate $\hat{u}_{i,0}$ (and its variance $\hat{v}_{i,0}^2 = \text{Var}(\hat{u}_{i,0})$) is then given by

$$\hat{u}_{i,0} = Y_i \quad \text{and} \quad \hat{v}_{i,0}^2 = \hat{\sigma}^2 \quad (7)$$

where an estimated variance $\hat{\sigma}^2$ has been plugged in place of σ^2 since the variance of errors are supposed to be unknown (see Section III-D). At the next iteration, a larger window $\Delta_{i,1}$ with $\Delta_{i,0} \subset \Delta_{i,1}$ centered at \mathbf{x}_i is considered. Every point \mathbf{x}_j from $\Delta_{i,1}$ gets a weight¹ $\pi_{i \sim j,1}$ defined by comparing pairs of $p \times p$ regularized patches $\hat{\mathbf{u}}_{i,0} = \left(\hat{u}_{i,0}^{(1)}, \dots, \hat{u}_{i,0}^{(p^2)} \right)^T$ and

¹Here, the subscript $i \sim j$ means “ $\mathbf{x}_j \in \Delta_{i,\cdot}$, and the index j runs through the neighborhood of \mathbf{x}_i .”

$\hat{\mathbf{u}}_{j,0} = (\hat{u}_{j,0}^{(1)}, \dots, \hat{u}_{j,0}^{(p^2)})^T$ obtained at the first iteration. Note that p is fixed for all the pixels in the image. As usual, the points \mathbf{x}_j with a similar patch to $\hat{\mathbf{u}}_{i,0}$ will have weights close to 1 and 0 otherwise. Then we recalculate the estimate $\hat{\mathbf{u}}_{i,1}$ as the *weighted average* of data points lying in the neighborhood $\Delta_{i,1}$. We continue this way, growing with n the considered window $\Delta_{i,n}$ while $n \leq N_\Delta$ where N_Δ denotes the maximal number of iterations of the algorithm.

For each $n \geq 1$, the studied maximum likelihood (ML) estimator $\hat{\mathbf{u}}_{i,n}$ and its variance $\hat{v}_{i,n}^2$ can be then represented as

$$\hat{\mathbf{u}}_{i,n} = \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i \sim j, n} Y_j, \quad \hat{v}_{i,n}^2 = \hat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n}} [\pi_{i \sim j, n}]^2 \quad (8)$$

where the weights $\pi_{i \sim j, n}$ are continuous variables and satisfy the usual conditions $0 \leq \pi_{i \sim j, n} \leq 1$ and $\sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i \sim j, n} = 1$. In our modeling, they are computed from pairs of regularized $p \times p$ patches $\hat{\mathbf{u}}_{i,n-1}$ and $\hat{\mathbf{u}}_{j,n-1}$ obtained at iteration $n-1$ and p is fixed for all the pixels in the image. In what follows, n will coincide with the iteration and we will use $\hat{n}(\mathbf{x}_i)$ to designate the index of the “best” window $\hat{\Delta}(\mathbf{x}_i) \stackrel{\text{def}}{=} \hat{\Delta}_{i, \hat{n}(\mathbf{x}_i)}$ and the “best” estimate $\hat{\mathbf{u}}(\mathbf{x}_i) \stackrel{\text{def}}{=} \hat{\mathbf{u}}_{i, \hat{n}(\mathbf{x}_i)}$. Finally, among all nonrejected window $\Delta_{i,n}$ from \mathcal{N}_Δ , the optimal window is chosen as

$$\hat{\Delta}(\mathbf{x}_i) = \arg \max_{\Delta_{i,n} \in \mathcal{N}_\Delta} \left\{ |\Delta_{i,n}| : |\hat{\mathbf{u}}_{i,n} - \hat{\mathbf{u}}_{i,n'}| \leq \varrho \hat{v}_{i,n'} \right. \\ \left. \text{for all } 1 \leq n' < n \right\}$$

where ϱ is a positive constant. In the sequel, we will give some cues for choosing p and the threshold ϱ for practical imaging. Also, throughout this paper, we shall see the rational behind this pointwise statistical rule and the proposed strategy that updates the estimator when the neighborhood increases at each iteration. At this level, we just point out that this pointwise stopping rule guarantees the estimator is optimal in the sense that it enables to approximately minimize the pointwise L_2 risk—i.e., $[\mathbb{E}|\hat{\mathbf{u}}_{i,n} - \mathbf{u}(\mathbf{x}_i)|^2]^{1/2}$ —of the estimator (see Section III-B). With this adaptive choice of window which depends on the observations instead of a usual deterministic window and defined patch-based weights, this estimator is clearly not linear. Accordingly, the limit image cannot be easily predicted and mainly depends on control parameters. We can just confirm that the recovered image will be more regular than the input noisy image but will contain more discontinuities than the image smoothed with a rectangular window of $(2^{N_\Delta} + 1) \times (2^{N_\Delta} + 1)$ pixels. Moreover, for mathematical convenience, we decided to use a rectangular spatial window but the method can be easily extended to the case of a more usual Gaussian spatial window to give more influence to nearby pixels and to make the denoising method invariant to image rotation. Finally, the use of variable and overlapping windows contributes to the regularization performance with no block effect, enhances the flexibility of the resulting local regularizers and make them possible to cope well with spatial inhomogeneities in natural images.

A. Adaptive Weights

In order to compute the similarity between patches $\hat{\mathbf{u}}_{i,n}$ and $\hat{\mathbf{u}}_{j,n}$, a distance must be considered. In [7], [15], [18], and [58],

several authors showed that the L_2 distance $\|\hat{\mathbf{u}}_{i,n} - \hat{\mathbf{u}}_{j,n}\|^2$ is a reliable measure to compare image patches. To make a decision, we have rather used the following normalized distance

$$\text{dist}(\hat{\mathbf{u}}_{i,n-1}, \hat{\mathbf{u}}_{j,n-1}) = \frac{1}{2} \left[(\hat{\mathbf{u}}_{i,n-1} - \hat{\mathbf{u}}_{j,n-1})^T \hat{\mathbf{V}}_{i,n-1}^{-1} (\hat{\mathbf{u}}_{i,n-1} - \hat{\mathbf{u}}_{j,n-1}) \right. \\ \left. + (\hat{\mathbf{u}}_{j,n-1} - \hat{\mathbf{u}}_{i,n-1})^T \hat{\mathbf{V}}_{j,n-1}^{-1} (\hat{\mathbf{u}}_{j,n-1} - \hat{\mathbf{u}}_{i,n-1}) \right] \quad (9)$$

where $\hat{\mathbf{V}}_{\cdot, n-1}$ is $p^2 \times p^2$ diagonal matrix of the form (the symbol “.” is used to denote a spatial position)

$$\hat{\mathbf{V}}_{\cdot, n-1} = \begin{pmatrix} (\hat{v}_{\cdot, n-1}^{(1)})^2 & 0 & \dots & 0 \\ 0 & (\hat{v}_{\cdot, n-1}^{(2)})^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & (\hat{v}_{\cdot, n-1}^{(p^2)})^2 \end{pmatrix}$$

where $\hat{v}_{\cdot, n-1}^{(\ell)}$, $\ell = 1, \dots, p^2$, is the local standard deviation of the estimator $\hat{\mathbf{u}}_{\cdot, n-1}^{(\ell)}$, and the index ℓ is used to denote a spatial position in an image patch $\hat{\mathbf{u}}_{\cdot, n-1} = (\hat{u}_{\cdot, n-1}^{(1)}, \dots, \hat{u}_{\cdot, n-1}^{(\ell)}, \dots, \hat{u}_{\cdot, n-1}^{(p^2)})^T$. Moreover, we used a symmetrized distance to test both the hypotheses that \mathbf{x}_j belongs to the region $\Delta_{i,n}$ centered at \mathbf{x}_i and \mathbf{x}_i belongs to the region $\Delta_{j,n}$ centered in \mathbf{x}_j , at the same time. Accordingly, the hypothesis $\hat{\mathbf{u}}_{i,n-1}$ and $\hat{\mathbf{u}}_{j,n-1}$ are similar, is accepted if the distance is small, i.e., $\text{dist}(\hat{\mathbf{u}}_{i,n-1}, \hat{\mathbf{u}}_{j,n-1}) \leq \lambda_\alpha$. In our modeling, the parameter $\lambda_\alpha \in \mathbb{R}_+$ will be chosen as a quantile of a $\chi_{p^2, 1-\alpha}^2$ distribution with p^2 degrees of freedom, and controls the probability of type I error for the hypothesis of two points to belong to the same region

$$\mathbb{P}\{\text{dist}(\hat{\mathbf{u}}_{i,n-1}, \hat{\mathbf{u}}_{j,n-1}) \leq \lambda_\alpha\} = 1 - \alpha. \quad (10)$$

All these tests ($|\Delta_{i,n}|$ tests) have to be performed at a very high significance level, our experience suggesting to use a $1 - \alpha = 0.99$ -quantile. If $\text{dist}(\hat{\mathbf{u}}_{i,n-1}, \hat{\mathbf{u}}_{j,n-1})$ exceeds this critical threshold λ_α , then we have a significant difference between $\hat{\mathbf{u}}_{i,n-1}$ and $\hat{\mathbf{u}}_{j,n-1}$ and we reject the hypothesis that \mathbf{x}_i and \mathbf{x}_j belong to the same region. In other words, if $\text{dist}(\hat{\mathbf{u}}_{i,n-1}, \hat{\mathbf{u}}_{j,n-1})$ is large enough to have a probability smaller than 1% ($1 - \alpha = 0.01$), we suggest that the test gave a result that was significant at the 1% level, and we decide to reject the hypothesis that $\hat{\mathbf{u}}_{i,n-1}$ and $\hat{\mathbf{u}}_{j,n-1}$ are coming from the same region.

Henceforth, we introduce the following commonly used weight function

$$\pi_{i \sim j, n} = \frac{K(\lambda_\alpha^{-1} \text{dist}(\hat{\mathbf{u}}_{i,n-1}, \hat{\mathbf{u}}_{j,n-1}))}{\sum_{\mathbf{x}_j \in \Delta_{i,n}} K(\lambda_\alpha^{-1} \text{dist}(\hat{\mathbf{u}}_{i,n-1}, \hat{\mathbf{u}}_{j,n-1}))} \quad (11)$$

with $K(\cdot)$ denoting a monotone decreasing function, e.g., a kernel $K(z) = \exp(-z/2)$. Due to the fast decay of the exponential kernel, large distances between estimated patches lead to nearly zero weights. Note that the use of weights enables

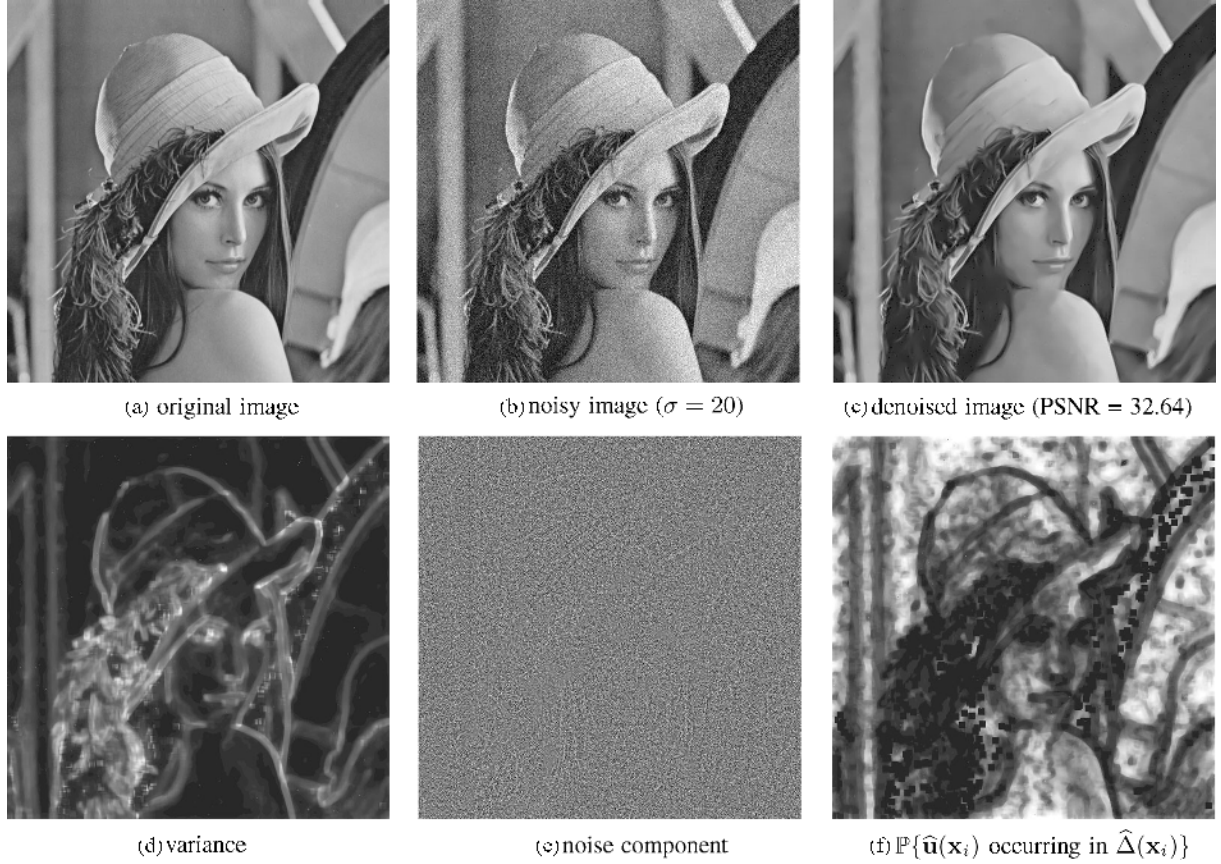


Fig. 1. Decomposition of the noisy *Lena* image ($\sigma = 20$).

to relax the structural assumption that the neighborhood is roughly modeled by a square window.

B. “Ideal” Smoothing Window

In this section, we address the problem of the automatic selection of the window $\Delta_{i,\cdot}$, adapted for each pixel \mathbf{x}_i . It is well understood that the local smoothness varies significantly from point to point in the image and *global risk* measures cannot wholly reflect the performance of estimators at a point. Then, a classical way to measure the performance of the estimator $\hat{u}_{i,n}$ to its target value $u(\mathbf{x}_i)$ is to choose the *local L_2 risk*, which is explicitly decomposed into the sum of the squared bias $\hat{b}_{i,n}^2$ and variance $\hat{v}_{i,n}^2$ as

$$\mathbb{E} \left[|\hat{u}_{i,n} - u(\mathbf{x}_i)|^2 \right]^{1/2} = \left| \hat{b}_{i,n}^2 + \hat{v}_{i,n}^2 \right|^{1/2}. \quad (12)$$

Our goal is to minimize this *local L_2 risk* with respect to the size of the window $\Delta_{i,n}$, at each pixel in the image. Actually, the optimal solution explicitly depends on the smoothness of the “true” function $u(\mathbf{x}_i)$ which is unknown, and so, of less practical interest [29], [53]. A natural way to bring some further understanding of the situation is then to individually analyze the behavior of the bias and variance terms when $\Delta_{i,n}$ increases or decreases with n as follows.

- The bias term $\hat{b}_{i,n} = \mathbb{E}[\hat{u}_{i,n} - u(\mathbf{x}_i)]$ is nonrandom and characterizes the accuracy of approximation of

Algorithm Patch-based image denoising algorithm

Let $\{p, \alpha, \rho, N_\Delta\}$ be the parameters.

Initialization: compute $\hat{\sigma}^2$ and $\{\hat{u}_{i,0}, \hat{v}_{i,0}^2\}$ for each $\mathbf{x}_i \in G$ and set $n = 1$.

Repeat

- for each $\mathbf{x}_i \in G$
 - compute

$$\begin{aligned} \pi_{i \sim j, n} &= \frac{K(\lambda_\alpha^{-1} \text{dist}(\hat{\mathbf{u}}_{i, n-1}, \hat{\mathbf{u}}_{j, n-1}))}{\sum_{\mathbf{x}_j \in \Delta_{i, n}} K(\lambda_\alpha^{-1} \text{dist}(\hat{\mathbf{u}}_{i, n-1}, \hat{\mathbf{u}}_{j, n-1}))} \\ \hat{u}_{i, n} &= \sum_{\mathbf{x}_j \in \Delta_{i, n}} \pi_{i \sim j, n} Y_j, \\ \hat{v}_{i, n}^2 &= \hat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i, n}} [\pi_{i \sim j, n}]^2 \end{aligned}$$

- choose the window as

$$\hat{\Delta}(\mathbf{x}_i) = \arg \max_{\Delta_{i, n} \in \mathcal{N}_\Delta} \{|\Delta_{i, n}| : |\hat{u}_{i, n} - \hat{u}_{i, n'}| \leq \rho \hat{v}_{i, n'}, \text{ for all } 1 \leq n' < n\}.$$

If this rule is violated at iteration n , we do not accept $\hat{u}_{i, n}$ and keep the estimate $\hat{u}_{i, n-1}$ as the final estimate at \mathbf{x}_i , i.e. $\hat{u}(\mathbf{x}_i) \leftarrow \hat{u}_{i, n-1}$ and $\hat{v}(\mathbf{x}_i) = n-1$. This estimate is unchanged at the next iterations and \mathbf{x}_i is “frozen”.

- increment n
- while $n \leq N_\Delta$.
-

Fig. 2. Patch-based image denoising algorithm.

the function u at the point \mathbf{x}_i . As it explicitly depends on the unknown function $u(\mathbf{x}_i)$, its behavior is doubtful. Nevertheless, if we use the geometric inequality

$|\mathbf{x}_j - \mathbf{x}_i| \leq (\sqrt{2}/2)|\Delta_{i,n}|^{1/2}$ and assume that there exists a real constant $0 < C_1 < \infty$ (i.e., u is Lipschitz continuous) such that $|u(\mathbf{x}_j) - u(\mathbf{x}_i)| \leq C_1|\mathbf{x}_j - \mathbf{x}_i|$, then

$$\begin{aligned} |\hat{b}_{i,n}| &= \left| \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i \sim j, n} [\mathbb{E}[Y_j] - u(\mathbf{x}_i)] \right| \\ &\leq \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i \sim j, n} |u(\mathbf{x}_j) - u(\mathbf{x}_i)| \\ &\leq C_1 \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i \sim j, n} |\mathbf{x}_j - \mathbf{x}_i| \\ &\leq \frac{C_1}{\sqrt{2}} |\Delta_{i,n}|^{1/2}. \end{aligned}$$

Accordingly, $|\hat{b}_{i,n}|^2$ is of the order $O(|\Delta_{i,n}|)$ and typically increases when $\Delta_{i,n}$ grows (see also [29]).

- The behavior of the variance term is just opposite. The errors are independent and the stochastic term $\hat{v}_{i,n}^2$ can be exactly computed on the basis of observations. Since $0 \leq \pi_{i \sim j, n} \leq 1$ and $\sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i \sim j, n} = 1$, it follows that:

$$\frac{\hat{\sigma}^2}{|\Delta_{i,n}|} \leq \hat{v}_{i,n}^2 \leq \hat{\sigma}^2.$$

In addition, we can reasonably assume that there exists a constant $0 \leq \gamma^2 \leq 1$ such that $\hat{v}_{i,n}^2 \approx \hat{\sigma}^2 |\Delta_{i,n}|^{-\gamma^2}$. Accordingly, as $\Delta_{i,n}$ grows, more data is used to construct the estimate $\hat{u}_{i,n}$, and so $\hat{v}_{i,n}^2$ decreases.

Therefore, the bias and standard deviation terms are monotonous functions with opposite behavior. In order to approximately minimize the local L_2 risk of the estimator with respect to $|\Delta_{i,n}|$, a natural idea would be to minimize an upper bound of the form

$$\mathbb{E}|\hat{u}_{i,n} - u(\mathbf{x}_i)|^2 \leq \frac{C_1^2}{2} |\Delta_{i,n}| + \frac{\hat{\sigma}^2}{|\Delta_{i,n}|^{\gamma^2}}.$$

An approximation of the optimal window size can be then obtained as

$$|\Delta^*(\mathbf{x}_i)| = \arg \min_{|\Delta_{i,n}|} \frac{C_1^2}{2} |\Delta_{i,n}| + \frac{\hat{\sigma}^2}{|\Delta_{i,n}|^{\gamma^2}}$$

which amounts to solving the following equation:

$$\frac{C_1^2}{2} - \frac{\gamma^2 \hat{\sigma}^2}{|\Delta_{i,n}|^{\gamma^2+1}} = 0.$$

However, the closed-form solution given by

$$|\Delta^*(\mathbf{x}_i)| = \left[\frac{2\gamma^2 \hat{\sigma}^2}{C_1^2} \right]^{1/(\gamma^2+1)}$$

cannot be used in practice since C_1 and γ are unknown. Nevertheless, for the optimal value $|\Delta^*(\mathbf{x}_i)|$, it can be easily shown that the ratio between the optimal bias $b^*(\mathbf{x}_i)$ and the optimal

standard deviation $v^*(\mathbf{x}_i)$ is image independent (see [29] and [53])

$$\frac{|b^*(\mathbf{x}_i)|}{v^*(\mathbf{x}_i)} \leq \gamma.$$

Accordingly, an ideal choice of the window will be the largest window $\Delta_{i,n}$ such that $\hat{b}_{i,n}$ is still not larger than $\gamma \hat{v}_{i,n}$, for some real value $0 \leq \gamma^2 \leq 1$. In practice, the bias is not observable and γ is unknown. Henceforth, we can adopt the following strategy which consists in decomposing the estimator $\hat{u}_{i,n}$ as [29], [35], [53]

$$\hat{u}_{i,n} = u(\mathbf{x}_i) + \hat{b}_{i,n} + \varsigma_i \quad (13)$$

where $\varsigma_i \sim \mathcal{N}(0, \mathbb{E}[\varsigma_i^2])$ is a stochastic component. Therefore, $\mathbb{E}[\hat{u}_{i,n}] = u(\mathbf{x}_i) + \hat{b}_{i,n}$ and the use of (12) yields

$$\begin{aligned} \mathbb{E}[\varsigma_i^2] &= \mathbb{E}[|\hat{u}_{i,n} - u(\mathbf{x}_i) - \hat{b}_{i,n}|^2] \\ &= \mathbb{E}[|\hat{u}_{i,n} - u(\mathbf{x}_i)|^2] - 2\mathbb{E}[\hat{u}_{i,n} - u(\mathbf{x}_i)]\hat{b}_{i,n} + \hat{b}_{i,n}^2 \\ &= \mathbb{E}[|\hat{u}_{i,n} - u(\mathbf{x}_i)|^2] - \hat{b}_{i,n}^2 \\ &\stackrel{\text{def}}{=} \hat{v}_{i,n}^2. \end{aligned}$$

By definition, the following inequality:

$$|\hat{u}_{i,n} - u(\mathbf{x}_i)| \leq \hat{b}_{i,n} + \varkappa \hat{v}_{i,n} \quad (14)$$

then holds with a high probability and $0 < \varkappa < +\infty$. Combining (14) and the inequality $\hat{b}_{i,n} \leq \gamma \hat{v}_{i,n}$ yields

$$|\hat{u}_{i,n} - u(\mathbf{x}_i)| \leq (\gamma + \varkappa) \hat{v}_{i,n}. \quad (15)$$

Finally, we modify correspondingly the definition of the *ideal* window as

$$\Delta^*(\mathbf{x}_i) = \sup_{\Delta_{i,n} \in \mathcal{N}_\Delta} \{|\Delta_{i,n}| : |\hat{u}_{i,n} - u(\mathbf{x}_i)| \leq (\gamma + \varkappa) \hat{v}_{i,n}\}. \quad (16)$$

The crucial point is that this inequality depends no longer on $\hat{b}_{i,n}$, but is yet related to the unknown function $u(\mathbf{x}_i)$. In Section III-C, we shall see that a *data-driven* window selector based on this definition of $\Delta^*(\mathbf{x}_i)$ can actually be derived.

C. Data-Driven Local Window Selector

In our approach, the collection of estimators $\{\hat{u}_{i,1}, \dots, \hat{u}_{i,n}\}$ is naturally ordered in the direction of increasing $|\Delta_{i,n}|$ where $\hat{u}_{i,n}$ can be thought as the best possible estimator with the smallest variance. The estimator $\hat{u}_{i,0} = Y_i$ is also discarded since its variance is too high. A selection procedure can be then described based on pairwise comparisons of an essentially one-dimensional family of competing estimators $\hat{u}_{i,n}$. In this modeling, the differences $\hat{u}_{i,n} - \hat{u}_{i,n'}$ are Gaussian random variables with known variances $\text{Var}(\hat{u}_{i,n} - \hat{u}_{i,n'}) \leq \hat{v}_{i,n'}^2$ with $1 \leq n' < n$ (see [31, Appendix A1] for the proof), and expectations equal to



Fig. 3. Estimated noise components by applying restoration methods to the noisy (WGN) *Lena* image ($\sigma = 20$). (a) Total variation (TV) minimizing process [48]; (b) bilateral filtering (BF) [55]; (c) anisotropic diffusion (AD) [43]; (d) Wiener filtering (WF); (e) robust adaptive window approach (RAWA) [30]; (f) adaptive weights smoothing (AWS) [45]; (g) *nonlocal means*; (NLMMeans) [7]; (h) wavelet-based denoising (WBD) [44]; (i) our patch-based denoising method.

the bias differences $\hat{b}_{i,n} - \hat{b}_{i,n'}$. From the definition of $\Delta^*(\mathbf{x}_i)$ [see (16)], we derive

$$|\hat{u}_{i,n'} - \hat{u}_{i,n}| \leq (2\gamma + \kappa)\hat{v}_{i,n'}, \quad 1 \leq n' < n$$

and, among all good candidates $\hat{u}_{i,n}$ satisfying this inequality, one chooses the one with the smallest variance $\hat{v}_{i,n}^2$. Following the above discussion, a window selector will be then based on the following pointwise rule [26], [28], [34], [35]

$$\hat{\Delta}(\mathbf{x}_i) = \arg \max_{\Delta_{i,n} \in \mathcal{N}_{\Delta}} \left\{ |\Delta_{i,n}| : |\hat{u}_{i,n} - \hat{u}_{i,n'}| \leq \varrho \hat{v}_{i,n'} \right. \\ \left. \text{for all } 1 \leq n' < n \right\}$$

where $\varrho = (2\gamma + \kappa)$. This rule actually ensures the balance between the stochastic term and the bias term, and means that we take the largest window such that the estimators $\hat{u}_{i,n}$ and $\hat{u}_{i,n'}$ are not too different, in some sense, for all $1 \leq n' < n$. Hence, if an estimated point $\hat{u}_{i,n'}$ appears far from the previous ones, this means that the bias is already too large and the window $\Delta_{i,n}$ is not a good one. For each pixel, the detection of this transition enables to determine the critical size of the window. Also, this rule updates the intersection of estimated confidence intervals at each iteration and tests if a new candidate estimator belongs to this current confidence interval, for each point. This idea underlying our construction definitely belongs to Lepskii [34], [35].

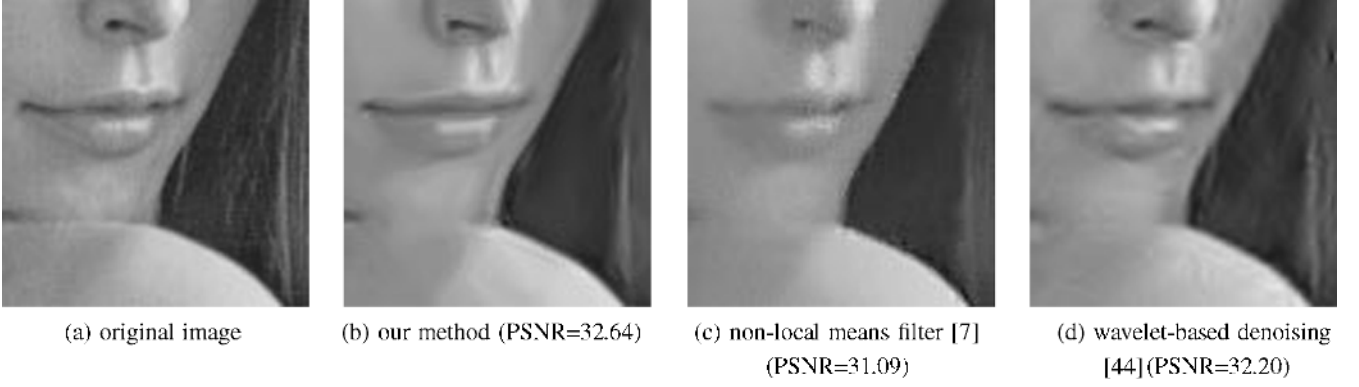


Fig. 4. Comparisons with the *nonlocal means* filter [7] and a wavelet-based denoising method [44] when applied to the noisy (WGN) *Lena* image ($\sigma = 20$).

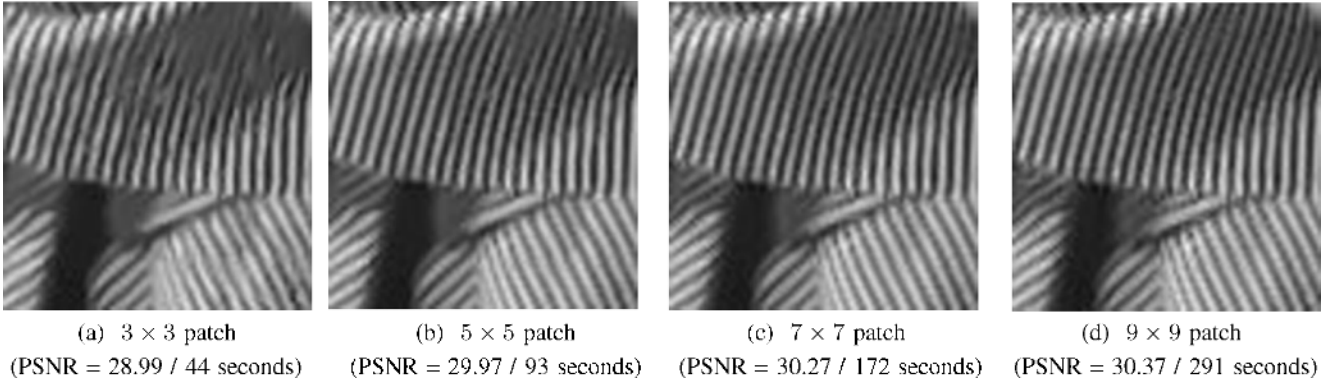


Fig. 5. Performances (PSNR values and timings in seconds) for different patch sizes when the method ($\alpha = 0.01$, $\varrho = 3$, and $N_\Delta = 4$) is applied to the noisy (WGN) 512×512 *Barbara* image ($\sigma = 20$).

D. Implementation

The key ingredient of the estimation procedure is an increasing sequence of nested square windows. At the initialization, we naturally choose $|\Delta_{i,0}| = 1$, set the fixed size of $p \times p$ patches and choose the number of iterations N_Δ . In addition, the estimation procedure relies on the preliminary estimation of the noise variance $\hat{\sigma}^2$ robustly estimated from input data as [5]

$$\hat{\sigma} = 1.4826 \operatorname{med}(|\mathbf{r}| - \operatorname{med}|\mathbf{r}|)$$

where $\mathbf{r} = \{r_1, r_2, \dots, r_{|G|}\}$ is the set of local residuals of the entire image defined as [we note Y_{i_1, i_2} the observation Y_i at point $i = (i_1, i_2)$]

$$r_i = \frac{[2Y_{i_1, i_2} - (Y_{i_1+1, i_2} + Y_{i_1, i_2+1})]}{\sqrt{6}}$$

and the constant $1/\sqrt{6}$ is used to ensure $\mathbb{E}[r_i^2] = \hat{\sigma}^2$ in homogeneous regions. To complete the procedure, we choose $\varrho \in [2, 3]$ in order to get a good accuracy for the pointwise estimator (see [31] for more details) and λ_α as a $1 - \alpha = 0.99$ -quantile of a $\chi_{p^2, 1-\alpha}^2$ distribution. Once the ϱ and λ_α parameters are determined using these statistical arguments, the remainder of the algorithm given in Fig. 2, is completely automatic. Finally, the

complexity of the whole procedure is bounded and of the order $p \times p \times |G| \times (|\Delta_{\cdot, 1}| + \dots + |\Delta_{\cdot, N_\Delta}|)$ if the image contains $|G|$ pixels.

IV. EXPERIMENTAL RESULTS

Our results were measured by the peak signal-to-noise ratio (PSNR) in decibels (dB) defined as

$$\text{PSNR} = 10 \log_{10} \frac{255^2}{\text{MSE}}, \quad \text{MSE} = \frac{\sum_{\mathbf{x}_i \in G} (u_o(\mathbf{x}_i) - \hat{u}(\mathbf{x}_i))^2}{|G|}$$

where u_0 is the noise-free original image. We have done simulations on a commonly-used set of images available at http://decsai.ugr.es/~javier/denoise/test_images/ and described in [46]. In all our experiments, we have chosen image patches of 9×9 pixels and set the algorithm parameters as follows: $\lambda_{0.01} = \chi_{81, 0.99}^2 = 113.5$, $\varrho = 3$ and $N_\Delta = 4$ (see also [31]). The processing of a 256×256 image required typically about 1 minute ($p = 9$) on a PC (2.0 Ghz, Pentium IV) using a C++ implementation of the algorithm.

The potential of the estimation method is mainly illustrated with the 512×512 *Lena* image [Fig. 1(a)] corrupted by an additive white-Gaussian noise (WGN) [Fig. 1(b)], PSNR=22.13 dB, $\sigma = 20$). In Fig. 1(c), the noise is reduced in a natural manner and significant geometric features, fine textures, and original contrasts are visually well recovered with no undesirable artifacts (PSNR = 32.64 dB). The noise component is shown in Fig. 1(e) (magnification factor of 2) and has been estimated by

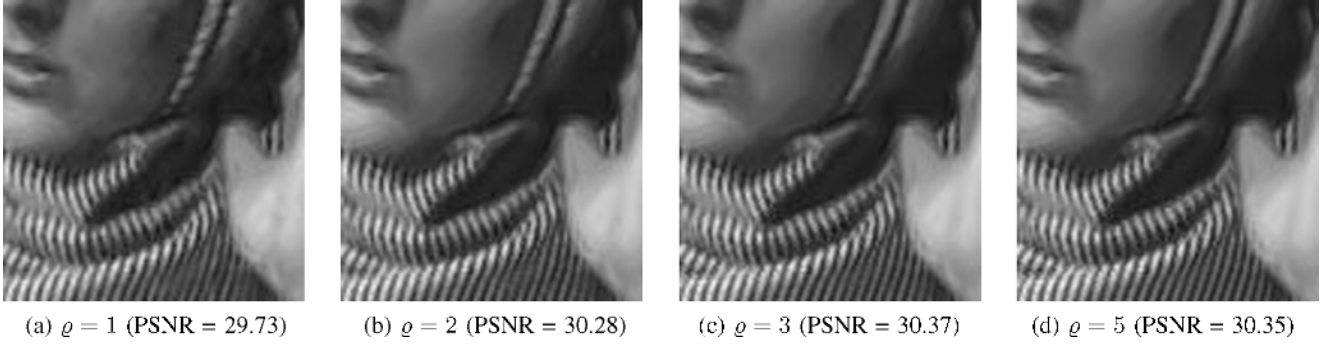


Fig. 6. Performances (PSNR values) for different values for ρ when the method ($\alpha = 0.01$, 9×9 patch and $N_{\Delta} = 4$) is applied to the noisy (WGN) 512×512 Barbara image ($\sigma = 20$).



Fig. 7. Performances (PSNR values) for different values for α when the method ($\rho = 3$, 9×9 patch and $N_{\Delta} = 4$) is applied to the noisy (WGN) 512×512 Barbara image ($\sigma = 20$).

calculating the difference between the noisy image [Fig. 1(b)] and the recovered image [Fig. 1(c)]. The estimated noise component contains few geometric structures and is similar to a simulated white Gaussian noise. To better appreciate the accuracy of the denoising process, the variance of the pointwise estimator is shown in Fig. 1(d) where dark values correspond to high-confidence estimates. As expected, pixels with a low level of confidence are located in the neighborhood of image discontinuities. Fig. 1(f) shows the probability of a patch $\hat{\mathbf{u}}(\mathbf{x}_i)$ occurring in $\hat{\Delta}(\mathbf{x}_i)$, i.e.,

$$\mathbb{P}\{\hat{\mathbf{u}}(\mathbf{x}_i) \text{ occurring in } \hat{\Delta}(\mathbf{x}_i)\} \stackrel{\text{def}}{=} \frac{\#\Omega(\mathbf{x}_i)}{|\hat{\Delta}(\mathbf{x}_i)|}$$

where the set $\Omega(\mathbf{x}_i)$ is used to denote $\{\mathbf{x}_j \in \hat{\Delta}(\mathbf{x}_i) : \text{dist}(\hat{\mathbf{u}}(\mathbf{x}_i), \hat{\mathbf{u}}(\mathbf{x}_j)) \leq \lambda_{\alpha}\}$. This amounts to count the number of detected image patches in the optimal window $\hat{\Delta}(\mathbf{x}_i)$ which are similar to the central optimal patch centered at \mathbf{x}_i . In Fig. 1(f), dark values correspond low probabilities of occurrence and, it is confirmed that repetitive patterns in the neighborhood of image discontinuities are located along image level lines.

We have compared the performance of our method to several competitive methods: total variation (TV) minimizing process

[48], bilateral filtering (BF) [55], anisotropic diffusion (AD) using a diffusivity function of the type $(1 + (|\nabla u|^2)/K^2)^{-1}$ [43], and Wiener filtering (WF) (Matlab function wiener2). Fig. 3(a)–(d) shows the results of the four tested methods. The global control parameters of these algorithms were tuned (we have to try several values) to both eliminate noise and simultaneously to get the best PSNR value, and to give a good visual impression. Additionally, this noisy image has been restored using pointwise adaptive estimation methods [30], [45] which are not patch-based. Fig. 3(e)–(f) provides a visual comparison of image denoising with these two algorithms: the AWS algorithm [45] tends to oversmooth the image and to generate some artificial planar segments in homogeneous regions [Fig. 3(f)], whereas a variant of this approach (RAWA) [30] yields a similar result [Fig. 3(e)] to the image regularizing with the TV method [48] [see Fig. 3(a)]. Moreover, our approach is also compared to the *nonlocal means* algorithm (NLMeans) [7], [8] using 7×7 image patches and a fixed search window of 21×21 pixels: The visual impression and the numerical results are improved using our algorithm [see Figs. 3(i) and 4]. Fig. 4 shows that most details (see the mouth and the nose) are visually better reconstructed with no artifact when compared to previous methods. Finally, we point out that, visually and quantitatively (see Fig. 8), our unsupervised algorithm method favorably compares to any of these denoising algorithms, including the

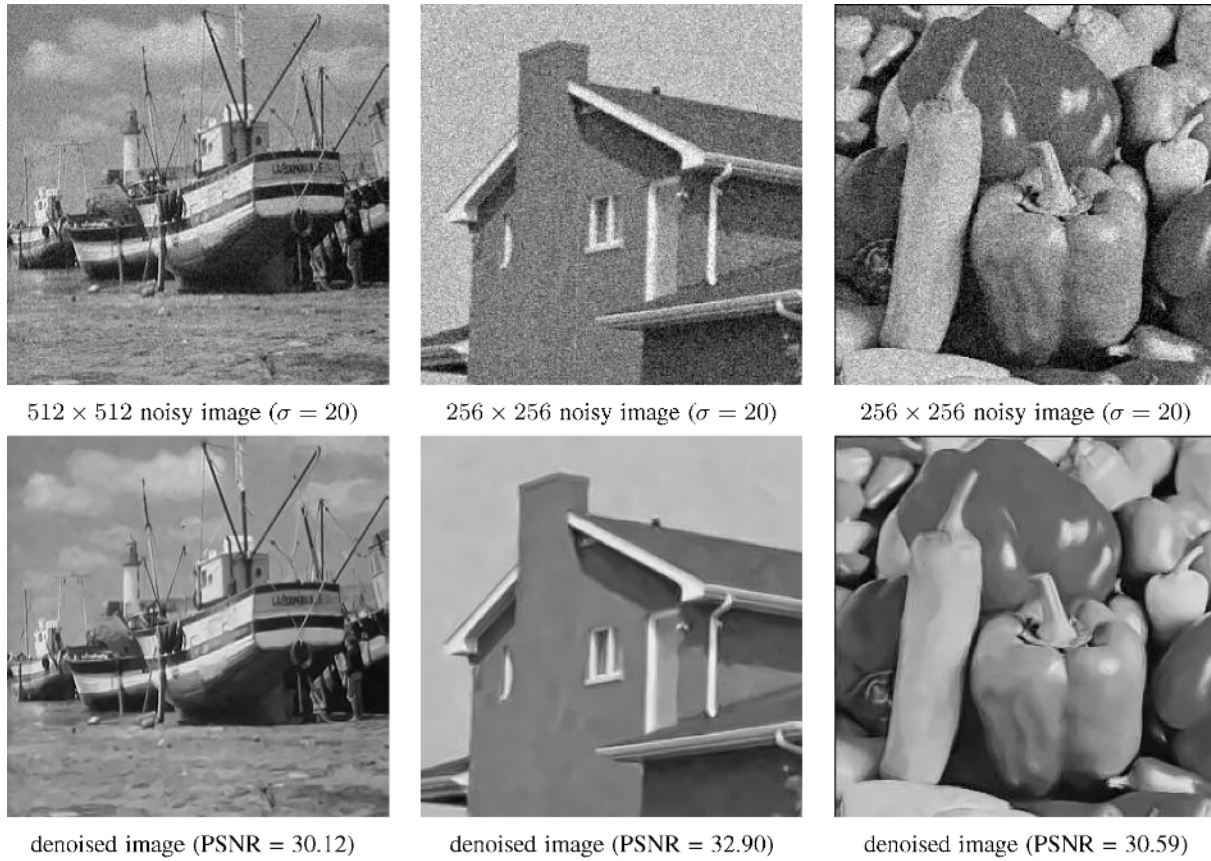
Fig. 8. Denoising of noisy (WGN) test images ($\sigma = 20$).

TABLE I
PERFORMANCES OF DENOISING ALGORITHMS WHEN APPLIED
TO TEST NOISY (WGN) IMAGES ($\sigma = 20$)

Image σ /PSNR	Lena 20/22.13	Barbara 20/22.18	Boats 20/22.17	House 20/22.11	Peppers 20/22.19
Our method (9×9 patch)	32.64	30.37	30.12	32.90	30.59
Buades et al. [7]	31.09	29.38	28.60	31.54	29.05
Ghazel et al. [22]	28.50	25.64	26.34	-	-
Kervrann [30]	30.54	26.50	28.01	30.70	28.23
Pizurica et al. [44]	32.20	29.53	29.93	-	30.30
Polzehl et al. [45]	29.74	26.05	27.74	30.31	28.40
Portilla et al. [46]	32.66	30.32	30.38	32.39	30.31
Roth et al. [47]	31.92	28.32	29.85	32.17	30.58
Rudin et al. [48]	30.48	27.07	29.02	31.03	28.51
Starck et al. [54]	31.95	-	-	-	-
Tomasi et al. [55]	30.26	27.02	28.41	30.01	28.88
Wiener filtering	28.51	26.99	27.97	28.74	28.10

more sophisticated wavelet-based denoising methods. We reported the best PSNR results we obtained using these methods in Table I. Both visually and in terms of PSNR, our method outperforms any of the tested methods. Indeed, we have also compared our method to the best available published results when very competitive methods [44], [46], [47] were applied to the same image dataset [46]. These results were taken from the

TABLE II
PSNR VALUES (DECIBELS) WHEN OUR PATCH-BASED DENOISING METHOD
($N_{\Delta} = 4$, $\alpha = 0.01$) WITH DIFFERENT PATCH SIZES AND SUB-SAMPLING
(FACTOR 2) IS APPLIED TO NOISY (WGN) IMAGES ($\sigma = 20$)

Patch size	$\lambda_{0.01}$	Lena 512 ²	Barbara 512 ²	Boats 512 ²	House 256 ²	Peppers 256 ²
9 pts/ 3×3	21.67	32.13	28.99	29.86	32.69	30.86
25 pts/ 5×5	44.31	32.52	29.97	30.15	33.05	30.98
49 pts/ 7×7	74.92	32.63	30.27	30.17	33.03	30.80
81 pts/ 9×9	113.5	32.64	30.37	30.12	32.90	30.59
25 pts/ 9×9	44.31	32.27	29.84	29.64	32.46	30.26
36 pts/ 11×11	58.62	32.26	29.84	29.51	32.57	29.52

corresponding publications. If the PSNR gains are marginal for some images, the visual difference can be significant as shown in Figs. 3(h)–(i) and 4 where less artifacts are visible using our method. To complete the experiments, Table II shows the PSNR values using our patch-based denoising method when applied to this set of test images for a wide range of noise variance. This table can be used for comparison purposes with previously published denoising methods [46], [47].

Moreover, we have also examined some complementary aspects of our approach. Table III shows the PSNR values obtained by varying the patch size and subsampling (factor 2). Note the PSNR values are close for every patch size and the optimal patch size depends on the image contents. Images with smooth parts can be satisfying regularized with small 3×3 or

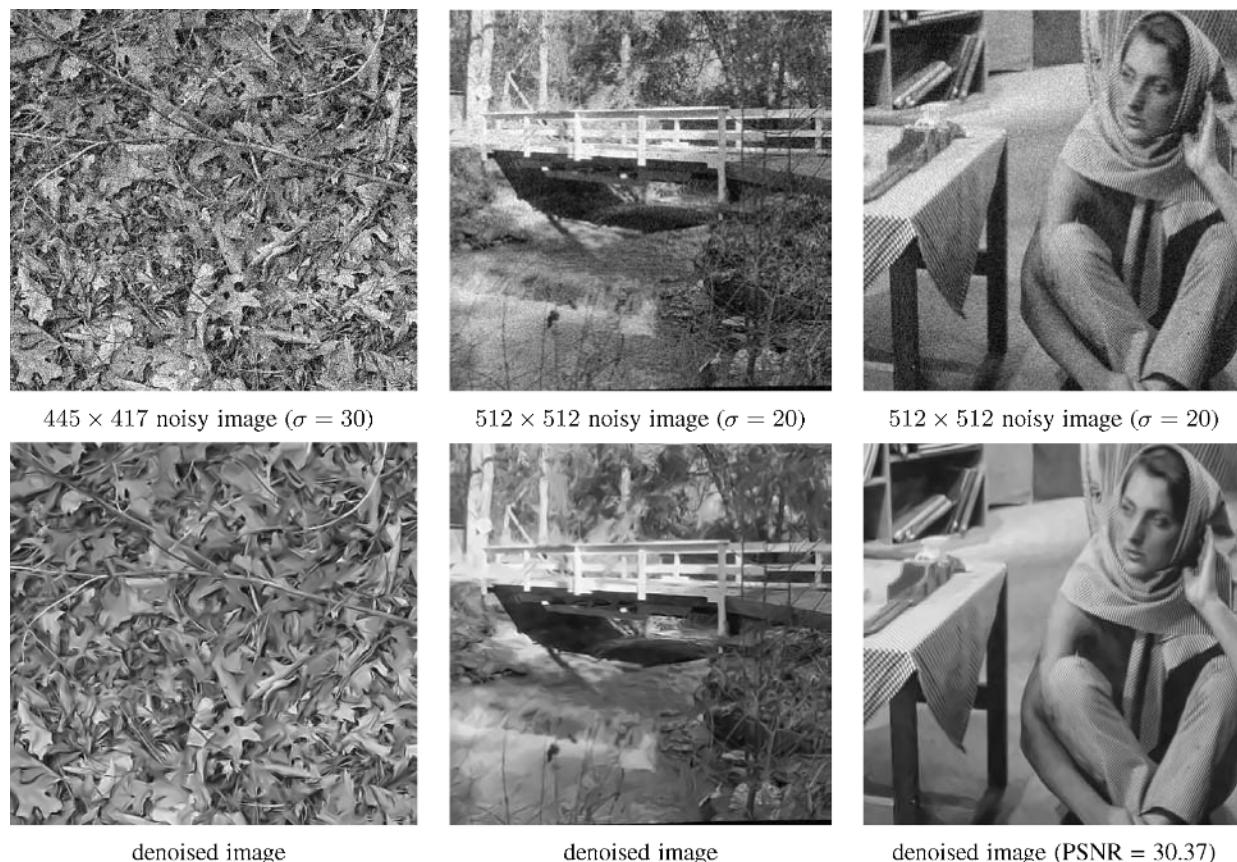
Fig. 9. Denoising of noisy (WGN) textured 512×512 images.

TABLE III
PERFORMANCES OF OUR DENOISING METHOD ($p = 9, N_{\Delta} = 4, \alpha = 0.01$)
WHEN APPLIED TO TEST NOISY (WGN) IMAGES

σ /PSNR	Lena	Barbara	Boats	House	Peppers
	512^2	512^2	512^2	256^2	256^2
5 / 34.15	37.91	37.12	36.14	37.62	37.34
10 / 28.13	35.18	33.79	33.09	35.26	34.07
15 / 24.61	33.70	31.80	31.44	34.08	32.13
20 / 22.11	32.64	30.37	30.12	32.90	30.59
25 / 20.17	31.73	29.24	29.20	32.22	29.73
50 / 14.15	28.38	24.09	25.93	28.67	25.29
75 / 10.63	25.51	22.10	23.69	25.49	22.31
100 / 8.13	23.32	20.64	21.78	23.08	20.51

5×5 image patches, the PSNR values being even higher than those obtained by using 9×9 patches, and the time computing is then reduced also. Some typical visual results obtained by varying p are shown in Fig. 5. Now, from our experiments, 9×9 image patches are probably more appropriate in most cases but smaller patches can be considered, especially for processing piecewise smooth images. Finally, from experiments shown in Figs. 6 and 7, it turns out the method is nearly parameter-free since by varying α and ρ in suitable ranges (ρ and α can be calibrated using statistical arguments as explained in [31]), image reconstruction is nearly unchanged.

The effects of the patch-based denoising approach are also illustrated on artificially corrupted textured images with an additive white-Gaussian noise. The set of parameters is unchanged for processing all these test images: $p = 9, N_{\Delta} = 4, \alpha = 0.01$. In most cases, a good compromise between the amount of smoothing and preservation of edges and textures is automatically reached (Fig. 9). In the last part of experiments, the patch-based denoising method has been used to restore a real noisy picture shown in Fig. 10(a). In that case, the noise variance $\hat{\sigma}^2$ is automatically estimated from image data. In Fig. 10(b), some anisotropic effects are visible, slightly enhancing coherence of lines in the image, and the noise component corresponding to fine texture is removed [Fig. 10(c)].

V. CONCLUSION

We have described a novel adaptive denoising algorithm where patch-based weights and variable window sizes are jointly used. An advantage of the method is that internal parameters can be easily chosen and are relatively stable. The algorithm is able to denoise both piecewise-smooth and textured natural images since they contain enough redundancy. Actually, the performance of our algorithm is very close, and in some cases even surpasses, to that of the already published denoising methods. Also we just mention that the algorithm can be easily parallelized since at iteration n , each pixel is processed independently. However, some problems may occur when the texture sample contains too many texels making hard to find close matches for the neighborhood context window.

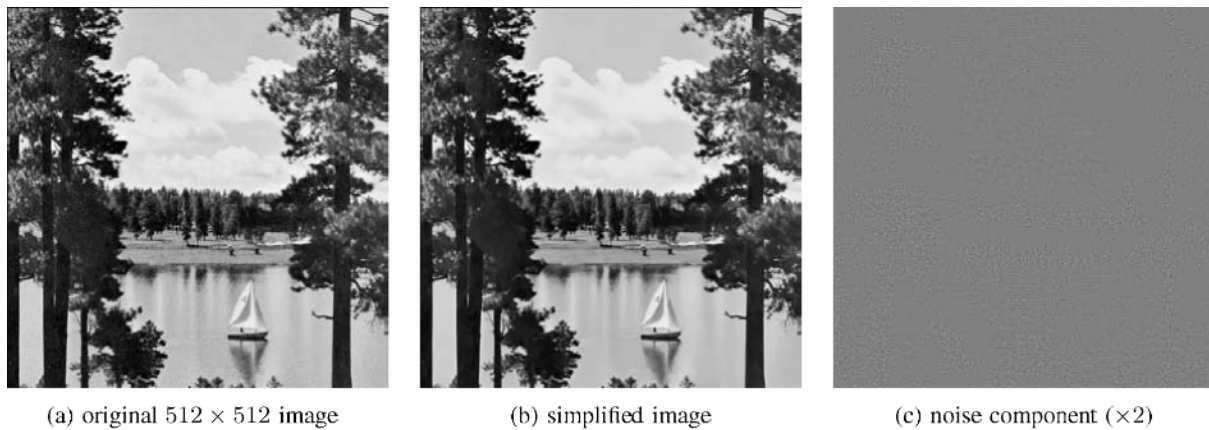


Fig. 10. Simplification of the *sailboat* 512×512 image ($\hat{\sigma} = 7.70$).

These problems can usually be eliminated by providing a larger image patch. In the future, we plan to study automatic patch-size selection with respect to the signal-to-noise ratio and the scale of textures to better adapt to image contents.

REFERENCES

- [1] J. F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, "Image decomposition into a bounded variation component and an oscillating component," *J. Math. Imag. Vis.*, vol. 22, no. 1, 2005.
- [2] D. Barash, "A fundamental relationship between bilateral filtering, adaptive smoothing and the nonlinear diffusion equation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 844–847, Jun. 2002.
- [3] D. Barash and D. Comaniciu, "A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift," *Image Vis. Comput.*, vol. 22, no. 1, pp. 73–81, 2004.
- [4] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 421–432, Mar. 1998.
- [5] M. J. Black and G. Sapiro, "Edges as outliers: Anisotropic smoothing using local image statistics," presented at the Scale-Space Conf. Kerkyra, Greece, 1999.
- [6] T. Brox and J. Weickert, "A TV flow based local scale measure for texture discrimination," in *Proc. Eur. Conf. Computer Vision*, Prague, Czech Republic, 2004, vol. 2, pp. 578–590.
- [7] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2004.
- [8] —, "Image denoising by non-local averaging," presented at the Computer Vision Pattern Recognition Conf. San Diego, CA, 2005.
- [9] F. Catte, P.-L. Lions, J.-M. Morel, and T. Coll, "Image selective smoothing and edge-detection by nonlinear diffusion," *SIAM J. Numer. Anal.*, vol. 29, no. 1, pp. 182–193, 1992.
- [10] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1–2, pp. 89–97, 2004.
- [11] C. Chef d'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras, "Regularizing flows for constrained matrix-valued images," *J. Math. Imag. Vis.*, vol. 20, no. 1–2, pp. 147–162, 2004.
- [12] Y. Cheng, "Mean-shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
- [13] D. Comaniciu, V. Ramesh, and P. Meer, "The variable bandwidth mean-shift and data-driven scale selection," presented at the Int Conf. Computer Vision Vancouver, BC, Canada, 2001.
- [14] D. Comaniciu and P. Meer, "Mean-shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [15] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.
- [16] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
- [17] D. L. Donoho and I. M. Johnston, "Denoising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [18] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," presented at the Int. Conf. Computer Vision Kerkyra, Greece, 1999.
- [19] M. Elad, "On the bilateral filter and ways to improve it," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1141–1151, Oct. 2002.
- [20] A. Fitzgibbon, Y. Wexler, and A. Zisserman, "Image-based rendering using image-based priors," presented at the Int. Conf. Computer Vision Nice, France, 2003.
- [21] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, 2000.
- [22] M. Ghazel, G. H. Freeman, and E. R. Vrscay, "Fractal image denoising," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1560–1578, Dec. 2003.
- [23] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Texture preserving variational denoising using an adaptive fidelity term," presented at the VLSCM Conf. Nice, France, 2003.
- [24] —, "Estimation of optimal PDE-based denoising in the SNR sense," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2269–2280, Aug. 2006.
- [25] F. Godtliebsen, E. Sjøtved, and J. S. Marron, "A nonlinear Gaussian filter applied to images with discontinuities," *J. Nonparametric Statist.*, vol. 8, pp. 21–43, 1997.
- [26] A. Goldenshluger and A. Nemirovsky, "On spatial adaptive estimation of nonparametric regression," *Math. Meth. Statist.*, vol. 6, no. 2, pp. 135–170, 1997.
- [27] G. Gomez, J. L. Marroquin, and L. E. Sucar, "Probabilistic estimation of local scale," presented at the Proc. Int. Conf. Pattern Recognition Barcelona, Spain, 2000.
- [28] A. Juditsky, "Wavelet estimators: Adapting to unknown smoothness," *Math. Meth. Statist.*, vol. 1, pp. 11–20, 1997.
- [29] V. Katkovnik, K. Egiazarian, and J. Astola, "Adaptive window size image denoising based on intersection of confidence intervals (ICI) rule," *J. Math. Imag. Vis.*, vol. 16, no. 3, pp. 223–235, 2002.
- [30] C. Kervann, "An adaptive window approach for image smoothing and structures preserving," presented at the Proc. Eur. Conf. Computer Vision Prague, Czech Republic, 2004.
- [31] C. Kervann and J. Boulanger, "Local Adaptivity to Variable Smoothness for Exemplar-Based Image Denoising And Representation INRIA Res. Rep., RR-5624, 2005.
- [32] J. S. Lee, "Digital image smoothing and the sigma filter," *Comput. Vis. Graph. Image Process.*, vol. 24, pp. 255–269, 1983.
- [33] E. Le Pennec and S. Mallat, "Sparse geometric image representation with bandelets," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 423–438, Apr. 2005.
- [34] O. Lepskii, "On a problem of adaptive estimation on white Gaussian noise," *Probl. Appl.*, vol. 35, pp. 454–466, 1990.
- [35] O. V. Lepskii, E. Mammen, and V. G. Spokoiny, "Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors," *Ann. Statist.*, vol. 25, no. 3, pp. 929–947, 1997.
- [36] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 117–154, 1998.
- [37] M. Maurizot, P. Bouthemy, B. Delyon, A. Juditski, and J.-M. Odobez, "Determination of singular points in 2D deformable flow fields," presented at the IEEE Int. Conf. Image Processing Washington, DC, 1995.

- [38] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, ser. University Lecture vol. 22. Providence, RI: AMS, 2001, vol. 22.
- [39] P. Mrazek and M. Navara, "Selection of optimal stopping time for nonlinear diffusion filtering," *Int. J. Comput. Vis.*, vol. 52, no. 2/3, pp. 189–203, 2003.
- [40] P. Mrazek, J. Weickert, and A. Bruhn, *On Robust Estimation and Smoothing With Spatial and Tonal Kernels*. Bremen, Germany: Univ. Bremen, 2004.
- [41] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and variational problems," *Commun. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, 1989.
- [42] S. Osher, A. Solé, and L. Vese, "Image decomposition and restoration using total variation minimization and the H^{-1} norm," *Multiscale Model. Simul.*, vol. 1, no. 3, pp. 349–370, 2003.
- [43] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–239, Jul. 1990.
- [44] A. Pizurica and W. Philips, "Estimating probability of presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Process.*, vol. 3, no. 15, pp. 654–665, Mar. 2006.
- [45] J. Polzehl and V. Spokoiny, "Adaptive weights smoothing with application to image restoration," *J. Roy. Statist. Soc. B.*, vol. 62, no. 2, pp. 335–354, 2000.
- [46] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [47] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors with applications," presented at the Computer Vision Pattern Recognition Conf. San Diego, CA, 2005.
- [48] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.
- [49] P. Saint-Marc, J. S. Chen, and G. Médioni, "Adaptive smoothing: A general tool for early vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 514–529, Jun. 1991.
- [50] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vis.*, vol. 37, no. 2, pp. 151–172, 2000.
- [51] S. M. Smith and M. Brady, "SUSAN—A new approach to low-level image processing," *Int. J. Comput. Vis.*, vol. 23, no. 1, pp. 45–78, 1997.
- [52] N. Sochen, R. Kimmel, and A. M. Bruckstein, "Diffusions and confusions in signal and image processing," *J. Math. Imag. Vis.*, vol. 14, no. 3, pp. 237–244, 2001.
- [53] L. Stankovic, "Performance analysis of the adaptive algorithm for bias-to-variance trade-off," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1228–1234, May 2004.
- [54] J. L. Starck, E. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.
- [55] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," presented at the Int. Conf. Computer Vis. Bombay, India, 1998.
- [56] J. van de Weijer and R. van den Boomgaard, "Local mode filtering," presented at the Computer Vision Pattern Recognition Kauai, HI, 2001.
- [57] T. Weismann, E. Ordentlich, G. Seroussi, S. Verdu, and M. Weinberger, "Universal discrete denoising: Known channel," *IEEE Trans. Inf. Theory.*, to be published.
- [58] D. Zhang and Z. Wang, "Image information restoration based on long-range correlation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 5, pp. 331–341, May 2002.
- [59] S. C. Zhu, Y. Wu, and D. Mumford, "Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling," *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 107–126, 1998.

Charles Kervrann was born 1969. He received the Ph.D. degree in signal processing and telecommunications from the university of Rennes 1, Rennes, France, in 1995.

In 1998, he was appointed to a full-time Researcher at the Applied Mathematics and Informatics Department, National Institute for Agricultural Research (INRA), Rennes. From 1995 to 1997, he was an Expert Engineer with the Temics project, INRIA. In 2003, he was on secondment at INRIA in the Vista Project. His interests include statistical models for image processing, tracking in image sequences, biological imaging, and video microscopy.

Jérôme Boulanger was born in 1980. He graduated with a three-year engineering degree from the National School of Physics of Marseille (ENSPM), Marseille, France, in 2003. During the final year of this degree, he received the DEA specialized in image processing. He is currently pursuing the Ph.D. degree at IRISA, Rennes, France.

His main research interests are in statistical models for image sequence analysis in biological imaging.