Local adaptivity to variable smoothness for exemplar-based image denoising and representation

Charles Kervrann, Jérôme Boulanger

To cite this version:

Charles Kervrann, Jérôme Boulanger. Local adaptivity to variable smoothness for exemplar-based image denoising and representation. [Research Report] PI 1733, 2005, pp.60. inria-00000166

HAL Id: inria-00000166
https://inria.hal.science/inria-00000166

Submitted on 18 Jul 2005

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES

PUBLICATION
INTERNE
N^{o} 1733

LOCAL ADAPTIVITY TO VARIABLE SMOOTHNESS FOR EXEMPLAR-BASED IMAGE DENOISING AND REPRESENTATION

CHARLES KERVRANN AND JÉRÔME BOULANGER

ISSN 1166-8687

IRISA
CAMPUS UNIVERSITAIRE DE BEAULIEU - 35042 RENNES CEDEX - FRANCE

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES
Campus de Beaulieu – 35042 Rennes Cedex – France
Tél. : (33) 02 99 84 71 00 – Fax : (33) 02 99 84 71 71
http://www.irisa.fr

Local adaptivity to variable smoothness for exemplar-based image denoising and representation

Charles Kervrann * and Jérôme Boulanger **

Systèmes cognitifs
Projet Vista

Publication interne n ° 1733 — Juillet 2005 — 60 pages

Abstract: A novel adaptive and exemplar-based approach is proposed for image restoration and representation. The method is based on a pointwise selection of small image patches of fixed size in the variable *neighborhood* of each pixel. The main idea is to associate with each pixel the weighted sum of data points within an adaptive neighborhood. This method is general and can be applied under the assumption that the image is a locally and fairly stationary process. In this paper, we focus on the problem of the adaptive neighborhood selection in a manner that it balances the accuracy of approximation and the stochastic error, at each spatial position. Thus, the new proposed pointwise estimator automatically adapts to the degree of underlying smoothness which is unknown with minimal *a priori* assumptions on the function to be recovered. Finally, we propose a practical and simple algorithm with no hidden parameter for image denoising. The method is applied to both artificially corrupted and real images and the performance is very close, and in some cases even surpasses, to that of the already published denoising methods. Also, the method is demonstrated to be valuable for applications in fluorescence microscopy.

Key-words: exemplar-based methods, estimation, bias-variance trade-off, restoration, denoising, nonlinear filtering, detection, fluorescence microscopy

(Résumé : tsvp)

authors' affiliation: IRISA-INRIA / INRA - Département de Mathématiques et Informatique Appliquées (MIA) (Jouy-en-Josas - 78352)

[*] charles.kervrann@irisa.fr
[**] jerome.boulanger@irisa.fr

Centre National de la Recherche Scientifique
(UMR 6074) Université de Rennes 1 – Insa de Rennes

Institut National de Recherche en Informatique
et en Automatique – unité de recherche de Rennes

Adaptation locale pour la représentation et le débruitage d'image à base de motifs locaux

Résumé : Nous proposons une nouvelle méthode adaptative pour la restauration et la représentation d'image. L'idée est de sélectionner dans un voisinage adapté pour chaque pixel, des motifs qui sont des copies légèrement modifiées du motif centré au pixel considéré. La méthode de restauration, apparentée aux méthodes à noyaux pour la régression non-paramétrique, cherche alors à calculer, en chaque point, une moyenne pondérée des observations sélectionnées dans un voisinage variable spatialement. L'optimisation de la taille du voisinage repose ici sur un compromis biais/variance de l'estimateur. L'algorithme final, dirigé par les données, est d'une grande simplicité et nécessite à peine l'ajustement d'un faible nombre de paramètres. Nous présentons une comparaison avec des algorithmes conventionnels et des résultats expérimentaux qui mettent en évidence le potentiel de cette méthode pour traiter des situations o l'image est un processus localement stationnaire. Cette méthode est très efficace, puisque les performances obtenues dépassent la plupart des méthodes existantes. Elle a également été validée sur des images de microscopie de fluorescence en bio-imagerie.

Mots clés : méthode basée exemple, estimation, compromis bias-variance, restauration, débruitage, filtrage non-linéaire, détection, microscopie à fluorescence

Contents

1 Introduction

Traditionally, the problem of image recovering is to reduce undesirable distortions and noise while preserving important features such as homogeneous regions, discontinuities, edges and textures. In image regions corresponding to the interior of an object, linear filtering produces desirable results, that is a pixel value which is more representative of the region in which it lies. However, in the neighborhood of discontinuities, linear filtering such as the Gaussian filter, removes noise but blurs edges significantly. This undesirable effect can be reduced by taking into account local geometries and statistics during the filtering process. Popular image restoration algorithms are therefore nonlinear to reduce the amount of smoothing near abrupt changes:

- Most of the more efficient regularization methods are based on discrete [6] or continuous [56, 64] 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 to data. Thus, related partial differential equations (PDE) and variational methods, including anisotropic diffusion [59, 77, 8] and total variation (TV) minimization [64], have shown impressive results to tackle the problem of edge-preserving smoothing [59, 14, 77, 15].

- For reasons of efficiency in computer vision, other smoothing algorithms aggregate information over a neighborhood of fixed size, based on two basic criteria: a spatial criterion and a brightness criterion. The so-called *bilateral filtering* [75, 3] and previous local nonlinear filters (e.g. Lee 's filter [45], *Susan* filter [69]) use this generic principle and can be then considered as classical *neighborhood filters*, involving the local weighted averaging of input data over a spatial neighborhood. The weight functions are typically Gaussian kernels and are globally controlled by setting the standard deviations in both spatial and brightness domains. The relationships between nonlinear Gaussian filters, and iterative *mean-shift* algorithm, local mode filtering, clustering, local M-estimators, nonlinear diffusion, regularization approaches combining nonlocal data and nonlocal smoothness terms, have been recently investigated in [79, 26, 4, 55, 72].

As effective as these smoothing techniques, we note they have a relatively small number of parameters that control the global amount of smoothing being performed. The problem is that there is no satisfying way to retrieve these smoothing parameters from data in practical imaging and they are usually chosen to give a good visual impression [75, 3]. Furthermore, when local characteristics of the data differ significantly across the domain, tuning these global smoothing parameters is probably not satisfying. Some efforts have been then reported to determine local scales of significant image features and to detect non-stationarities for image regularization [51, 8, 18, 33, 44]. Another competitive approach consists in decomposing the image into its primary noise, texture and bounded variation (BV) components [53, 58, 1], which actually can be hard to compute in practice.

What makes image restoration a difficult task, 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 non-parametric models to capture all the regularities and geometries seen in local patterns. In contrast to the above-cited methods, another line of work consists then in modeling non-local pairwise interactions from training data [82] or a library of natural image patches [29, 41, 63]. 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 [82, 46], image denoising [63], image reconstruction and super-resolution [29] and image rendering [28]). Also, it has been experimentally confirmed that more intuitive exemplar-based approaches are fearsome for 2D texture synthesis [25] and image inpainting [20]. In our framework, we will also assume that small image patches in the variable neighborhood of a pixel contains the essential process required for local restoration. Unlike most existing exemplar-based MRF methods that use training sets and optimization algorithms for learning [29, 63], the proposed restoration approach is unsupervised and 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 in [12, 13]. The idea is to estimate the underlying image at a point from similar copies of a central pattern detected in a local neighborhood. We 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. In addition, we address the central problem of choosing the smoothing window (neighborhood) which can different at each pixel to cope with spatial inhomogeneities across the image domain. Accordingly, we propose to use a kind of change-point detection procedure, initiated by Lepskii for 1D signals [48]. The Lespkii's principle, also based on "wavelets" ideas [23, 24, 42], is a procedure which aims at minimizing the pointwise risk of the estimator and amounts to balance the accuracy of approximation and the stochastic error at each spatial position. This so-called *pointwise adaptive estimation* approach has been described in its general form and in great details in [48, 49], and the interested readers should of course have a look at these milestone papers [48, 49]. In short, our approach can be viewed as an application of this idea of *pointwise adaptive estimation* [48, 50] combined with exemplar-based techniques for image restoration. We just point out that the proposed approach requires no training procedure and no adaptive partitioning scheme such as quad-trees as already investigated in fractal denoising [32]. However, it shares some common points with the recent *non-local means* algorithm [12, 13] and other exemplar-based methods [25, 20]. Other related works to our approach are nonlinear Gaussian filters [35, 75, 3, 79, 55] and statistical regularization schemes [61, 43], but are enhanced via incorporating either a variable window scheme or exemplar-based weights.

The remainder of the paper is organized as follows. Related studies are presented in Section 2. In Section 3, we introduce the image modeling and some notations. In Section 4, we formulate the problem of the selection of the best possible window and present the adaptive estimation procedure, whose complexity is actually controlled by simply restricting the size of the larger window. Theoretical results are given in Section 5. In Section 6, we propose a practical algorithm with no hidden parameter for image denoising. In Section 7, we demonstrate the ability of the method to restore artificially corrupted images with additive white Gaussian noise (WGN). Also, the method is applied to denoise real fluorescence microscopy images in bio-imaging. Conclusions and perspectives are presented in Section 8.

2 Related work for image denoising

In this section, the relationships between the proposed technique and other image restoration methods are discussed. A large number of methods have been proposed for image denoising and regularization as recently described in [12]. To our knowledge, the more competitive methods are recent wavelet-based methods. However, we do not draw into comparison both the wavelet-based methods [24, 74, 62, 22, 60, 47] since they process the input data in the transformed domain, and the fractal denoising methods [32] also rooted in wavelet theory. In Section 5, we have just reported the experimental results when these wavelet-based methods are applied to a commonly-used image dataset in image denoising [62]. In the current section, we mainly focus on nonlinear Gaussian filters, nonlinear diffusion, M-estimators from robust statistics and some related energy minimization methods since they share some common points with the proposed method (see also [12] for a recent survey).

Fixed-window methods yield good results when all the pixels in the window come from the same population as the central pixel. However, difficulties arise when the square (or circular) window overlaps a discontinuity. Filtering with a window (or spatial kernel) that is symmetric around the central pixel results in averaging of edge values, and therefore blurring of the edge. In such cases, a possible strategy would be to substitute the border pixel with a pixel inside the object. Nitzberg & Shiota [57] (and later Fischl & Schwartz [27]) proposed an offset term that displaces spatial kernel centers away from the presumed edge location, thus enhancing the contrast between adjacent regions without blurring their boundaries. The so-called *"offset-filtering"* requires the generation of a vector field over the image domain which specifies an appropriate displacement at each point. Intuitively, the displacement direction is calculated from the dominant local gradient direction. Unfortunately, approaches based on gradient information are known to be quite sensitive to noise. The dominant directions can be more robustly estimated from eigenvectors of the second moment matrix, which are essentially first and second order derivatives. This reduces the sensitivity to noise but requires the user-defined setting of a global scale of features to be preserved during diffusion [78] A more commonly-used strategy consists then in disregarding some data points in the fixed window that should be classified as outliers in a bimodal distribution. In our framework, we will exploit a such concept originated from robust statistics to filter out some undesirable data points in the estimation window [7, 8], and to improve the estimation of image discontinuities [7, 8].

An other class of nonlinear filters aims at estimating a connected component of pixels, that can be of arbitrary shape, containing the point of interest. Such nonlinear filters specialized in Gaussian noise elimination have re-appeared in recent papers [75, 9, 3]. Most of them use a certain window of neighboring pixels to modify the observed value at a given pixel. One primarily and typical filter is the sigma filter [45] described as follows: define $E_i = \{\mathbf{x}_j \in \Delta_i : |Y_i - Y_j|^2 \leq 2g^2\}$, where Δ_i is the window centered at pixel $\mathbf{x}_i$, Y_i is the observation at $\mathbf{x}_i$ and g is a smoothing parameter to be determined and depending on the signal-to-noise ratio (SNR); then the output of the filter is

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

where $L_g(Y_i - Y_j) = 1$ if $\mathbf{x}_j \in E_i$ and 0 otherwise. It can be seen that the weight function of the sigma filter is discontinuous. A continuous version of this weight might be more plausible and choosing $L_g(Y_i - Y_j) = \exp(-(Y_i - Y_j)^2/2g^2)$ gives the well-known nonlinear Gaussian filter [35]. 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* (e.g. Lee 's [45], Susan [69] and Saint-Marc 's [65] filters) of the form

$$\widehat{u}(\mathbf{x}_i) \;=\; \frac{\sum_{\mathbf{x}_j \in \Delta_i} L_g(Y_i - Y_j)\; K_h(\mathbf{x}_i - \mathbf{x}_j)\; Y_j}{\sum_{\mathbf{x}_j \in \Delta_i} \; L_g(Y_i - Y_j) K_h(\mathbf{x}_i - \mathbf{x}_j)}. \tag{2}$$

where $K_h(\cdot) = (1/h)K(\cdot/h)$ and $L_g(\cdot) = (1/g)L(\cdot/g)$ are rescaled versions of non-negative kernel functions K and L. By definition, the continuous kernel functions K and L fulfill: $\int K(t)dt = 1$, $\int L(t)dt = 1$ and have ℓ vanishing moments: $\int t^l K(t)dt = 0$, $\int t^l L(t)dt = 0$ for $1 \leq l \leq \ell$. The bandwidths h and g control the amount of averaging. These "weight" functions depends on both the spatial and intensity difference between the central pixel and its neighbors. The weight functions $L_g(\cdot)$ and $K_h(\cdot)$ are typically Gaussians kernels where and 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. However, both must be set manually according to the image contents. Note, in particular, that as $g \to \infty$, the bilateral filter approaches a Gaussian filter of standard deviation h, and as both $g, h \to \infty$, the bilateral filter approaches the mean filter. In short, smoothing over some neighborhood is based on a spatial criterion and a brightness criterion in order to select similar data points. Thus, this generic principle, also developed by Yaroslavsky [81], was adopted in [9] where the authors describe an efficient method for choosing an arbitrarily shaped connected window, in an manner that varies at each pixel. More recently, Buades *et al.* proposed the so-called *non-local means* filter [12, 13] defined as

$$\widehat{u}(\mathbf{x}_i) \;=\; \frac{\sum_{\mathbf{x}_j \in \Delta_i} L_g(\mathbf{Y}_i - \mathbf{Y}_j)\; K_h(\mathbf{x}_i - \mathbf{x}_j)\; Y_j}{\sum_{\mathbf{x}_j \in \Delta_i} \; L_g(\mathbf{Y}_i - \mathbf{Y}_j) K_h(\mathbf{x}_i - \mathbf{x}_j)} \tag{3}$$

where $\mathbf{Y}_i$ denotes a vector of pixel values taken in the neighborhood of a point $\mathbf{x}_i$. 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*. It is worth noting that, if the size of the patch is reduced to one pixel, the *non-local means* filter, also controlled by a small number of smoothing parameters g and h, is strictly equivalent to (2). Finally, it has been shown that using example image patches greatly improves image denoising [12, 13], and will be also considered in our framework.

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 & Comaniciu [4] have showed that bilateral filtering represents a weighted averaging algorithm which turns out to be an implementation of anisotropic diffusion [59, 77], controlled by a global scale parameter. Elad [26] established further how the bilateral filter is algorithmically related to anisotropic diffusion [59] and robust estimation [7] 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 [72, 73]. Furthermore, Barash & Comaniciu showed that kernel density estimation applied into the joint spatial-range domain yields a powerful processing paradigm - the *mean-shift* procedure [19] - also related to bilateral filtering but having additional flexibility [4]. The link between iterative *mean-shift* algorithm, local mode filtering, clustering, local M-estimators, nonlinear diffusion, regularization approaches were already analyzed in [16, 79, 76, 26, 55, 4]. Also, all these methods have been casted into a unified framework for functional minimization combining nonlocal data and nonlocal smoothness terms in [55]. In particular, Mrazek *et al.* brought to the fore the large amount of structural similarities between the iterated bilateral filter

$$\widehat{u}_{n+1}(\mathbf{x}_i) \;=\; \frac{\sum_{\mathbf{x}_j\in\Delta_i} L_g(\widehat{u}_n(\mathbf{x}_i)-\widehat{u}_n(\mathbf{x}_j))\; K_h(\mathbf{x}_i-\mathbf{x}_j)\;\widehat{u}_n(\mathbf{x}_j)}{\sum_{\mathbf{x}_j\in\Delta_i} L_g(\widehat{u}_n(\mathbf{x}_i)-\widehat{u}_n(\mathbf{x}_j))\; K_h(\mathbf{x}_i-\mathbf{x}_j)} \tag{4}$$

and the local M-smoother

$$\widehat{u}_{n+1}(\mathbf{x}_i) \;=\; \frac{\sum_{\mathbf{x}_j\in\Delta_i} w_g(Y_j-\widehat{u}_n(\mathbf{x}_i))\; K_h(\mathbf{x}_i-\mathbf{x}_j)\; Y_j}{\sum_{\mathbf{x}_j\in\Delta_i} w_g(Y_j-\widehat{u}_n(\mathbf{x}_i))\; K_h(\mathbf{x}_i-\mathbf{x}_j)} \tag{5}$$

where $w_g(s^2) = \rho'_g(s^2)$ and $\rho(s^2)$ 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. Note that the use of a chain of nonlinear filters has been also proposed in [2], and the bandwidths g and h vary at each iteration according to specific and deterministic rules to balance edge detection and noise removal. To complete the state of the art of related regularization methods to our approach, we must mention the successful of the Total Variation (TV) minimizing process of Rudin-Osher-Fatemi [64] which simultaneously computes a piecewise smooth solution and estimate discontinuities. This method aims at minimizing an energy functional comprised of the TV norm and the fidelity of this image to the noisy input image. In contrast to neighborhood filters, the regularization is extremely local and involves only the pixel values and derivatives at point $\mathbf{x}_i$. In the discrete setting, this energy functional, also casted into the Bayesian framework [38], is related to MRF methods and can be stated as follows

$$\sum_{i=1}^{|G|} |\nabla u(\mathbf{x}_i)| \; + \; \frac{\lambda_{\scriptscriptstyle TV}}{2} \sum_{i=1}^{|G|} (Y_i - u(\mathbf{x}_i))^2 \tag{6}$$

where $|G|$ is the number of pixels of the discrete image domain G, $\nabla u(\mathbf{x}_i)$ is a discrete approximation of the gradient of the function u and $\lambda_{\scriptscriptstyle TV}$ is interpreted as Lagrangian multiplier, which globally balances the contribution of the TV norm and the fidelity term. The steepest descent of this TV norm is then related to the mean curvature motion and, further, to the robust anisotropic diffusion if a robust M-estimator norm is substituted to the L_1 norm (see [7, 5]). In (6), the fidelity term is actually used to automatically stop the diffusion process and to avoid a flat image. In this paper, we will compare our method with the Rudin-Osher-Fatemi method since it performs well in image denoising.

Finally, all cited restoration methods have a relatively small number of smoothing parameters (e.g. bandwidths g and h in (2), $\lambda_{\scriptscriptstyle TV}$ in (6)) that control the global amount of smoothing being performed. In implementing these methods, the first question to be asked is how should the smoothing parameters be chosen ? The major drawback of global methods is that there is no satisfying way to retrieve the smoothing parameters from data. A number of authors have then turned to various tools derived with statistical motivations such as bandwidth selection could be interpreted and exploited for global parameter selection in the diffusion process [16, 79, 76, 26, 55, 4, 33]. However, it is also theoretically confirmed that automatically determining a bandwidth for spherical kernel is a difficult problem [35], and the bandwidths g and h involved in the bilateral filtering are usually chosen to give a good visual impression and are heuristically chosen [75, 3]. Furthermore, when local characteristics of the data differ significantly across the domain, selecting either optimal bandwidths or optimal Lagrangian multipliers 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, some mechanisms using a spatially varying parameter $\lambda_{\scriptscriptstyle TV}(\mathbf{x}_i)$ and local variances have been proposed to adapt restoration over image regions according to their contents [33]. In [37], the local amount of Gaussian smoothing is computed in terms of variance in a space-scale framework, through the minimal description length criterion (MDL). The local variance is actually useful for localization of significant image features [8, 33, 37]. An alternative way to select the local scale is to maximize a measure of edge strength over scale space [51] but the resulting scale computed from image derivatives, is sensitive to signal-to-noise ratio. More recently, TV flow has been suggested since it includes a non-explicit scale parameter useful for detecting the scale of image features [10]. Also, some efforts have been reported to locally vary the bandwidths in kernel regression [18, 68]. In [68], the authors determine local bandwidths using Parzen windows to mimic local density for image segmentation. This is a variant of the *plug-in* idea usually used in the statistics literature, which is fast and easy to compute. However the plug-in approach is problematic since it is known to be highly sensitive to noise in images and to the choice of a global initial bandwidth.

In this paper, we will focus on this problem which is an open research issue, and propose a stable scheme to select the best possible size and shape for local regression. In addition, we propose to use example image patches to take into account complex spatial interactions in images. In contrast to exemplar-based approaches for image modeling [82, 63], we propose an unsupervised framework that uses no library of image patches and no computational intensive training algorithms [29, 63]. Our adaptive smoothing works in the joint spatial-range domain as the *non-local means* filter [13] 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 presented in the remainder of the paper.

3 Image model and basic idea

In order to describe our estimation method, let us first introduce some useful notations. Consider the following image model

$$Y_i = u(\mathbf{x}_i) + \epsilon_i, \quad i = 1, \ldots, |G| \tag{7}$$

where $\mathbf{x}_i \in \mathbb{R}^d$, $d \geq 2$, represents the spatial coordinates of the discrete image domain G 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 variance, i.e. $\mathrm{Var}(\epsilon_i) \overset{def}{=} \sigma^2$. Our aim is to recover $u : \mathbb{R}^d \rightarrow \mathbb{R}$ from noisy observations Y_i. In what follows, we suppose no explicit smoothness assumptions on u is available and it will be only assumed that u is compactly supported and bounded

Nevertheless, 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 will assume that the unknown image $u(\mathbf{x}_i)$ can be approximated by 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 estimated patch $\mathbf{u}_j$ to the reference image patch $\mathbf{u}_i$ will have larger weights in the average. This amounts to suppose there exists some stationarity in the 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 [82, 29, 41, 63]. We only focus on image patches as non-local image features, and adapt kernel regression techniques for image restoration.

Here, 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 $k = p^2$ elements where the pixels are concatenated along a fixed lexicographic ordering. As with all exemplar-based techniques, the size of image patches must be specified in advance according to how stochastic the user believes the image to be [25, 20]. In our experiments, we shall see that 7×7 or 9×9 patches are 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 computer vision 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. Hence, the choice of the set $\mathcal{N}_\Delta$ of candidate neighborhoods will play the key role. For the sake of parsimony, the set of admissible neighborhoods will be arbitrarily chosen as a geometric grid of nested square windows

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

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 $\mathcal{N}_\Delta$. For technical reasons, we will require the following conditions: $\Delta_{i,n}$ is centered at $\mathbf{x}_i$ and $\Delta_{i,n} \subset \Delta_{i,n+1}$ and assume that the set $\mathcal{N}_\Delta$ is finite. In the next section, we will precise this adaptive estimation procedure and 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 estimator as possible. This optimization will be mainly accomplished by starting, at each pixel, with a small window $\Delta_{i,0}$ as a *pilot* estimate, and growing $\Delta_{i,n}$ with n.

4 Adaptive estimation procedure

The proposed procedure is iterative and works as follows [61, 44].

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 $\widehat{u}_{i,0}$ (and its variance $\widehat{v}^2_{i,0} \overset{def}{=} \mathrm{Var}(\widehat{u}_{i,0})$) is then

$$\widehat{u}_{i,0} \;=\; Y_i \qquad \text{and} \qquad \widehat{v}^2_{i,0} \;=\; \widehat{\sigma}^2 \tag{8}$$

where an estimated variance $\widehat{\sigma}^2$ has been plugged in place of σ^2 since the variance of errors is supposed to be unknown (see Section 6). 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 $\mathbf{x}_j \in \Delta_{i,1}$ gets a weight $\pi_{i\sim j,1}$ defined by comparing pairs of current estimated patches $\widehat{\mathbf{u}}_{i,0} = (\widehat{u}^{(1)}_{i,0}, \cdots, \widehat{u}^{(k)}_{i,0})^T$ and $\widehat{\mathbf{u}}_{j,0} = (\widehat{u}^{(1)}_{j,0}, \cdots, \widehat{u}^{(k)}_{j,0})^T$ of size $k = p \times p$, obtained at the first iteration. Here, the subscript $i \sim j$ means "$\mathbf{x}_j \in \Delta_{i,\cdot}$ *and the index j runs through the neighborhood of* $\mathbf{x}_i$*"*. As usual, the points $\mathbf{x}_j$ with a similar patch $\widehat{\mathbf{u}}_{j,0}$ to $\widehat{\mathbf{u}}_{i,0}$ will have weights close to 1 and 0 otherwise. Then we recalculate the estimate $\widehat{u}_{i,1}$ 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 $\widehat{u}_{i,n} \overset{def}{=} \widehat{u}_{\Delta_{i,n}}(\mathbf{x}_i)$ and its variance $\widehat{v}^2_{i,n} \overset{def}{=} \mathrm{Var}(\widehat{u}_{i,n})$ can be represented as

$$\widehat{u}_{i,n} \;=\; \sum_{\mathbf{x}_j\in\Delta_{i,n}} \pi_{i\sim j,n} Y_j, \quad \text{and} \quad \widehat{v}^2_{i,n} \;=\; \widehat{\sigma}^2 \sum_{\mathbf{x}_j\in\Delta_{i,n}} [\pi_{i\sim j,n}]^2 \tag{9}$$

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 $\sum_{\mathbf{x}_j\in\Delta_{i,n}} \pi_{i\sim j,n} = 1$. Here, we suggest to compute the weights from pairs of current restored patches $\widehat{\mathbf{u}}_{i,n-1}$ and $\widehat{\mathbf{u}}_{j,n-1}$ obtained at iteration $n-1$ to capture complex interactions between pixels (see Fig. 1). In what follows, n will coincide with the iteration and we will use $\widehat{n}(\mathbf{x}_i)$ to designate the index of the "best" window chosen among all non-rejected window $\Delta_{i,n}$ from $\mathcal{N}_\Delta$, as

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

where ϱ is a positive constant. 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.

With this adaptive choice of window which depends on the observations instead of a usual deterministic window and defined exemplar-based weights, this estimator is clearly not linear and uses no explicit assumption on the smoothness of the unknown function u. Moreover, the best possible estimator $\widehat{u}(\mathbf{x}_i)$ is computed from the whole path of values $\{\widehat{u}_{i,n}\}$ as

$$\widehat{u}(\mathbf{x}_i) \stackrel{def}{=} \widehat{u}_{i,\widehat{n}(\mathbf{x}_i)} = \sum_{\mathbf{x}_j\in\widehat{\Delta}(\mathbf{x}_i)} \pi_{i\sim j,\widehat{n}(\mathbf{x}_i)}\, Y_j. \tag{10}$$

The use of variable and overlapping windows contributes to the restoration performance with no block effect and make them possible to cope well with spatial inhomogeneities in natural images. In the following, we use a spatial rectangular kernel (i.e. square windows) for mathematical convenience and to refer to the sigma filter [45], but the method can be naturally extended to the case of a more usual spatial Gaussian kernel [35, 75, 17, 3, 79]. In equation (9), it worth noting the weight function $\pi_{i\sim j,n}$ does not directly depend on input data but are only calculated from neighboring local estimates. This contrasts with traditional local M-estimators (see (5)), non-linear Gaussian filters (see (2)) and the *non-local means* filter (see (3)).

4.1 Adaptive weights

As already mentioned, we may decide on the basis of current estimates $\widehat{\mathbf{u}}_{i,n-1}$ and $\widehat{\mathbf{u}}_{j,n-1}$, if a pixel $\mathbf{x}_j$ belongs to the region $\Delta_{i,n-1}$. In order to compute the similarity between patches $\widehat{\mathbf{u}}_{i,n}$ and $\widehat{\mathbf{u}}_{j,n}$, an objective distance must be considered. In [25, 20, 12, 13], several authors showed that the L_2 distance $\|\widehat{\mathbf{u}}_{i,n-1} - \widehat{\mathbf{u}}_{j,n-1}\|^2$ is a reliable measure to compare image patches. To make a decision, we have rather used the following normalized distance

$$\begin{aligned}\text{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1}) \;=\;& \frac{1}{2}\;\Big[(\widehat{\mathbf{u}}_{i,n-1} - \widehat{\mathbf{u}}_{j,n-1})^T\, \widehat{\mathbf{V}}_{i,n-1}^{-1}\, (\widehat{\mathbf{u}}_{i,n-1} - \widehat{\mathbf{u}}_{j,n-1}) \\ &+ \;(\widehat{\mathbf{u}}_{j,n-1} - \widehat{\mathbf{u}}_{i,n-1})^T\, \widehat{\mathbf{V}}_{j,n-1}^{-1}\, (\widehat{\mathbf{u}}_{j,n-1} - \widehat{\mathbf{u}}_{i,n-1})\Big]\end{aligned} \tag{11}$$

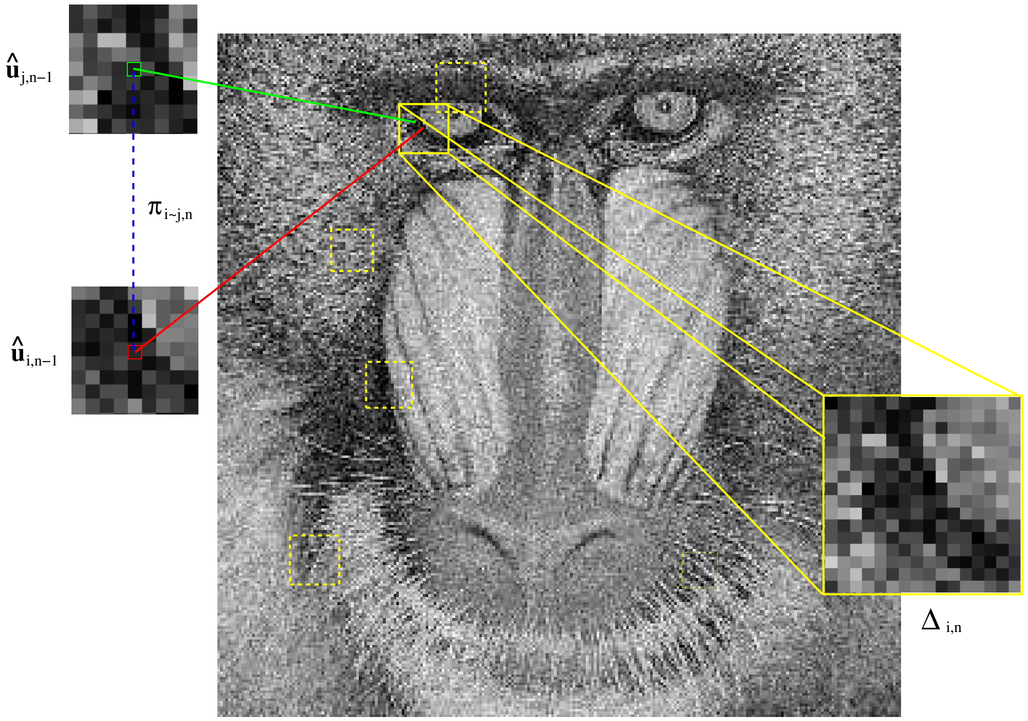


Figure 1: Description of the exemplar-based restoration approach.

where $\widehat{\mathbf{V}}_{i,n-1}$ and $\widehat{\mathbf{V}}_{j,n-1}$ are diagonal matrices with the k non-zero diagonal elements equal to variances $\{\widehat{v}^2_{i,n-1}\}$ and $\{\widehat{v}^2_{j,n-1}\}$ associated to the current estimated patch $\widehat{\mathbf{u}}_{i,n-1}$ and $\widehat{\mathbf{u}}_{j,n-1}$. Accordingly, The hypothesis $\widehat{\mathbf{u}}_{i,n-1}$ and $\widehat{\mathbf{u}}_{j,n-1}$ are similar is accepted if the distance is small, i.e. $\text{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1}) \leq \lambda_\alpha$. In our modeling, the parameter $\lambda_\alpha \in \mathbb{R}_+$ is chosen as a quantile of a $\chi^2_{k,1-\alpha}$ distribution with k degrees of freedom, and controls the probability of type I error for the hypothesis of two points to belong to the same region:

$$\mathbb{P}\left\{\text{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1}) \leq \lambda_\alpha\right\} = 1 - \alpha. \tag{12}$$

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. Henceforth, we introduce the following commonly-used weight function

$$\pi_{i\sim j,n} \;=\; \frac{K\left(\lambda_\alpha^{-1}\,\text{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1})\right)}{\displaystyle\sum_{\mathbf{x}_j \in \Delta_{i,n}} K\left(\lambda_\alpha^{-1}\,\text{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1})\right)} \tag{13}$$

with $K(\cdot)$ denoting a monotone decreasing function, e.g. a kernel $K(x) = \exp(-x/2)$. Due to the fast decay of the exponential kernel, large distances between estimated patches lead to nearly zero weights, λ_α acting as an automatic threshold. Note that the use of weights enables to relax the structural assumption the neighborhood is roughly modeled by a square window, and is an alternative strategy to a more natural geometric strategy which consists in estimating the anisotropy and direction of a rectangular or elliptic window (see [80] for instance). To complete the description, the optimal estimator (9) requires the determination of the best window adapted for each pixel. This difficult problem is theoretically addressed in the next section.

4.2 An "ideal" smoothing window

In this section, we address the problem of 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 usual *global risks* cannot wholly reflect the performance of estimators at a point. Then, a classical way to measure the performance of the estimator $\widehat{u}_{i,n}$ 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 $\widehat{b}^2_{i,n}$ and variance $\widehat{v}^2_{i,n}$:

$$\left[\mathbb{E}|\widehat{u}_{i,n} - u(\mathbf{x}_i)|^2\right]^{1/2} \;=\; |\widehat{b}^2_{i,n} + \widehat{v}^2_{i,n}|^{1/2}. \tag{14}$$

Our goal is to minimize this local L_2 risk with respect to the size of the window $\Delta_{i,n}$, and for each pixel in the image. It is worth noting that the weights depend on previous estimates and a closed-form solution for the ideal window cannot be calculated. 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 (see [67, 39, 43, 71]). 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 $\widehat{b}_{i,n} = \mathbb{E}\left[\widehat{u}_{i,n} - u(\mathbf{x}_i)\right]$ is nonrandom and characterizes the accuracy of approximation of the function u at the point $\mathbf{x}_i$ by the smoothing window. As it explicitly depends on the unknown function $u(\mathbf{x}_i)$, it is usually not very useful by itself. Nevertheless, approximations to the bias can be derived as follows: first, we assume there exists a real constant $C_1 < \infty$ for all $\mathbf{x}_i \in G$ such that $|u(\mathbf{x}_j) - u(\mathbf{x}_i)| \leq C_1|\mathbf{x}_j - \mathbf{x}_i|$ (i.e. u is Lipschitz continuous). Also, we use the geometric inequality

$|\mathbf{x}_j - \mathbf{x}_i| \leq \frac{\sqrt{2}}{2}|\Delta_{i,n}|^{1/2}$ for 2D images. Then, it follows that

$$\begin{aligned}|\widehat{b}_{i,n}| \; = \; \left| \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n} \; [\mathbb{E}[Y_j] - u(\mathbf{x}_i)] \right| \; & \leq \; \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n} \; |u(\mathbf{x}_j) - u(\mathbf{x}_i)| \\ & \leq \; C_1 \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n} \; |\mathbf{x}_j - \mathbf{x}_i| \\ & \leq \; \frac{C_1 |\Delta_{i,n}|^{3/2}}{\sqrt{2}}\end{aligned}$$

and so $|\widehat{b}_{i,n}|^2$ is of the order $O(|\Delta_{i,n}|^3)$. Thus the squared bias is small when $|\Delta_{i,n}|$ is small and typically increases when $\Delta_{i,n}$ grows. As expected, small windows give a less biased estimator.

- The behavior of the variance term is just opposite. The errors are independent and the stochastic term $\widehat{\upsilon}^2_{i,n}$ 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

  $$\frac{\widehat{\sigma}^2}{|\Delta_{i,n}|} \; \leq \; \widehat{\upsilon}^2_{i,n} \; \leq \; \widehat{\sigma}^2.$$

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

In conclusion, the bias and standard deviation 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

$$\mathbb{E}|\widehat{u}_{i,n} - u(\mathbf{x}_i)|^2 \; \leq \; \frac{C_1^2}{2}|\Delta_{i,n}|^3 + \frac{\widehat{\sigma}^2}{|\Delta_{i,n}|^\beta}.$$

This equation summarizes the well-known trade-off between bias and variance and the size of the optimal window can be easily calculated as

$$|\Delta^\star(\mathbf{x}_i)| \; = \; \left[\frac{2\beta\widehat{\sigma}^2}{3C_1^2}\right]^{\frac{1}{\beta+3}}.$$

Note that this result cannot be used in practice since C_1 and β are unknown. However, for the optimal value $|\Delta^\star(\mathbf{x}_i)|$, it can be easily shown that the ratio between the optimal bias $b^\star(\mathbf{x}_i)$ and the optimal standard deviation $\upsilon^\star(\mathbf{x}_i)$ is image independent (see [43, 71]):

$$\frac{|b^\star(\mathbf{x}_i)|}{\upsilon^\star(\mathbf{x}_i)} \; \leq \; \sqrt{\frac{\beta}{3}} \; \overset{def}{=} \; \gamma.$$

Accordingly, an ideal choice of the window will be the largest window $\Delta_{i,n}$ such that $\widehat{b}_{i,n}$ is still not larger than $\gamma\widehat{v}_{i,n}$, for some real value $\gamma \in \mathbb{R}_+$:

$$\Delta^\star(\mathbf{x}_i) \;=\; \sup_{\Delta_{i,n} \in \mathcal{N}_\Delta} \{|\Delta_{i,n}| : \widehat{b}_{i,n} \leq \gamma\widehat{v}_{i,n}\}.$$

Since $\Delta^\star(\mathbf{x}_i)$ is based on the full knowledge of the bias (and then of the unknown function u) rather than a data-dependent selection, it represents an ideal we cannot expect to attain. Since, in the ideal case, the ideal window $\Delta^\star(\mathbf{x}_i)$ exactly balances the bias and variance terms, i.e. $b^\star(\mathbf{x}_i) = \gamma v^\star(\mathbf{x}_i)$, the corresponding *pointwise ideal risk* is then of the form

$$\Gamma_i(u^\star, u) \overset{def}{=} \left[\mathbb{E}|u^\star(\mathbf{x}_i) - u(\mathbf{x}_i)|^2\right]^{1/2} \;=\; \sqrt{1+\gamma^2}\; v^\star(\mathbf{x}_i). \tag{15}$$

In practice, the bias is not observable and the trade-off between bias and variance cannot be obtained by sweeping out the measured bias $\widehat{b}_{i,n}$ and variance $\widehat{v}^2_{i,n}$ indexed by the smoothing window $\Delta_{i,n}$ at point $\mathbf{x}_i$. Henceforth, we need more precise characterizations to derive a selection procedures: first, following the standard decomposition of the estimator $\widehat{u}_{i,n}$ [50]

$$\widehat{u}_{i,n} = u(\mathbf{x}_i) + \widehat{b}_{i,n} + \nu_i \tag{16}$$

where $\nu_i \sim \mathcal{N}(0, \mathbb{E}[\nu_i^2])$, we have $\mathbb{E}[\widehat{u}_{i,n}] = u(\mathbf{x}_i) + \widehat{b}_{i,n}$ and $\mathbb{E}[\nu_i^2] = \mathbb{E}[|\widehat{u}_{i,n} - u(\mathbf{x}_i) - \widehat{b}_{i,n}|^2] \overset{def}{=} \widehat{v}^2_{i,n}$. Therefore, the following inequality

$$|\widehat{u}_{i,n} - u(\mathbf{x}_i)| \leq \widehat{b}_{i,n} + \varkappa\,\widehat{v}_{i,n} \tag{17}$$

holds with a high probability and $0 < \varkappa < \infty$. In this inequality, the right hand side is comprised of two terms, the deterministic dynamic error $\widehat{b}_{i,n}$ which is completely determined by u, and the stochastic error which is completely independent of u. Finally, since $\widehat{b}_{i,n} \leq \gamma\widehat{v}_{i,n}$, we modify correspondingly the definition of the *ideal* window as

$$\Delta^\star(\mathbf{x}_i) \;=\; \sup_{\Delta_{i,n} \in \mathcal{N}_\Delta} \{|\Delta_{i,n}| : |\widehat{u}_{i,n} - u(\mathbf{x}_i)| \leq (\gamma + \varkappa)\,\widehat{v}_{i,n}\}. \tag{18}$$

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

4.3 A data-driven local window selector

In the pointwise estimation approach, we strongly suppose $\widehat{v}_{i,n}$ decreases as n increases and the ordering relation $\widehat{u}_{i,n'} \preccurlyeq \widehat{u}_{i,n}$ that implies $\widehat{v}_{i,n} \leq \widehat{v}_{i,n'}$, can be introduced. If this assumption is not fulfilled for the original set $\mathcal{N}_\Delta$, i.e. there is $\Delta_{i,n'} \subset \Delta_{i,n}$ with the property $\widehat{v}_{i,n'} > \widehat{v}_{i,n}$, then we simply exclude the window $\Delta_{i,n'}$ from $\mathcal{N}_\Delta$ at point $\mathbf{x}_i$. The collection of estimators $\{\widehat{u}_{i,0}, \ldots, \widehat{u}(\mathbf{x}_i)\}$ is then naturally ordered in the direction of increasing $|\Delta_{i,n}|$

where $\widehat{u}(\mathbf{x}_i)$ can be thought as the best possible estimator. Accordingly, we propose a selection procedure based on pairwise comparisons of an essentially one-dimensional family of competing estimators $\widehat{u}_{i,n}$ as described below.

Actually, the random variables $\widehat{u}_{i,n} - \widehat{u}_{i,n'}$ are Gaussian random variables with expectations equal to the bias differences $\widehat{b}_{i,n} - \widehat{b}_{i,n'}$ and variances to $\text{Var}(\widehat{u}_{i,n} - \widehat{u}_{i,n'}) \leq \widehat{v}^2_{i,n'}$ (see the proof in Appendix A.1.). From (17) and (18), it follows

$$\begin{aligned}
|\widehat{u}_{i,n'} - \widehat{u}_{i,n}| \leq |\widehat{b}_{i,n'} - \widehat{b}_{i,n} + \varkappa \text{Var}(\widehat{u}_{i,n} - \widehat{u}_{i,n'})| &\leq |\widehat{b}_{i,n'} - \widehat{b}_{i,n}| + \varkappa \widehat{v}_{i,n'} \\
&\leq |\widehat{b}_{i,n'}| + |\widehat{b}_{i,n}| + \varkappa \widehat{v}_{i,n'} \\
&\leq \gamma \widehat{v}_{i,n'} + \gamma \widehat{v}_{i,n} + \varkappa \widehat{v}_{i,n'} \\
&\leq (2\gamma + \varkappa)\widehat{v}_{i,n'}.
\end{aligned} \tag{19}$$

Now if a threshold $\varrho = (2\gamma + \varkappa)$ is properly chosen, none of the variables $|\widehat{u}_{i,n'} - \widehat{u}_{i,n}|$ will exceed the value ϱ with a high probability. Thus, the fulfillment of the event

$$\left| \frac{\widehat{u}_{i,n} - \widehat{u}_{i,n'}}{\widehat{v}_{i,n'}} \right| \leq \varrho, \quad 1 \leq n' \leq n, \tag{20}$$

will suggest that $\widehat{u}_{i,n}$ can be a good candidate for the "true" estimator. Among all such candidates, one naturally choose the one with the smallest variance $\widehat{v}^2_{i,n}$. The exact choice of the threshold ϱ then becomes a balancing act based on large deviation calculations: ϱ should be large enough to guarantee a sufficiently high probability of (20) and, at the same time, small enough to provide a good control of $\widehat{u}_{i,n} - \widehat{u}(\mathbf{x}_i)$. However, it worth noting that setting the threshold ϱ remains an open issue in practice as already mentioned in [43] and in a recent reliability study for a fine adjustment of $\varkappa$ and γ in signal processing applications [71]. In the next section, we shall see how this parameter can be estimated from image data. Following the above discussion, a window selector is then based on the following pointwise rule [49, 52, 42, 50, 36]:

$$\widehat{\Delta}(\mathbf{x}_i) \;=\; \arg \max_{\Delta_{i,n} \in \mathcal{N}_\Delta} \{|\Delta_{i,n}| : \quad |\widehat{u}_{i,n} - \widehat{u}_{i,n'}| \leq \varrho\, \widehat{v}_{i,n'} \;\text{ for all }\; 1 \leq n' < n\}. \tag{21}$$

This choice ensures the balance between the stochastic term and bias and means that we take the largest window such that the estimators $\widehat{u}_{i,n}$ and $\widehat{u}_{i,n'}$ are not too different, in some sense, for all $1 \leq n' \leq n$. Hence, if an estimated point $\widehat{u}_{i,n'}$ 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. In this case, rejecting $\widehat{u}_{i,n}$ in favor of $\widehat{u}_{i,n'}$, $n' < n$, as the procedure prescribes, would result in a reduction of the bias more substantial than then the increasing of the variance. For each pixel, the detection of this transition enables to determine the critical size that balances bias and variance. Out of all the windows which have not been thus rejected, the one corresponding to the smallest variance, namely $\widehat{\Delta}(\mathbf{x}_i)$ is used in the construction of the estimator $\widehat{u}(\mathbf{x}_i)$. This idea underlying our construction definitely belongs to Lepskii [48, 49].

5 Theoretical properties

5.1 Accuracy of the adaptive estimate

For adaptive estimation at a point $\mathbf{x}_i$, one must pay a price for adaptation. The extra factor is a function of $\varkappa$ and γ and is an unavoidable payment for pointwise adaptation. Under fair assumptions, the following proposition can be proved (see Appendix A.2).

PROPOSITION 1. *If the ideal risk is defined as*

$$\Gamma_i(u^\star, u) = \sqrt{1+\gamma^2}\, \upsilon^\star(\mathbf{x}_i)$$

and iif $n^\star(\mathbf{x}_i) \leq \widehat{n}(\mathbf{x}_i)$*, then there exists an optimal adaptive estimate* $\widehat{u}(\mathbf{x}_i)$ *with the inaccuracy of order* $\left[\frac{(2\gamma+\varkappa)}{\sqrt{1+\gamma^2}} + 1\right]$ *times the ideal risk. i.e.*

$$\left[\mathbb{E}_{n^\star(\mathbf{x}_i)\leq\widehat{n}(\mathbf{x}_i)} |\widehat{u}(\mathbf{x}_i) - u(\mathbf{x}_i)|^2\right]^{1/2} \leq \left[\frac{(2\gamma+\varkappa)}{\sqrt{1+\gamma^2}} + 1\right] \Gamma_i(u^\star, u).$$

It is clear that the ideal risk is unattainable in general but, iif $n^\star(\mathbf{x}_i) \leq \widehat{n}(\mathbf{x}_i)$, our estimator has (up to a constant term) the same risk as the risk of the estimator constructed with the ideal window $\Delta^\star(\mathbf{x}_i)$.

Clearly, the choice of the parameter ϱ plays an important role in the adaptation and must be carefully chosen. In order to calibrate this threshold, we need to evaluate the probability of the event $\{\widehat{n}(\mathbf{x}_i) = n\}$ at $\mathbf{x}_i$ and prove the following proposition:

PROPOSITION 2. *The event* $\{\widehat{n}(\mathbf{x}_i) = n\}$ *occurs at* $\mathbf{x}_i$ *with a probability*

$$\mathbb{P}\{\widehat{n}(\mathbf{x}_i) = n\} \leq \sum_{n'<n} 2\exp\left(-\frac{\varrho^2}{2}\right) \leq 2N_\Delta \exp\left(-\frac{\varrho^2}{2}\right).$$

Proof: see Appendix A.3.

From PROPOSITION 2, it follows

$$\begin{aligned} 1 - \mathbb{P}\{\widehat{n}(\mathbf{x}_i) \geq N_\Delta\} = \sum_{n=1}^{N_\Delta-1} \mathbb{P}\{\widehat{n}(\mathbf{x}_i) = n\} &\leq \sum_{n=1}^{N_\Delta-1}\sum_{n'=1}^{n} 2\exp\left(-\frac{\varrho^2}{2}\right) \\ &\leq \sum_{n=1}^{N_\Delta-1} 2n\exp\left(-\frac{\varrho^2}{2}\right) \\ &\leq N_\Delta(N_\Delta-1)\exp\left(-\frac{\varrho^2}{2}\right). \end{aligned}$$

Hence, if we fix the probability $\mathbb{P}\{\widehat{n}(\mathbf{x}_i) \geq N_\Delta\}$, an upper bound can be calculated to choose ϱ as

$$\varrho \leq \sqrt{2 \log \frac{N_\Delta(N_\Delta - 1)}{1 - \mathbb{P}\{\widehat{n}(\mathbf{x}_i) \geq N_\Delta\}}}. \tag{22}$$

Moreover, we point out that most images are piecewise smooth. They contain a small number of pixels with significant discontinuities. Therefore, the probability $\mathbb{P}\{\widehat{n}(\mathbf{x}_i) \geq N_\Delta\}$ is high at $\mathbf{x}_i \in G$. The threshold ϱ can be then adapted to image contents using the following approximation

$$\mathbb{P}\{\widehat{n}(\mathbf{x}_i) \geq N_\Delta\} \;\approx\; \mathbb{P}\{|r_i| \leq \widehat{\sigma}\} \;=\; \frac{\#\{\mathbf{x}_i : |r_i| \leq \widehat{\sigma}\}}{|G|}, \tag{23}$$

i.e. from the empirical distribution of pseudo-residuals $\{r_i\}$ defined by (25). In what follows, we adopt the upper bound (22) (and the approximation (23)) to derive a data-driven estimation of ϱ.

Finally, it remains to evaluate $\mathbb{P}\{\widehat{n}(\mathbf{x}_i) < n^\star(\mathbf{x}_i)\}$ and to prove that we nearly never under-estimate the optimal window $\Delta^\star(\mathbf{x}_i)$:

PROPOSITION 3. *The event* $\{\widehat{n}(\mathbf{x}_i) < n^\star(\mathbf{x}_i)\}$ *occurs at* $\mathbf{x}_i$ *with a probability*

$$\mathbb{P}\{\widehat{n}(\mathbf{x}_i) < n^\star(\mathbf{x}_i)\} \leq n^\star(\mathbf{x}_i)(n^\star(\mathbf{x}_i) - 1) \exp\left(-\frac{\varrho^2}{2}\right).$$

Proof: due to the definition of $\widehat{n}(\mathbf{x}_i)$, we have

$$\{\widehat{n}(\mathbf{x}_i) < n^\star(\mathbf{x}_i)\} \subseteq \bigcup_{n < n^\star(\mathbf{x}_i)} \bigcup_{n' < n} \{|\widehat{u}_{i,n} - \widehat{u}_{i,n'}| > \varrho \widehat{v}_{i,n'}\}.$$

and the probability of the event $\{n \leq n^\star(\mathbf{x}_i)\}$ occurs can bounded

$$\mathbb{P}\{\widehat{n}(\mathbf{x}_i) < n^\star(\mathbf{x}_i)\} \quad \leq \quad \sum_{n=1}^{n^\star(\mathbf{x}_i)-1} \sum_{n'=1}^{n} \mathbb{P}\{|u^\star(\mathbf{x}_i) - \widehat{u}_{i,n}| > \varrho \widehat{v}_{i,n}\}.$$

In Appendix A.3, we proved $\mathbb{P}\{|\widehat{u}_{i,n} - \widehat{u}_{i,n'}| > \varrho \widehat{v}_{i,n'}\} \leq 2\exp\left(-\varrho^2/2\right)$. Hence, we have

$$\mathbb{P}\{\widehat{n}(\mathbf{x}_i) < n^\star(\mathbf{x}_i)\} \quad \leq \quad \sum_{n=1}^{n^\star(\mathbf{x}_i)} \sum_{n'=1}^{n} 2\exp\left(-\frac{\varrho^2}{2}\right) \;=\; n^\star(\mathbf{x}_i)(n^\star(\mathbf{x}_i) - 1) \exp\left(-\frac{\varrho^2}{2}\right)$$

$\square$

This last inequality can be checked to hold true for any positive values of ϱ. The probability of the event $\{\widehat{n}(\mathbf{x}_i) < n^\star(\mathbf{x}_i)\}$ is small, provided that the value $n^\star(\mathbf{x}_i)(n^\star(\mathbf{x}_i) - 1)\exp(-\varrho^2/2)$ is sufficiently small. Accordingly, if $2.5 \leq \varrho \leq 3$ (as shown in our experiments) and if we choose $N_\Delta = 4$, the probability of the event $\{\widehat{n}(\mathbf{x}_i) \geq n^\star(\mathbf{x}_i))\}$ occurs with a high probability.

5.2 Invariance properties

We prove some additional invariance properties of our estimator. In particular it satisfies the following properties:

- *Average gray level invariance:* We study an idealized situation when the underlying image is a unique flat region. In this situation, we can show the estimate $\widehat{u}(\mathbf{x}_i)$ coincide with the mean values of observations Y_i for all $\mathbf{x}_i \in G$ with a probability close to 1:

  $$\lim_{\widehat{\Delta}(\mathbf{x}_i)\to G} \sum_{\mathbf{x}_j \in \widehat{\Delta}(\mathbf{x}_i)} \pi_{i\sim j,\widehat{n}(\mathbf{x}_i)}\, Y_j \;=\; \frac{1}{|G|}\sum_{\mathbf{x}_j\in G} Y_j$$

  Additionally, the deviations $\widehat{u}(\mathbf{x}_i) - u_0$ are of the order $|G|^{-1/2}$:

  $$\lim_{\widehat{\Delta}(\mathbf{x}_i)\to G} \widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \widehat{\Delta}(\mathbf{x}_i)} [\pi_{i\sim j,\widehat{n}(\mathbf{x}_i)}]^2 \;=\; \frac{\widehat{\sigma}^2}{|G|}$$

  We may apply the Bernstein's inequality which tells us that the deviations of $\widehat{u}(\mathbf{x}_i) - u_0$ are bounded:

  $$\mathbb{P}\{|\widehat{u}(\mathbf{x}_i) - u_0| \geq \delta\} \;\leq\; 2\exp\left(-\frac{3|G|^2\delta^2}{6\widehat{\sigma}^2 + 2C_2|G|\delta}\right) \underset{|G|\to\infty}{\longrightarrow} 0$$

  for any $\delta \in \mathbb{R}_+$ since we can find $C_2 \in \mathbb{R}_+$ such that $|\widehat{u}(\mathbf{x}_i) - \mathbb{E}(\widehat{u}(\mathbf{x}_i))| \leq C_2$.

- *Extremum principle:* The estimator fulfills the extremum principle which offers the practical advantage that, if we start with an input image within the range $[0, 255]$, we will never obtain results with gray level outside this range. If $\inf_{\mathbf{x}_j\in G} Y_j \leq Y_j \leq \sup_{\mathbf{x}_j\in G} Y_j, \forall \mathbf{x}_j \in G$, then

  $$\begin{array}{ccccc} \displaystyle\sum_{\mathbf{x}_j\in\Delta_{i,n}} \pi_{i\sim j,n} \inf_{\mathbf{x}_j\in G} Y_j & \leq & \displaystyle\sum_{\mathbf{x}_j\in\Delta_{i,n}} \pi_{i\sim j,n} Y_j & \leq & \displaystyle\sum_{\mathbf{x}_j\in\Delta_{i,n}} \pi_{i\sim j,n} \sup_{\mathbf{x}_j\in G} Y_j \\ \inf_{\mathbf{x}_j\in G} Y_j & \leq & \widehat{u}_{i,n} & \leq & \sup_{\mathbf{x}_j\in G} Y_j \end{array}$$

Some other properties should be further investigated.

5.3 Complexity

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

6 implementation

In this section, we describe a practical algorithm given in Fig. 2, with a minimal number of calibrated parameters, based on the previous construction of the adaptive window and the corresponding estimators. The key ingredient of the procedure is an increasing sequence of nested square windows, centered at $\mathbf{x}_i$, of size $|\Delta_{i,n}| = (2^n+1)\times(2^n+1)$ pixels with $n = 1, \ldots, N_\Delta$. At the initialization, we naturally choose $|\Delta_{i,0}| = 1$ and set the fixed size of $p \times p$ patches and the level α of the hypothesis test for image patch comparison. In addition, the estimation procedure relies on the preliminary estimation of the noise variance $\widehat{\sigma}^2$ robustly estimated from input data as

$$\widehat{\sigma} = 1.4826\ \mathrm{med}(|\,\mathbf{r} - \mathrm{med}|\mathbf{r}|\,|) \tag{24}$$

where $\mathbf{r} = \{r_1, r_2, \ldots, r_{|G|}\}$ 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 $i = (i_1, i_2)$) (see [30]):

$$r_i = [2Y_{i_1,i_2} - (Y_{i_1+1,i_2} + Y_{i_1,i_2+1})]/\sqrt{6}. \tag{25}$$

and the constant $1/\sqrt{6}$ is used to ensure $\mathbb{E}[r_i^2] = \widehat{\sigma}^2$ in homogeneous regions. A stopping rule (e.g Csiszár's I-divergence [21]) could be used to save computing time if two successive solutions are very close and prevent an useless setting of the larger window size [44]. However, we manually set the number N_Δ of iterations to bound the numerical complexity (typically $N_\Delta = 4$). As expected, increasing N_Δ allows for additional variance reduction in homogeneous regions.

7 Experimental results

The proposed methodology is used for image denoising in various contexts, including object detection in bio-imaging. Our results were measured by the peak signal-to-noise ratio (PSNR) in decibels (db) defined as

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

where u_0 is the noise-free original image.

In our experiments, the noise variance $\widehat{\sigma}^2$ and the threshold ϱ are autonomously estimated from image data using (24) and (22) and we have chosen $p = 9$ and $N_\Delta = 4$. In all the experiments, we shall see $\varrho \leq 3$, which implies the optimal adaptive estimate $\widehat{u}(\mathbf{x}_i)$ is with the inaccuracy of order 3 times the ideal risk if $\gamma \leq 1$ (see PROPOSITION 1.). This value corresponds to a local neighborhood of maximal size of 17×17 pixels. Moreover, the choice of the critical values λ_α is also important. Large values improve stability of the model under homogeneity but result in a low sensitivity to parameter changes, while too small critical values lead to a large “false alarm” probability. In our experiments, this “false alarm” probability should not exceed the given level α set to 0.01 yielding to $\lambda_{0.01} = \chi^2_{81,0.99} = 113.5$.

Algorithm Exemplar-based image restoration algorithm

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

Initialization: compute $\widehat{\sigma}^2$, $\varrho$ and $\{\widehat{u}_{i,0}, \widehat{\upsilon}^2_{i,0}\}$ for each $\mathbf{x}_i \in G$.

Repeat
    • for each $\mathbf{x}_i \in G$
        – compute
            $\pi_{i\sim j,n} = \frac{K\left(\lambda_\alpha^{-1}\,\mathrm{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1})\right)}{\sum_{\mathbf{x}_j \in \Delta_{i,n}} K\left(\lambda_\alpha^{-1}\,\mathrm{dist}(\widehat{\mathbf{u}}_{i,n-1}, \widehat{\mathbf{u}}_{j,n-1})\right)}$
            $\widehat{u}_{i,n} = \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n}\, Y_j,$
            $\widehat{\upsilon}^2_{i,n} = \widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n}} [\pi_{i\sim j,n}]^2$
        – test the window using
            $\widehat{\Delta}(\mathbf{x}_i) = \arg\max_{\Delta_{i,n} \in \mathcal{N}_\Delta} \left\{|\Delta_{i,n}| : |\widehat{u}_{i,n} - \widehat{u}_{i,n'}| \le \varrho\, \widehat{\upsilon}_{i,n'}, \text{ for all } 1 \le n' < n\right\}.$
            If this rule is violated at iteration $n$, we do not accept $\widehat{u}_{i,n}$ and keep the estimate $\widehat{u}_{i,n-1}$ as
            the final estimate at $\mathbf{x}_i$, i.e. $\widehat{u}(\mathbf{x}_i) = \widehat{u}_{i,n-1}$ and $\widehat{n}(\mathbf{x}_i) = n - 1$. This estimate is unchanged
            at the next iterations and $\mathbf{x}_i$ is "frozen".
    • increment $n$
while $n = N_\Delta$.
```

Figure 2: Exemplar-based image restoration algorithm.

The processing of a 256×256 image required typically about 1.5 minute ($p = 9$) on a PC (2.6 Ghz, Pentium IV) using a C++ implementation of the algorithm.

7.1 Image restoration on artificially noisy images and comparison to state-of-the-art methods

We have done simulations on a commonly-used set of images available at `http://decsai.ugr.es/~javier/denoise/test_images/` and described in [62]. The potential of the estimation method is mainly illustrated with the 512×512 *lena* image (Fig. 3a) corrupted by an additive white-Gaussian noise (WGN) (Fig. 3b, PSNR = 22.13 db, $\sigma = 20$). In this experiment, we found $\mathbb{P}\{|r_i| \leq \widehat{\sigma}\} = 0.829$ from image data and derived $\varrho = 2.91$ from (22). In Fig. 3c, 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. 3b (magnification factor of 2) and has been estimated by calculating the difference between the noisy image (Fig. 3b) and the recovered image (Fig. 3c). The estimated noise component contains few geometric structures and is similar to a simulated white Gaussian noise. To better appreciate the accuracy of the restoration process, the variance of the pointwise estimator is shown in Fig. 3e 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. Figure 3f shows the probability of a patch $\widehat{\mathbf{u}}(\mathbf{x}_i)$ occurring in $\widehat{\Delta}(\mathbf{x}_i)$:

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

where the set $\Omega(\mathbf{x}_i)$ is used to denote $\{\mathbf{x}_j \in \widehat{\Delta}(\mathbf{x}_i) : \text{dist}(\widehat{\mathbf{u}}(\mathbf{x}_i), \widehat{\mathbf{u}}(\mathbf{x}_j)) \leq \lambda_\alpha\}$. In Fig. 3f, dark values correspond low probabilities of occurrence and, it is confirmed that repetitive patterns in the neighborhood of image discontinuities are mainly located along image level lines. We have compared the performance of our method to several competitive methods: Total Variation (TV) minimizing process [64], bilateral filtering [75], anisotropic diffusion (AD) using a diffusivity function of the type $(1 + |\nabla u|^2/K^2)^{-1}$ [59] and Wiener filtering (WF) (Matlab function `wiener2`). Figures 4a-d shows the results of the four tested methods. We stopped anisotropic diffusion after 150 iterations in order to avoid a over-smoothed image but a decorrelation criterion could be used to stop the diffusion process [54]. The TV minimizing method [64] completely eliminates small textures but also blurs edges when $\lambda_{TV} = 0.01$. If we modify the balance between the fidelity and regularizing terms by choosing $\lambda_{TV} = 0.05$, the image is denoised but smooth parts are not completely recovered. Accordingly, 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 (Fig. 4). Additionally, this noisy image has been restored using pointwise adaptive estimation methods [44, 61] which are not patch-based. Figs. 4e-f provides a visual comparison of image denoising with these two algorithms: the AWS algorithm [61] tends to oversmooth the image and to generate some artificial planar segments in homogeneous regions (Fig. 4f), whereas a variant of this approach [44] yields a similar result (Fig. 4e) to the image regularizing with the TV method [64] (see Fig. 4a). In Fig. 5, the corresponding recovered noise components are shown and most of them contain undesirable geometric structures. Moreover, our approach is also compared to the *non-local means* algorithm [12, 13] 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. 6-8). Our approach is also compared to another and recent patch-based approach applied to image denoising [63], that exploits ideas from sparse image coding and training images for learning Markov random field image priors. The PSNR values are reported in table I for the test images; In most cases, our *unsupervised* and simple method produces the best PSNR values. Finally, 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.

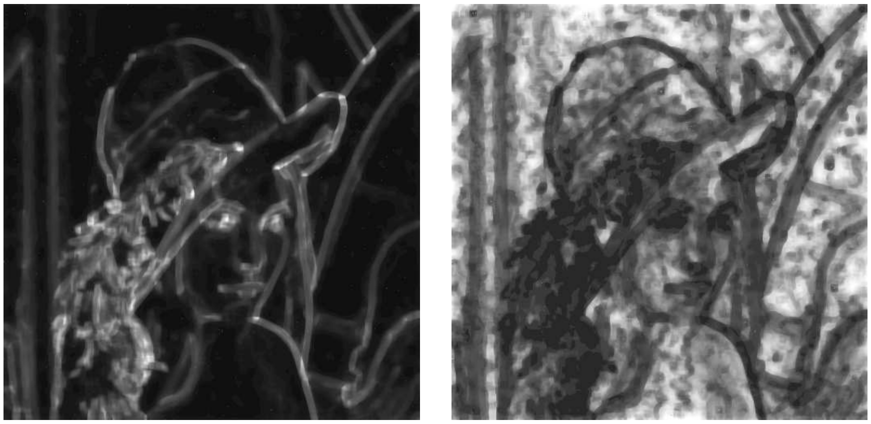
Moreover, we have also examined some complementary aspects of our approach using the artificially corrupted *barbara* image (Figs. 9-12, (WGN) $\sigma = 20$). In this experiment, we found $\mathbb{P}\{|r_i| \leq \widehat{\sigma}\} = 0.806$ using (23) and derived $\varrho = 2.871$ from (22). For illustration , figure 10 shows the four results obtained at each iteration of the algorithm. Finally, we varied the patch size and Fig. 11 shows that taking too small image patches can generate some visually undesirable flat zones during the restoration process (see Fig. 11 d-f). Note, that taking one point every two pixels (in both two directions) in a $p \times p$ patch ($k = p^2/4$) can be applied to produce natural regularized images (Fig. 11 a-c) and reduce the time computing. In addition, Table II shows the PSNR values obtained by varying the patch size and sub-sampling (factor 2). Note the PSNR values are close for every patch size and the optimal patch size depends on the image contents; a 9×9 patch seems appropriate in most cases and a smaller patch can be considered for processing piecewise smooth images. We have also compared our method to the best available published results when very competitive methods [63, 62, 60] were applied to the same image dataset [62]. These results were taken from the corresponding publications. We point out that, visually and quantitatively, our very simple and unsupervised algorithm method favorably compares to any of these denoising algorithms, including the more sophisticated wavelet-based denoising methods (e.g. see Fig. 6c, Fig. 7c and Fig. 8c). Note that our method yields an improved PSNR for a wide range of variance as compared to existing methods. If the PSNR gains are marginal for some images, the visual difference can be significant as shown Fig. 6 where less artifacts are visible using our method. To complete the experiments, Table III shows the PSNR values using our exemplar-based restoration method when applied to this set of test images for a wide range of noise variance. This table can be used for comparison with previously published denoising methods [62, 63].

In the second part of experiments, the effects of the exemplar-based restoration is approach are 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 (Figs. 13-15). However, the visual quality decreases if the image contains small textures to be preserved. In that case, the neighborhood contains a small number of similar patches and the image cannot be reliably denoised (e.g. see Fig. 13). The estimated noise component corresponding to the difference between the denoised image and the noisy image is shown for each processed image (Figs. 13e, 14e and 15e). The variance estimation and the probability $\mathbb{P}\{\widehat{\mathbf{u}}(\mathbf{x}_i)$ occurring in $\widehat{\Delta}(\mathbf{x}_i)\}$ is also shown to describe the image organization (Figs. 13f, 14f and 15f). The robustness to noise is illustrated in Fig. 16 by varying σ from 5 to 50.

In the last part of experiments, the exemplar-based restoration method has been used to denoise original noisy pictures shown in Fig. 17a, Fig. 18a and Fig. 19a. In that case, the noise variance $\widehat{\sigma}^2$ is automatically estimated from image data. The reconstruction of images

(a) original 512×512 image (b) noisy (WGN) image ($\sigma = 20$)

(c) denoised image (d) noise component ($\times 2$)

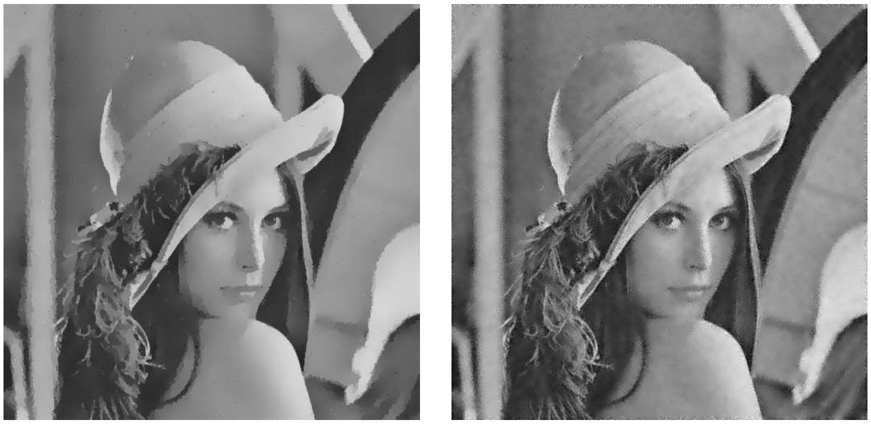


(e) variance of the estimator (f) $\mathbb{P}\{\widehat{\mathbf{u}}(\mathbf{x}_i)$ occurring in $\widehat{\Delta}(\mathbf{x}_i)\}$

Figure 3: Denoising of the noisy (WGN) *lena* 512×512 image ($\sigma = 20$).

(a) TV [64] (PSNR = 30.48)

(b) BF [75] (PSNR = 30.26)



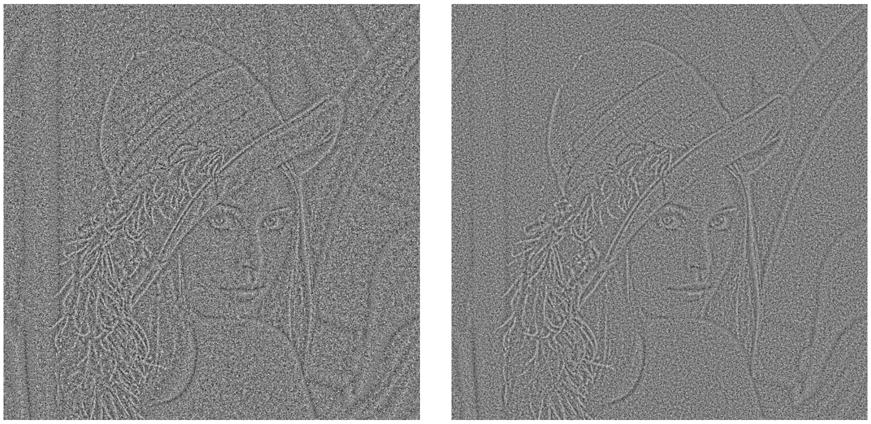
(c) AD [59] (PSNR = 28.83)

(d) WF (PSNR = 28.51)

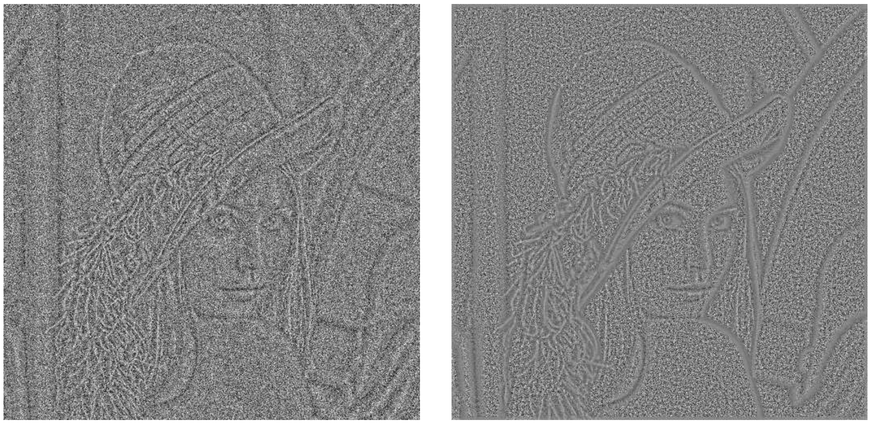
(e) RAWA [44] (PSNR = 30.52)

(f) AWS [61] (PSNR = 29.74)

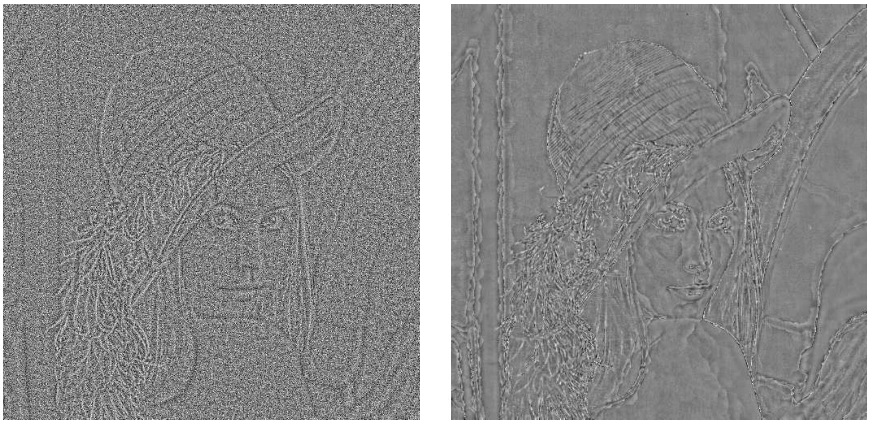
Figure 4: Comparison with restoration methods applied to the noisy (WGN) *lena* image ($\sigma = 20$): (a) Total Variation (TV) minimizing process [64], (b) bilateral filtering (BF) [75], (c) anisotropic diffusion (AD) [59], (d) Wiener filtering (WF), (e) adaptive weights smoothing (AWS) [61], (f) robust adaptive window approach (RAWA) [44].



(a) TV [64] (b) BF [75]



(c) AD [59] (d) WF



(e) RAWA [44] (f) AWS [61]

Figure 5: Estimated noise components by applying restoration methods to the noisy (WGN) *lena* image ($\sigma = 20$): (a) Total Variation (TV) minimizing process [64], (b) bilateral filtering (BF) [75], (c) anisotropic diffusion (AD) [59], (d) Wiener filtering (WF), (e) adaptive weights smoothing (AWS) [61], (f) robust adaptive window approach (RAWA) [44].

(a) exemplar-based restoration (PSNR=32.64)

(b) *non-local means* filter [12] (PSNR=31.09)

(c) wavelet-based denoising [60] (PSNR=32.20)

Figure 6: Comparisons with the *non-local means* algorithm [12] and a wavelet-based denoising method [60] when applied to the noisy (WGN) *lena* image ($\sigma = 20$).

(a) our method (PSNR=32.64)

(b) *non-local means* filter [12] (PSNR=31.09)

(c) wavelet-based denoising [60] (PSNR=32.20)

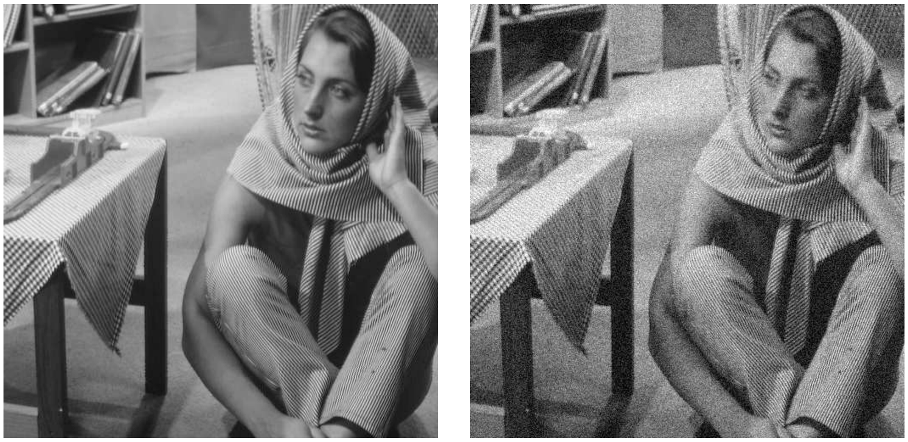
Figure 7: Comparisons with the *non-local means* filter [12] and a wavelet-based denoising method [60] when applied to the noisy (WGN) *lena* image ($\sigma = 20$).

(a) our method (PSNR=32.64)

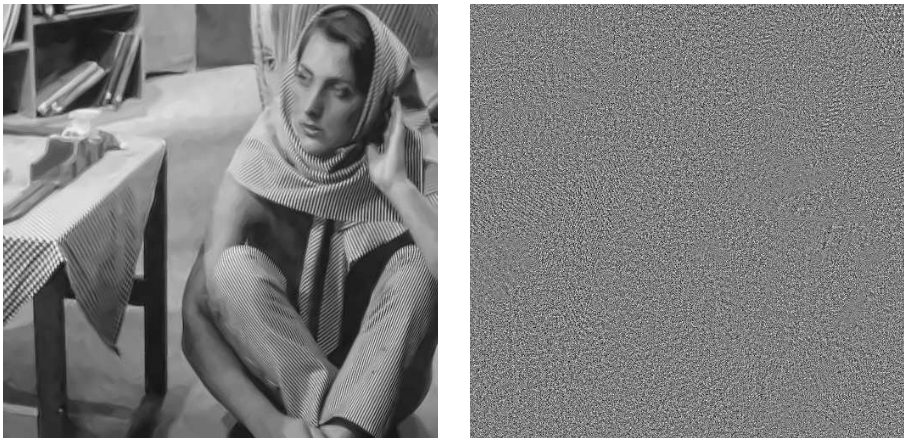
(b) *non-local means* filter [12] (PSNR=31.09)

(c) wavelet-based denoising [60] (PSNR=32.20)

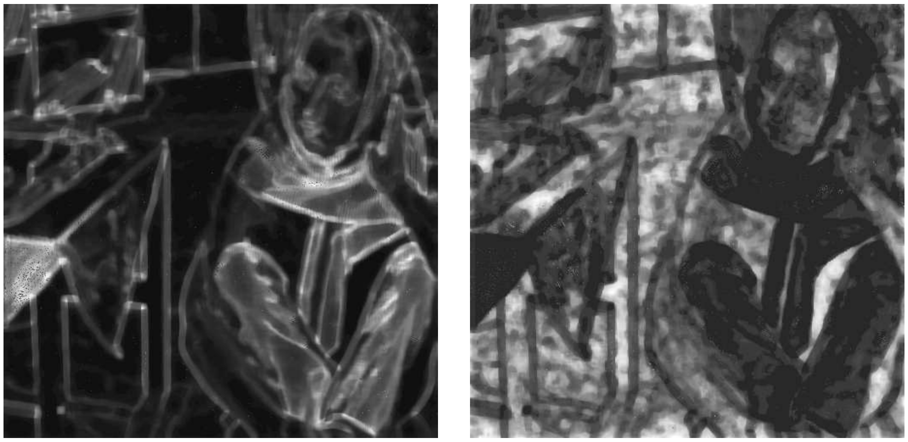
Figure 8: Comparisons with the *non-local means* filter [12] and a wavelet-based denoising method [60] when applied to the noisy (WGN) *lena* image ($\sigma = 20$).



(a) original 512×512 image (b) noisy (WGN) image ($\sigma = 20$)

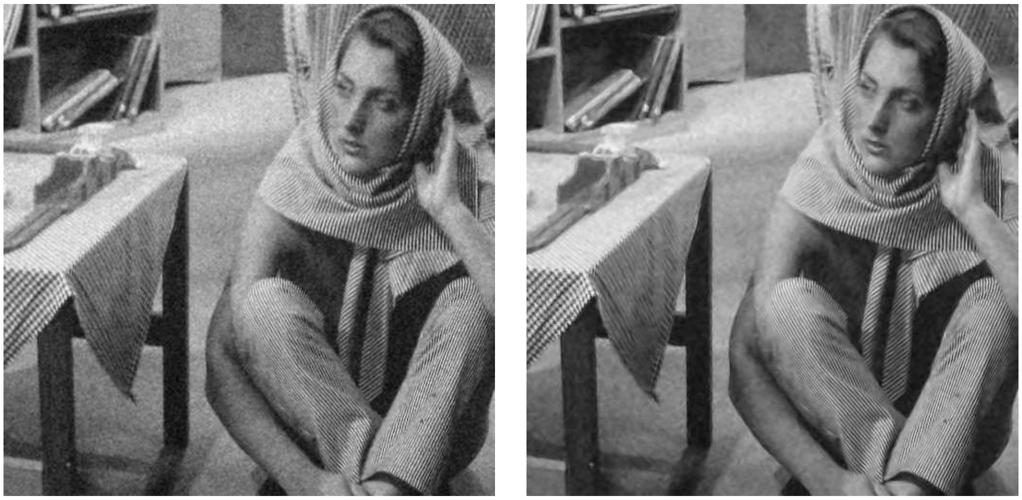


(c) denoised image (PSNR = 30.37) (d) noise component ($\times 2$)



(e) variance of the estimator (f) $\mathbb{P}\{\widehat{\mathbf{u}}(\mathbf{x}_i) \text{ occurring in } \widehat{\Delta}(\mathbf{x}_i)\}$

Figure 9: Densoing of the noisy (WGN) *barbara* 512×512 image ($\sigma = 20$).



(a) iteration #1 (b) iteration #2

(c) iteration #3 (d) iteration #4

Figure 10: Results obtained at each iteration of the algorithm ($N_\Delta = 4$) when applied to the noisy (WGN) *barbara* 512×512 image ($\sigma = 20$).

(a) 3×3 ($k = 9$) patch
(PSNR = 28.97)

(b) 5×5 ($k = 25$) patch
(PSNR = 29.97)

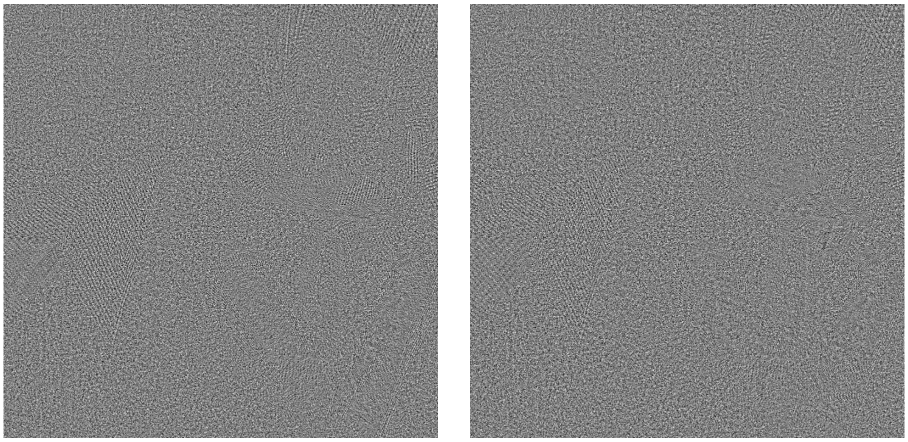
(c) 7×7 ($k = 49$) patch
(PSNR = 30.27)

(d) 9×9 ($k = 81$) patch
(PSNR = 30.37)

(e) sub-sampled 9×9 ($k = 25$) patch
(PSNR = 29.84)

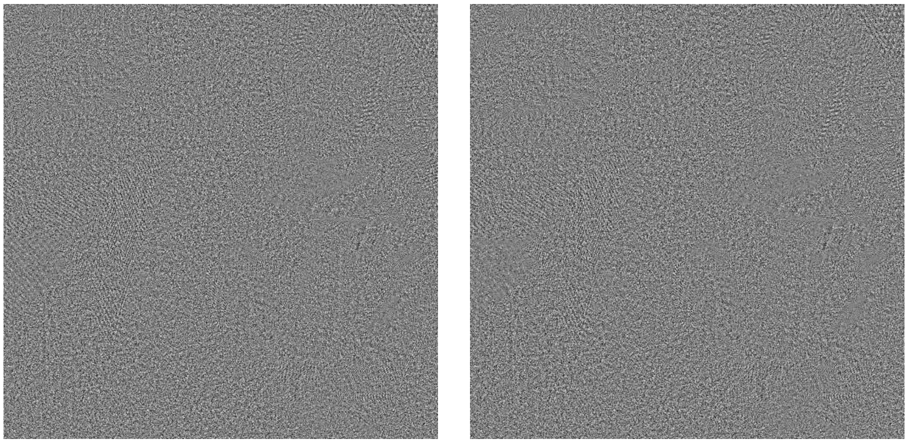
(f) sub-sampled 11×11 ($k = 36$) patch
(PSNR = 29.83)

Figure 11: Results with different patch sizes of $p \times p$ pixels and sub-sampling (factor 2) when the algorithm is applied to the noisy (WGN) *barbara* image ($\sigma = 20$).



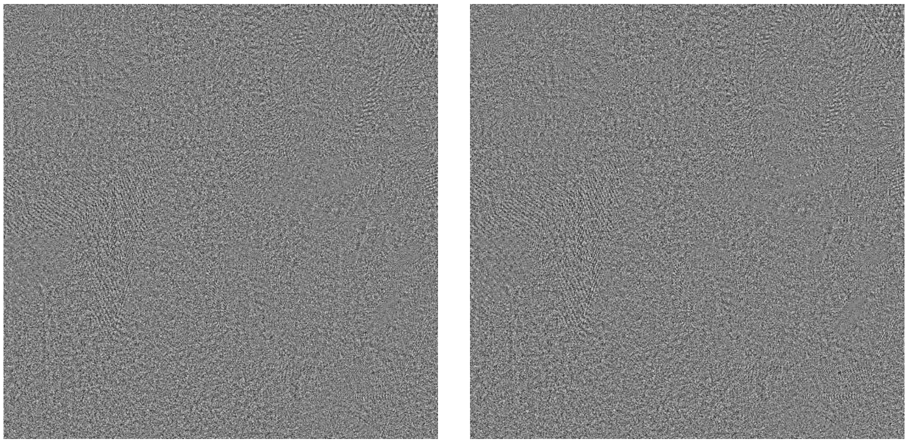
(a) 3×3 ($k = 9$) patch
(PSNR = 28.97) patch

(b) 5×5 ($k = 25$)
(PSNR = 29.97)



(c) 7×7 ($k = 49$) patch
(PSNR = 30.27)

(d) 9×9 ($k = 81$) patch
(PSNR = 30.37)



(e) sub-sampled 9×9 ($k = 25$) patch
(PSNR = 29.84)

(f) sub-sampled 11×11 ($k = 36$) patch
(PSNR = 29.83)

Figure 12: Noise components obtained with several patch sizes of $p \times p$ pixel and sub-sampling (factor 2) when the algorithm is applied to the noisy (WGN) *barbara* image ($\sigma = 20$).

Image	*Lena*	*Barbara*	*Boats*	*House*	*Peppers*
σ/PSNR	20 / 22.13	20 / 22.18	20 / 22.17	20 / 22.11	20 / 22.19
Our method (9 × 9 patch)	**32.64**	$\boxed{\mathbf{30.37}}$	**30.12**	$\boxed{\mathbf{32.90}}$	$\boxed{\mathbf{30.59}}$
Buades *et al.* [12]	31.09	29.38	28.60	31.54	29.05
Ghazel *et al.* [32]	28.50	25.64	26.34	-	-
Kervrann [44]	30.54	26.50	28.01	30.70	28.23
Pizurica *et al.* [60]	32.20	29.53	29.93	-	30.30
Polzehl *et al.* [61]	29.74	26.05	27.74	30.31	28.40
Portilla *et al.* [62]	$\boxed{32.66}$	30.32	$\boxed{30.38}$	32.39	30.31
Roth *et al.* [63]	31.92	28.32	29.85	32.17	30.58
Rudin *et al.* [64]	30.48	27.07	29.02	31.03	28.51
Starck *et al.* [74]	31.95	-	-	-	-
Tomasi *et al.* [75]	30.26	27.02	28.41	30.01	28.88
Wiener filering	28.51	26.99	27.97	28.74	28.10

Table 1: Performances of denoising algorithms when applied to test noisy (WGN) images ($\sigma = 20$).

k points / $p \times p$ patch	$\lambda_{k,0.99}$	*Lena*	*Barbara*	*Boats*	*House*	*Peppers*
		512×512	512×512	512×512	256×256	256×256
9 points / 3×3	21.67	32.13	28.97	29.86	32.69	30.86
25 points / 5×5	44.31	32.52	29.97	30.15	$\boxed{33.05}$	$\boxed{30.98}$
49 points / 7×7	74.92	32.63	30.27	$\boxed{30.17}$	33.03	30.80
81 points / 9×9	113.5	$\boxed{32.64}$	$\boxed{30.37}$	30.12	32.90	30.59
25 points / 9×9	44.31	32.27	29.84	29.64	32.46	30.26
36 points / 11×11	58.62	32.26	29.84	29.51	32.57	29.52

Table 2: PSRN values (db) when our examplar-based restoration method ($N_\Delta = 4, \alpha = 0.01$) with different patch size and sub-sampling (factor 2) is applied to noisy (WGN) images ($\sigma = 20$).

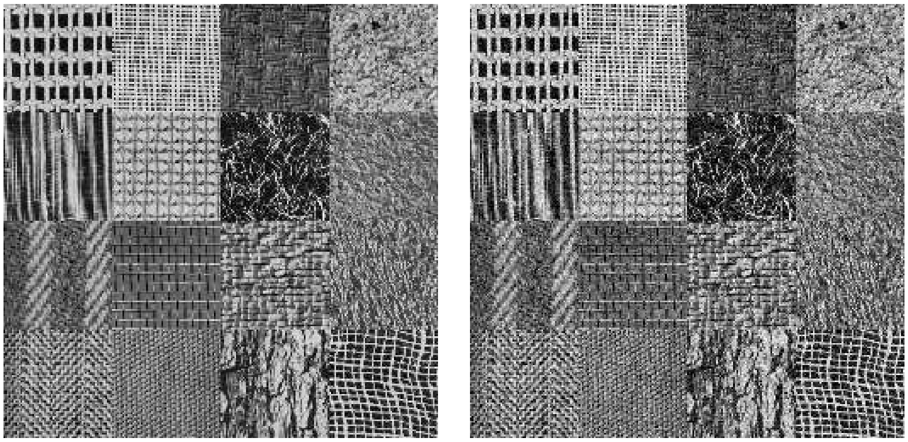
σ/PSNR	*Lena* 512×512	*Barbara* 512×512	*Boats* 512×512	*House* 256×256	*Peppers* 256×256
5 / 34.15	37.91	37.12	36.14	37.62	37.34
10 / 28.13	35.18	33.79	33.09	35.26	34.07
15 / 24.61	33.70	31.80	31.44	34.08	32.13
20 / 22.11	32.64	30.37	30.12	32.90	30.59
25 / 2017	31.73	29.24	29.20	32.22	29.73
50 / 14.15	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

Table 3: Performances of our exemplar-based restoration method ($p = 9, N_\Delta = 4, \alpha = 0.01$) when applied to test noisy (WGN) images.

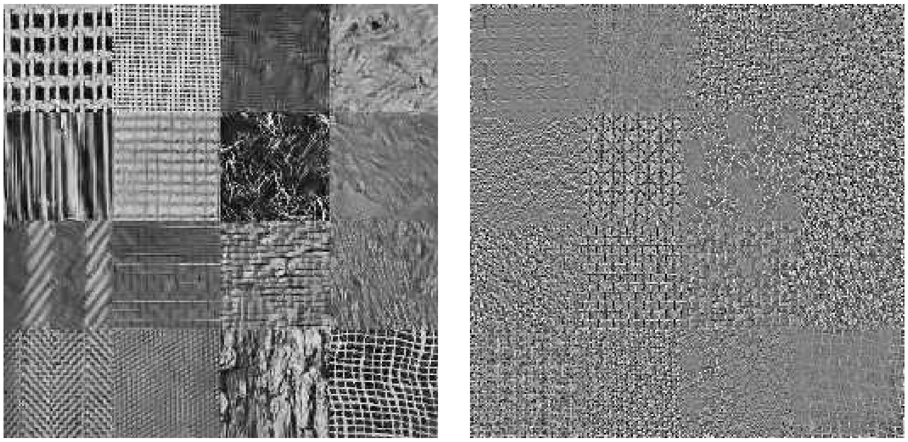
is respectively shown in Fig. 17a and Fig. 17b. Note that edges and geometric structures are well preserved and the noise component corresponding to the paper ageing (Fig. 18c) or fine texture in paints is removed (Figs. 17b and 19b).

7.2 Image denoising and detection of exceptional patterns in bio-imaging

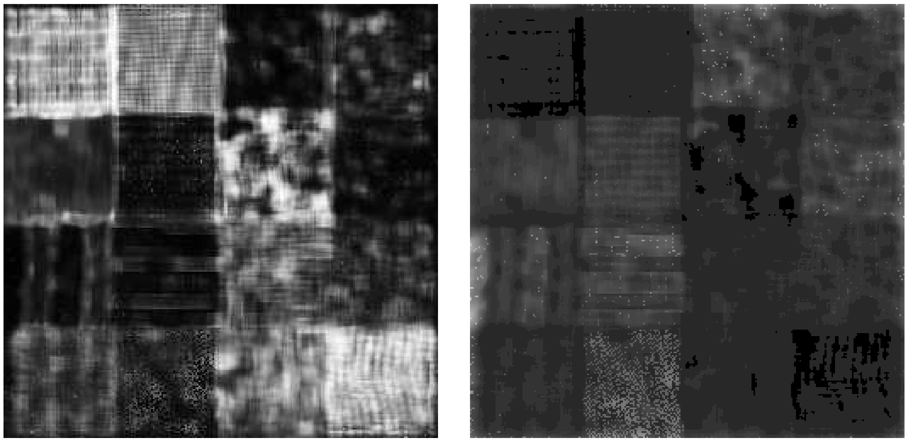
We have applied the exemplar-based restoration method to noisy 2D (360×445) images extracted from a temporal 3D+time sequence of 120 images, showing a large number of small fluorescently labeled vesicles in regions close to the Golgi apparatus (Fig. 20a, courtesy of Institut Curie). We mainly focus on the analysis of these vesicles that deliver cellular components to appropriate places within cells. Quantitative analysis of data obtained by fast 4D deconvolution microscopy allows to enlighten the role of specific Rab proteins. The role of Rab proteins is viewed as to organize membrane platforms serving for protein complexes to act at the required site withing the cell. Methods have been developed for a target protein - Rab6a' - involved in the regulation of transport from the Golgi apparatus to the endoplasmic reticulum. Typically, the state of Golgi membranes during mitosis is controversial, and the role of Golgi-intersecting traffic in Golgi inheritance is unclear and then investigated. In Fig. 20a, two types of objects are observed on three images from the 3D stack: i) organelle (Golgi), ii) static and moving vesicles. The signal-to-noise ratio is shown to be drastically improved resulting enhanced objects in Fig. 20b. Note the the variance of the noise has been first stabilized in order to transform Poisson noise corrupted data into Gaussian noise corrupted data [74, 11]. Moreover, we propose to focus on the detection of "rare" elements in this image. Traditionally, the "rare" elements are local distinctive features for which the intensity abruptly changes in the image. Here, this notion is intended to be quite general and includes spots, various curvature maxima and not repeated patterns in a variable neighborhood.



(a) original 512×512 image (b) noisy (WGN) image ($\sigma = 20$)



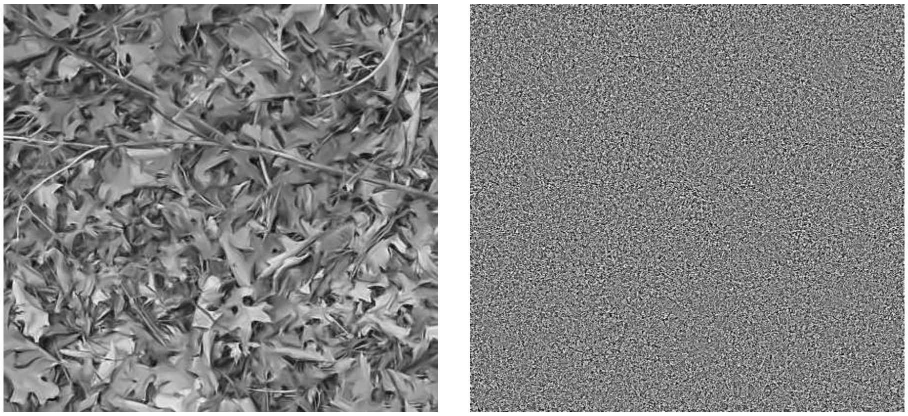
(c) denoised image (d) noise component ($\times 2$)



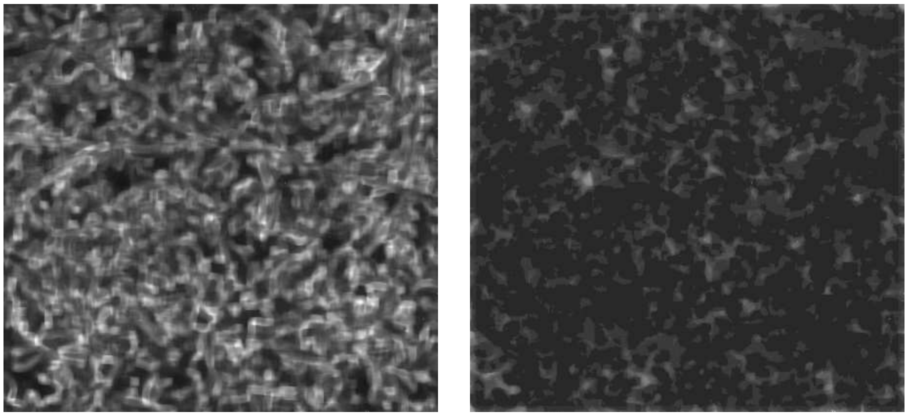
(e) variance of the estimator (f) $\mathbb{P}\{\widehat{\mathbf{u}}(\mathbf{x}_i)$ occurring in $\widehat{\Delta}(\mathbf{x}_i)\}$

Figure 13: Denoising of a noisy (WGN) 256×256 texture image ($\sigma = 20, \varrho = 2.71$).

(a) original 512×512 image (b) noisy (WGN) image ($\sigma = 30$)

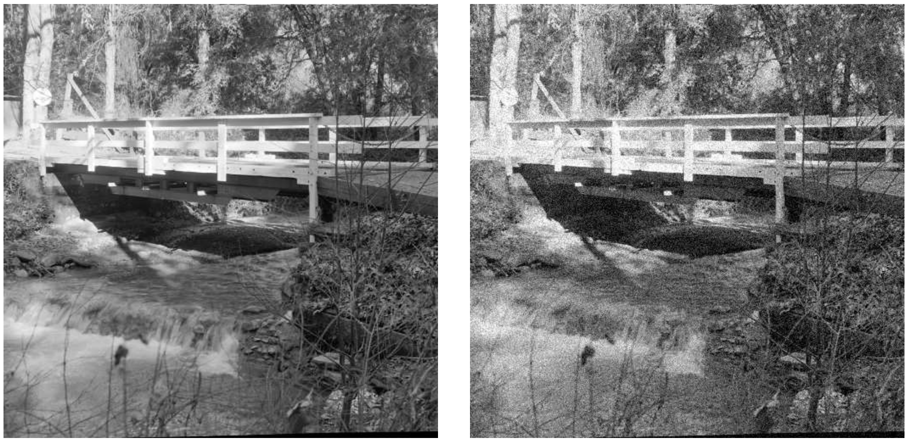


(c) denoised image (d) noise component ($\times 2$)

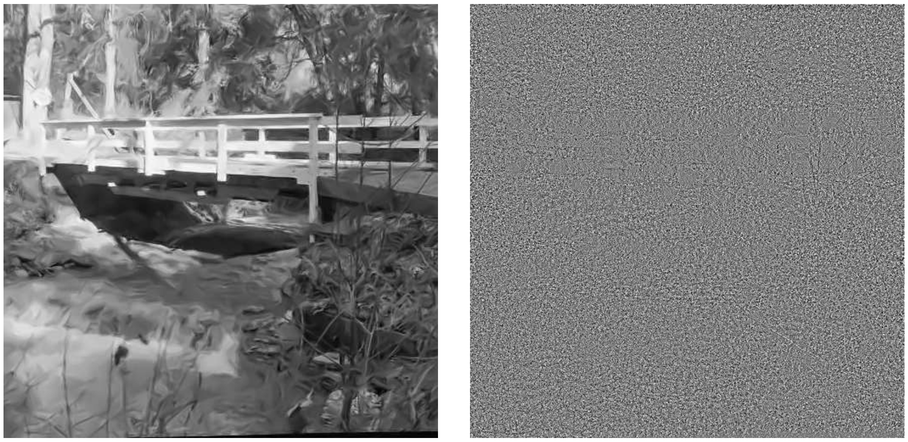


(e) variance of the estimator (f) $\mathbb{P}\{\widehat{\mathbf{u}}(\mathbf{x}_i)$ occurring in $\widehat{\Delta}(\mathbf{x}_i)\}$

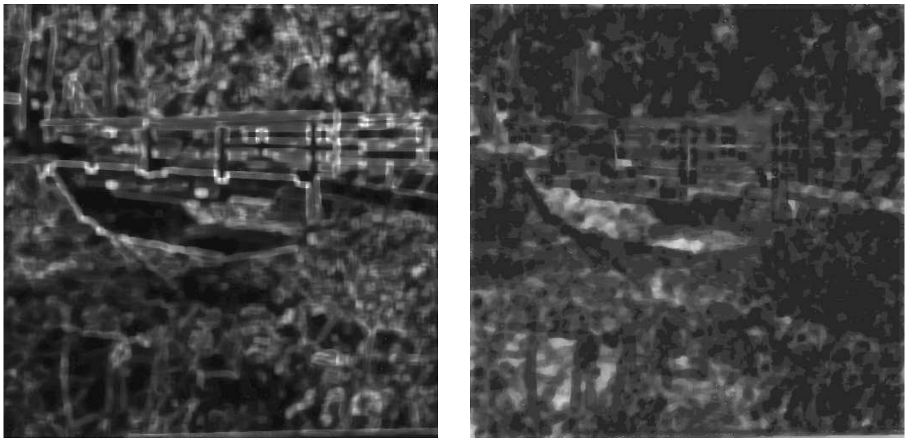
Figure 14: Denoising of a noisy (WGN) textured 445×417 image ($\sigma = 30, \varrho = 2.87$).



(a) original 512×512 image (b) noisy (WGN) image ($\sigma = 20$)

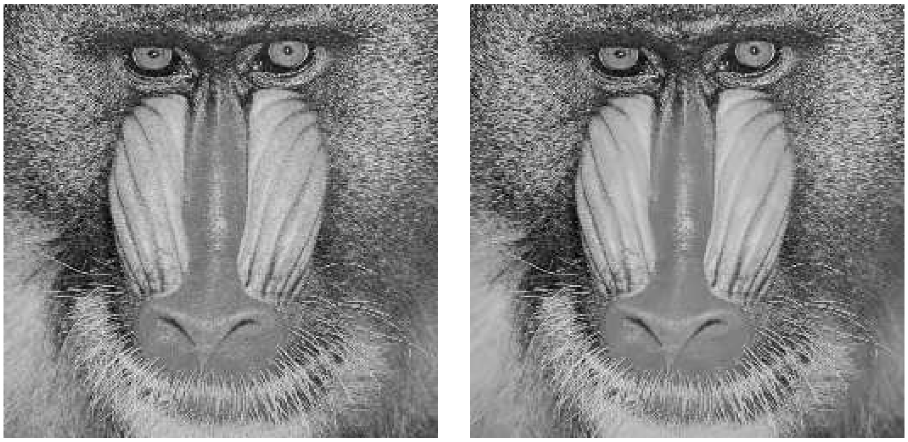


(c) denoised image (d) noise component ($\times 2$)

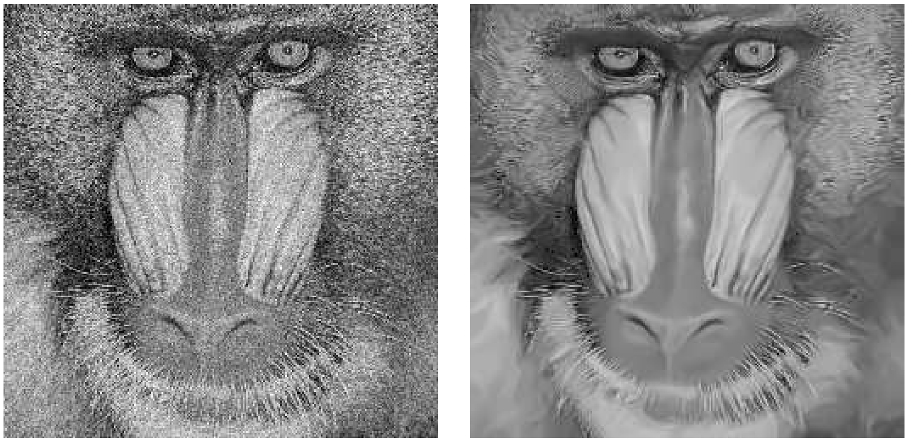


(e) variance of the estimator (f) $\mathbb{P}\{\widehat{\mathbf{u}}(\mathbf{x}_i) \text{ occurring in } \widehat{\Delta}(\mathbf{x}_i)\}$

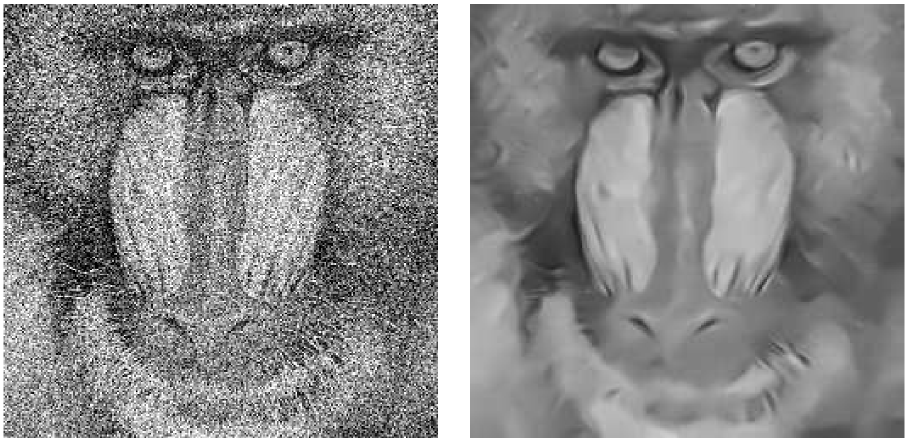
Figure 15: Denoising of a noisy (WGN) 512×512 *bridge* image ($\sigma = 20, \varrho = 2.87$).



(a) noisy image ($\sigma = 5$, PSNR = 34.15) (b) denoised image (PSNR = 34.22)

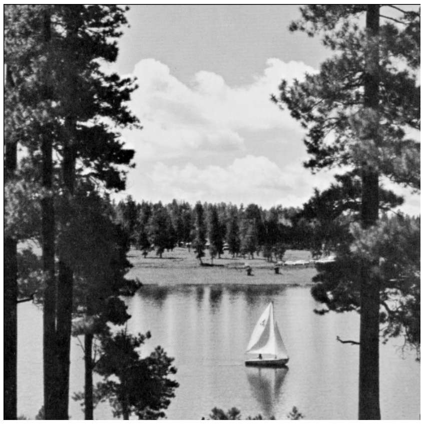


(c) noisy image ($\sigma = 20$, PSNR = 22.09) (d) denoised image (PSNR = 23.34)



(e) noisy image ($\sigma = 50$, PSNR = 14.05) (f) denoised image (PSNR = 20.04)

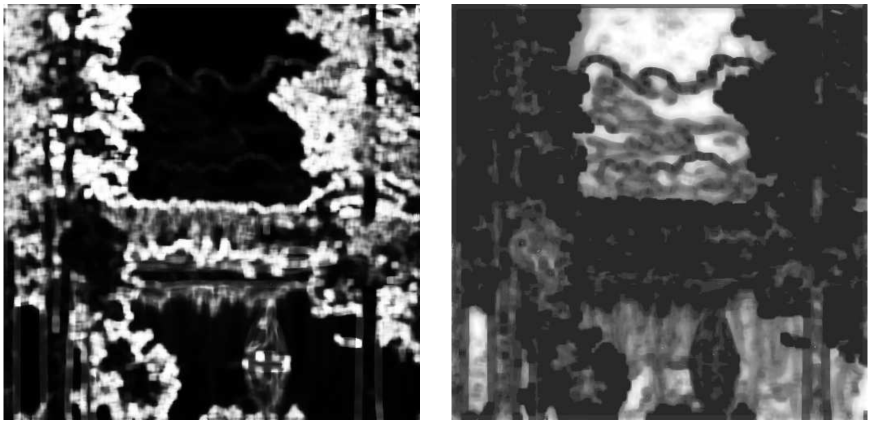
Figure 16: Results on a the noisy 256×256 *mandrilll* image artificially corrupted with different signal-to-ratio levels (WGN).



(a) original 512×512 image

(b) simplified image

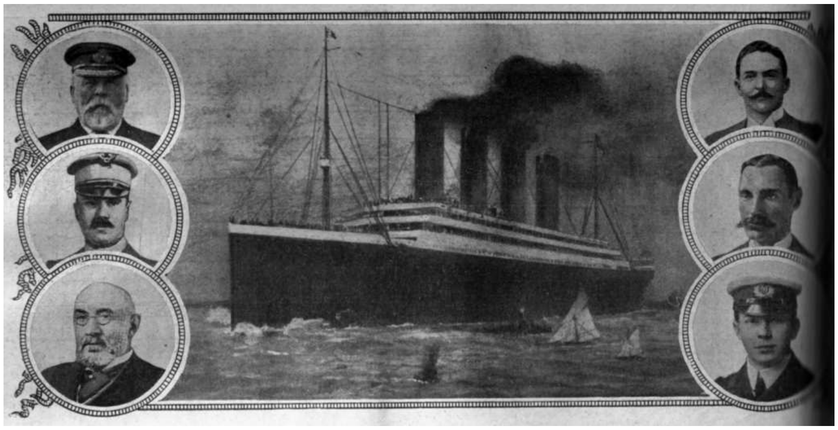
(c) noise component ($\times 2$)



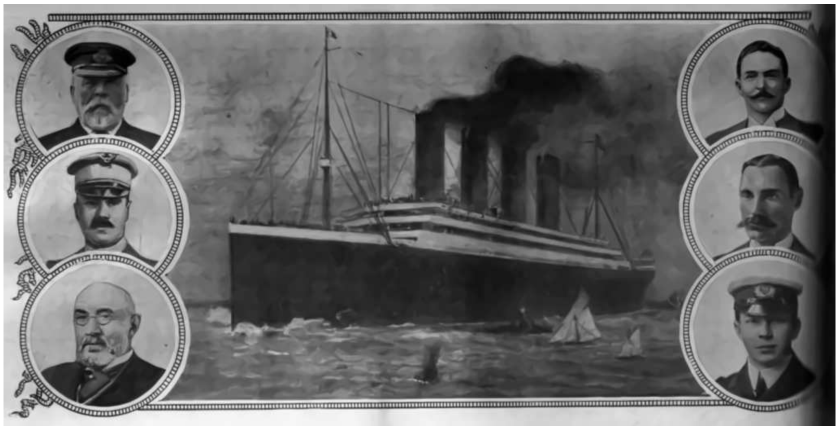
(d) variance of the estimator

(e) $\mathbb{P}\{\widehat{\mathbf{u}}(\mathbf{x}_i)$ occurring in $\widehat{\Delta}\}$

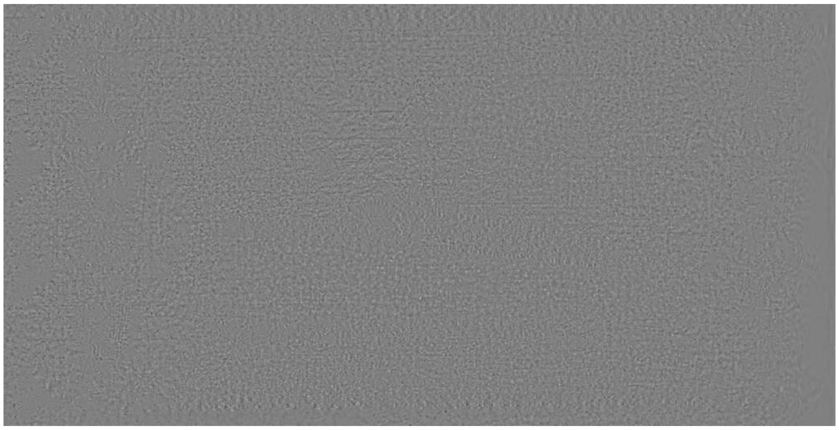
Figure 17: Simplification of the *sailboat* 512×512 image ($\widehat{\sigma} = 7.70, \varrho = 2.59$).



(a) original image

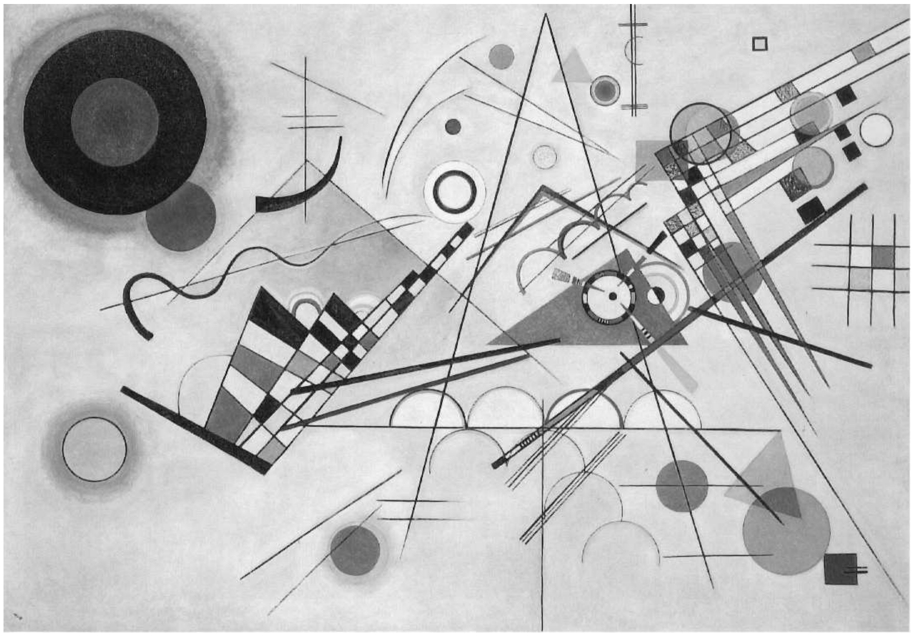


(b) restored image

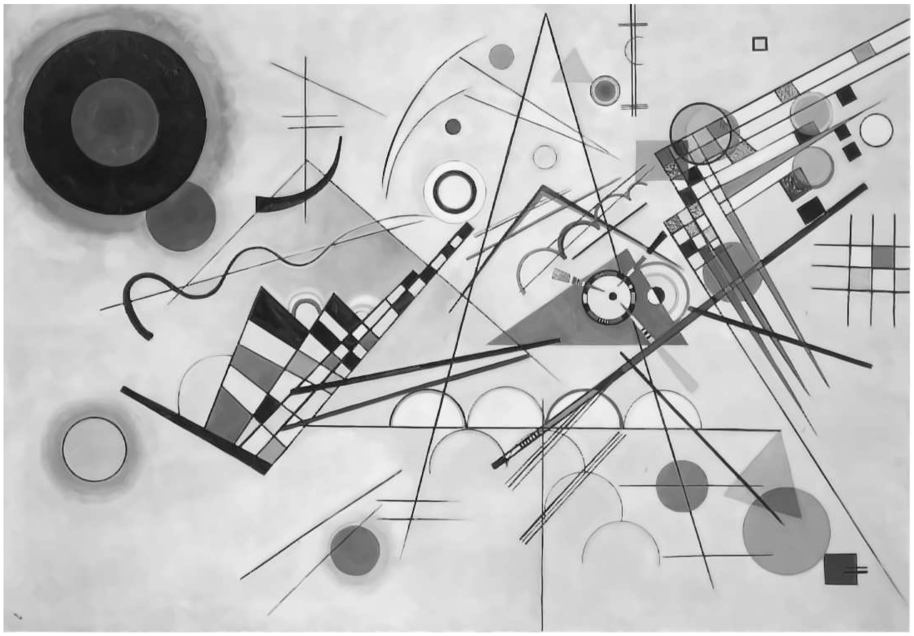


(c) noise component ($\times 2$)

Figure 18: Restoration of a picture.



(a) original image



(b) simplified image

Figure 19: Simplification of a paint.

As previously explained, if we define a set $\Omega(\mathbf{x}_i) = \{\mathbf{x}_j \in \widehat{\Delta}(\mathbf{x}_i) : \mathrm{dist}(\widehat{\mathbf{u}}(\mathbf{x}_i), \widehat{\mathbf{u}}(\mathbf{x}_j)) \leq \lambda_\alpha\}$ containing all occurrences of $\widehat{\mathbf{u}}(\mathbf{x}_i)$ in $\widehat{\Delta}(\mathbf{x}_i)$, then a patch $\widehat{\mathbf{u}}(\mathbf{x}_i)$ is stated as exceptional if

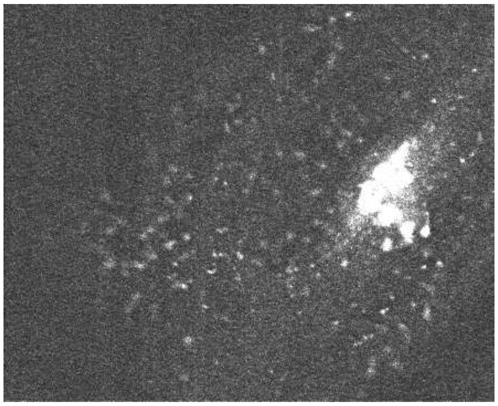
$$\mathbb{P}\{\widehat{\mathbf{u}}(\mathbf{x}_i) \text{ occurring in } \widehat{\Delta}(\mathbf{x}_i)\} \; = \; \frac{\#\Omega(\mathbf{x}_i)}{|\widehat{\Delta}(\mathbf{x}_i)|} \; \leq \; \varepsilon \qquad (26)$$

where $0 \leq \varepsilon \leq 1$ is a detection threshold. We generate a plausible list of exceptional patches from local minima of $\mathbb{P}\{\widehat{\mathbf{u}}(\mathbf{x}_i)$ occurring in $\widehat{\Delta}(\mathbf{x}_i)\}$. Finally clustering on locations of the minima which are lower than ε, produces a set of distinct exceptional patches. This approach does not involve extra computation. In what follows, the positions of interest points are superimposed on the input noisy image. In Fig. 20c, the user-defined parameter ε has been adjusted to 0.005 to return 218 detected objects, most of them corresponding to vesicles/spots. In addition, the reading of observed trajectories can be improved if the exemplar-based restoration method is applied to xt or yt projection images shown in Figs. 21 and 22. In future work, the challenge is to track these detected vesicles with high precision in movies representing several gigabytes of image data and collected and processed automatically to generate information on complex trajectories.

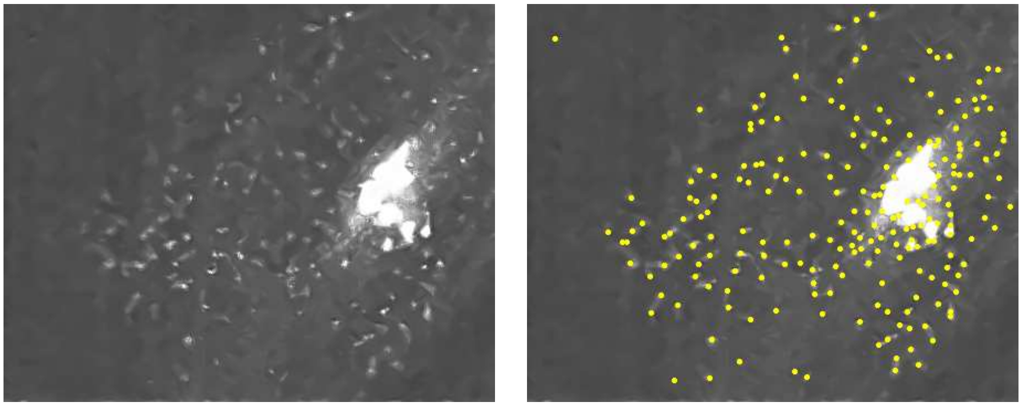
We have also tested the algorithm on 2D and 3D confocal fluorescence microscopy images that contain complex structures. In Fig. 23, a typical 2D image taken from the 3D stack of 80 images depicts glomeruli in the antennal lobes of the moth olfactory system (Fig. 23a). The smoothed 2D image have larger homogeneous areas than the original 2D image (Fig. 23b). The corresponding noise component image is shown in Fig. 23d. Here, this image decomposition is useful to extract regions/volumes of interest. Finally 125 interest points have been detected in the vicinity of point with high curvature, by setting $\varepsilon = 0.01$ and are mainly located on level lines with high curvature. Some of the current applications in biological studies are in neuron research. The 271×238 confocal image depicts neural cells (Fig. 24a). The image is denoised using the set of parameters used in the previous experiments. The denoised 2D image have larger homogeneous areas than the original 2D image (Fig. 24b) and can be more easily segmented. Finally, the same denoising process has been applied to a 2D image showing nuclei in a embryology specimen. In that case, 274“rare” elements ($\varepsilon = 0.05$) are located on the borders of nuclei corresponding to image level line points with high curvatures. These interest points can help to describe the spatial distribution of nuclei in the analyzed specimen. The validation of this methodology and the automatic adjustment of the threshold ε is an on-going research.

8 Summary and conclusions

We have described a novel feature-preserving adaptive restoration algorithm where local image patches and variable window sizes are jointly used. The proposed smoothing scheme provides an alternative method to the anisotropic diffusion, bilateral filtering and energy minimization methods. Our straightforward and unsupervised method yields a significant improvement in image denoising and achieves performances close to the best special-purpose wavelet-based denoising algorithms. We believe this method represents an important step forward for the use of neighborhood design that captures spatial dependencies in images. Unlike previous most exemplar-based methods that use learning algorithms, our method is unsupervised and fully automatic since control parameter are easily calibrated with statistical arguments. Experimental results show demonstrated its potential for large variety of images, including bio-imaging.

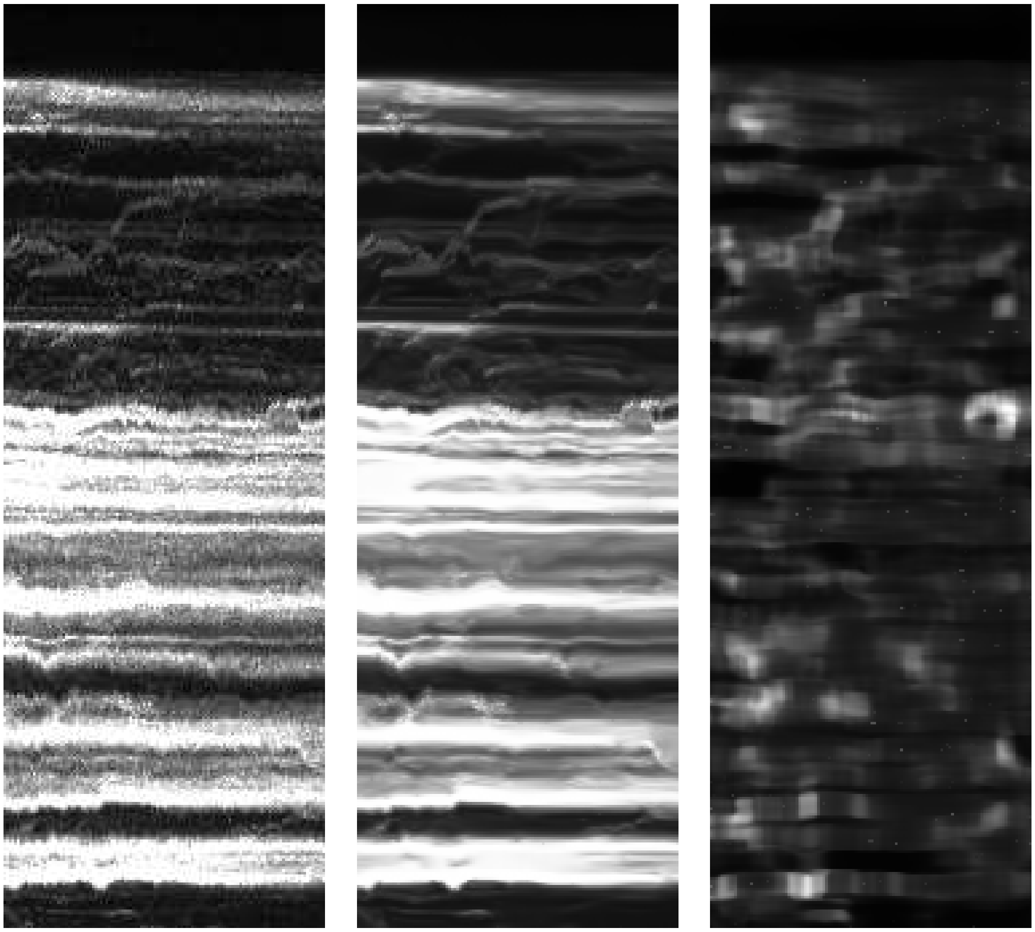


(a) noisy image



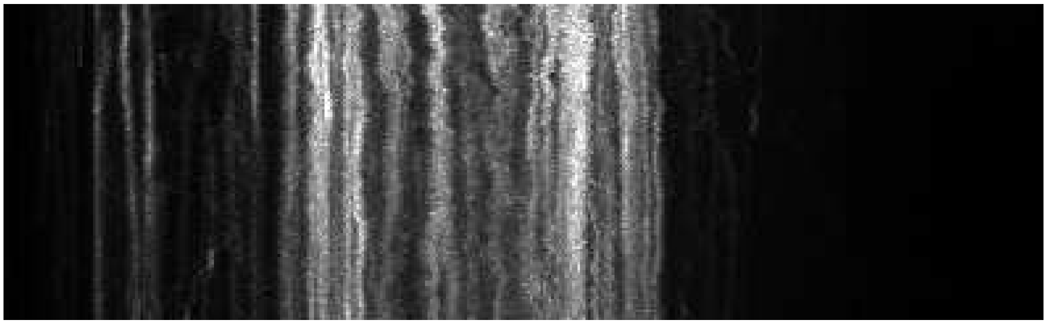
(b) denoised image (c) detection of spots

Figure 20: Results on a 2D image depicting small vesicles of transport (spots).

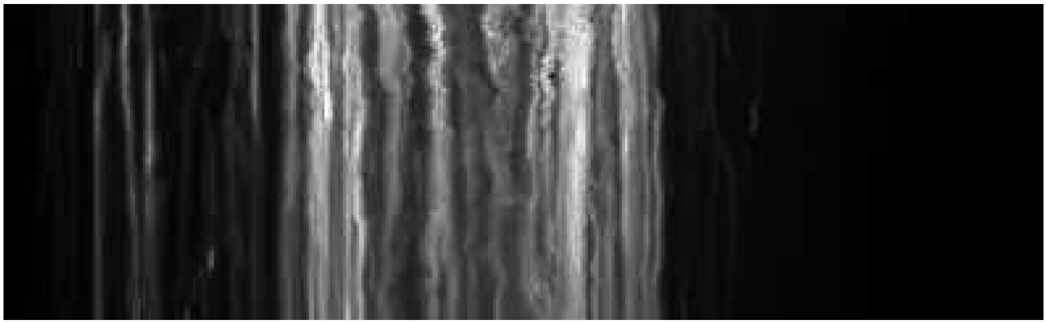


(a) noisy image (b) denoised image (c) variance of the estimator

Figure 21: Results on a 2D image depeciting trajectories of vesicles of transport in a spatio-temporal plane yt.



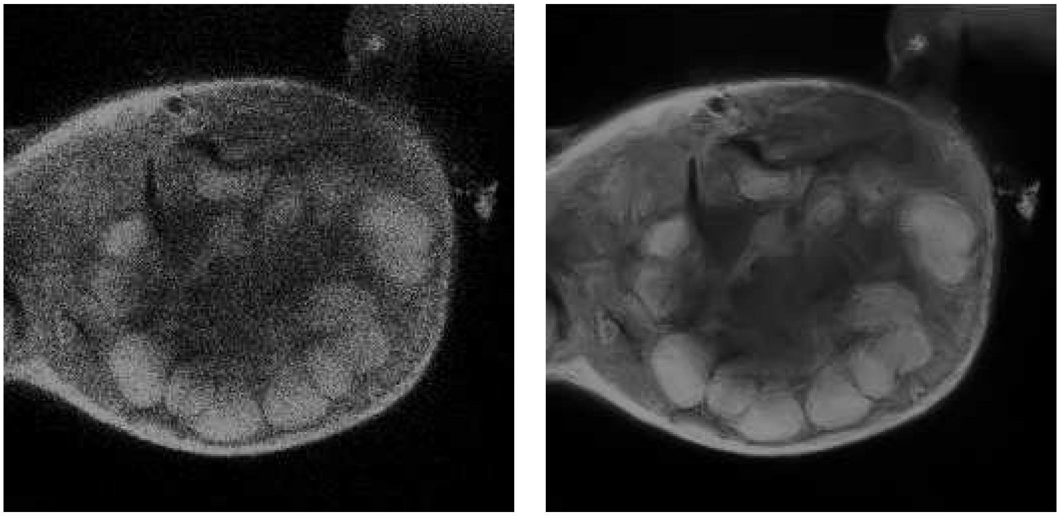
(a) noisy image



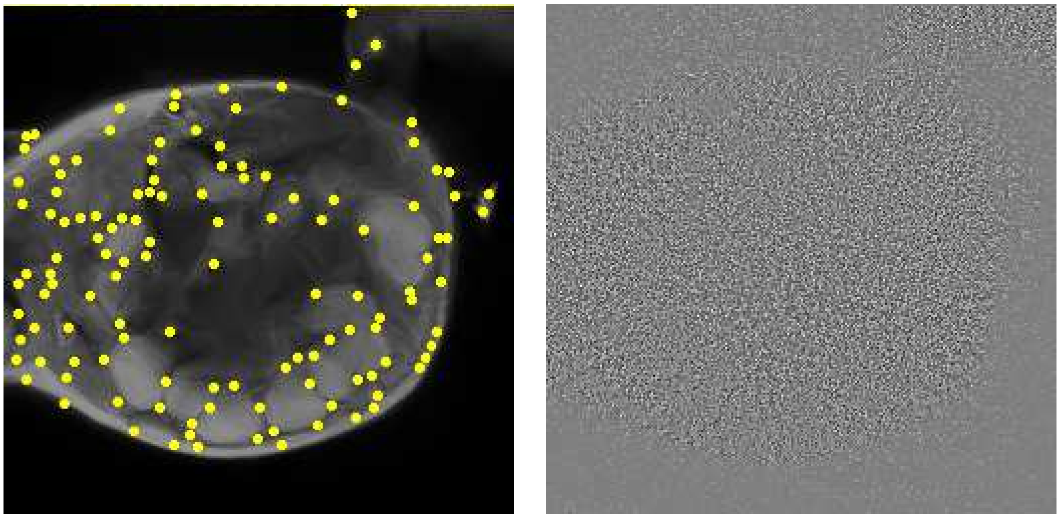
(b) denoised image

(c) variance of the estimator

Figure 22: Results on a 2D image depeciting trajectories of vesicles of transport in a spatio-temporal plane xt.

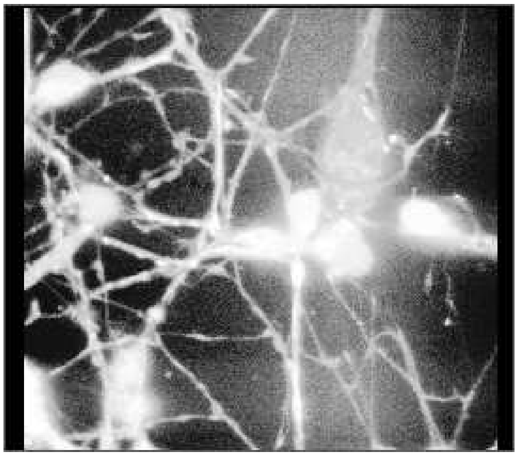


(a) noisy image (b) denoised image

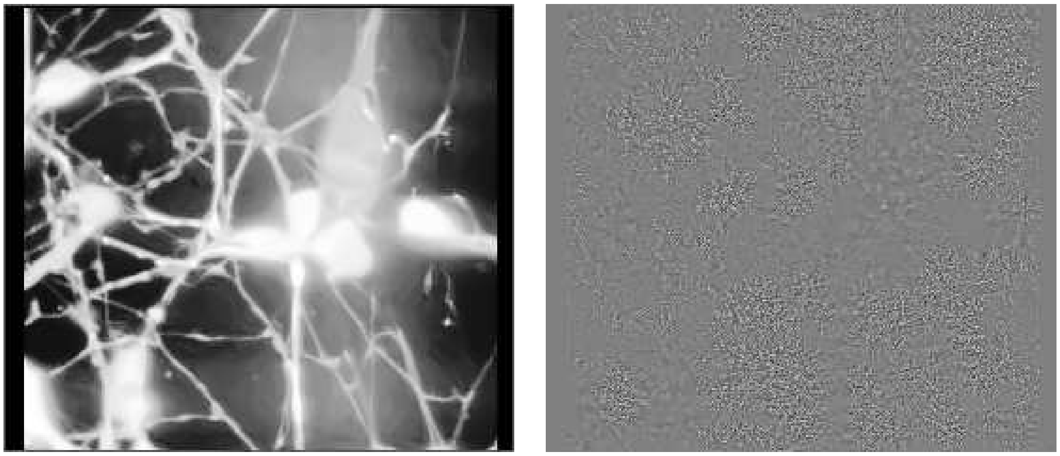


(c) detection of interest points (d) noise component (×2)

Figure 23: Results on a 2D image depicting glomeruli in the antennal lobes of the moth olfactory system.



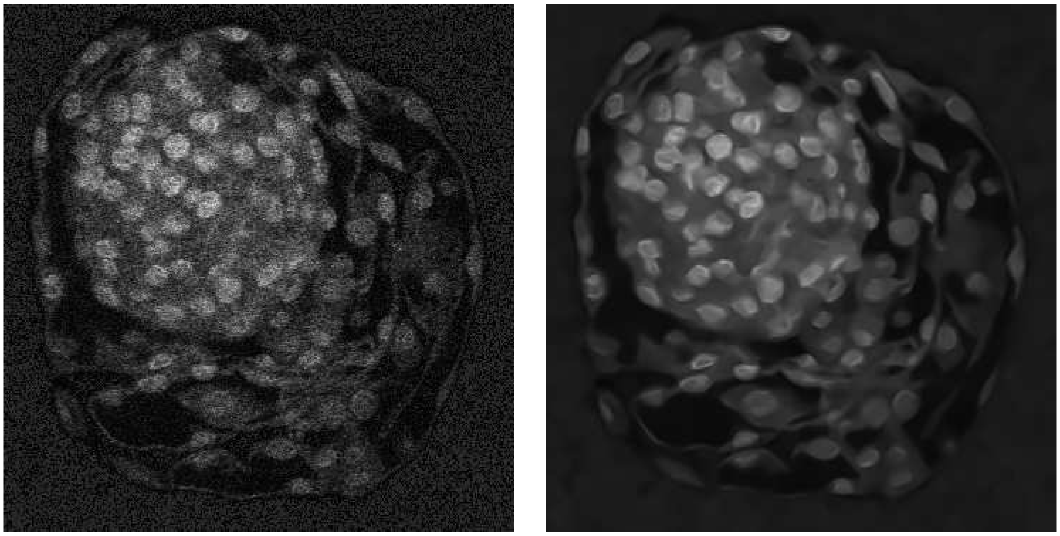
(a) noisy image



(b) denoised image

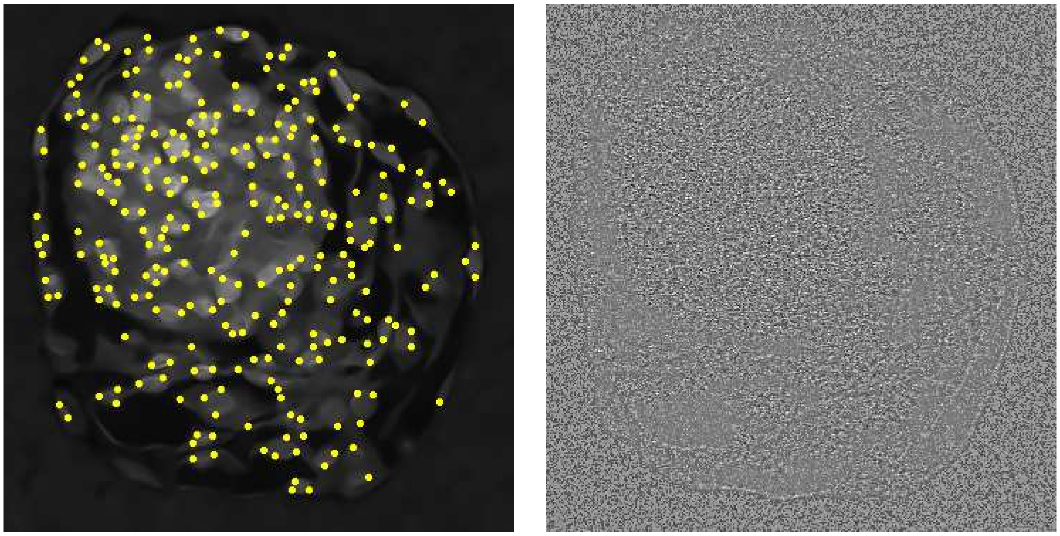
(c) noise component ($\times 2$)

Figure 24: Results on a 2D image depicting neural cells.



(a) noisy image

(b) denoised image



(c) detection of interest points

(d) noise component (×2)

Figure 25: Results on a 2D image depicting nuclei in a embryology spcecimen.

References

[1] J.F. Aujol, G. Aubert, L. Blanc-Feraud, A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *J. Math. Imaging Vis.*, 22(1), Janvier 2005.

[2] V. Aurich, J. Weule. Nonlinear Gaussian filters performing edge preserving diffusion. In *Proc. of 17-th DAGM Symposium*, Bielefeld, pp. 538-545, 1995.

[3] D. Barash. A fundamental relationship between bilateral filtering, adaptive smoothing and the nonlinear diffusion equation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 24(6): 844-847, 2002.

[4] D. Barash, D. Comaniciu. A Common framework for nonlinear diffusion, adaptive s-moothing, bilateral filtering and mean-shift. *Image and Video Computing*, 22(1): 73-81, 2004

[5] A. Ben Hamza, H. Krim. A variational approach to maximum a posteriori estimation for image denoising. In *Proc. EMMCVPR'01*, LNCS 2134, pp. 19-34, Sophia-Antipolis, France, 2001.

[6] A. Blake, A. Zisserman. *Visual Reconstruction*, MIT Press, 1987.

[7] M.J. Black, G. Sapiro, D.H. Marimont, D. Heeger. Robust anisotropic diffusion. *IEEE Trans. Image Process.*, 7(3): 421-432, 1998.

[8] M.J. Black, G. Sapiro. Edges as outliers: Anisotropic smoothing using local image statistics. In *Proc. Scale-Space'99*, LNCS 1682, pp. 259-270, Kerkyra, Greece, 1999.

[9] Y. Boykov, O. Veksler, R. Zabih. A variable window approach to early vision. *IEEE Trans. Patt. Anal. Mach. Intel.*, 20(12): 1283-1294, 1998.

[10] T. Brox, J. Weickert. A TV flow based local scale measure for texture discrimination. In *Proc. Eur. Conf. Comp. Vis. (ECCV'04)*, pp. 578-590, Vol 2, Prague, Czech Republic, 2004.

[11] J. Boulanger, C. Kervrann, P. Bouthemy. Adaptive spatio-temporal restoration for 4D fluorescence microscopic imaging. In *Proc. Int. Conf. Medical Image Comp. and Computer Assisted Intervention (MICCAI'05)*, Palm Springs, California, 2005.

[12] A. Buades, B. Coll, J.-M. Morel. A non-local algorithm for image denoising. *Preprint*, CMLA, Cachan, France, 2004. ((to appear in SIAM Multiscale Modeling and Simulation)

[13] A. Buades, B. Coll, J.-M. Morel. Image denoising by non-local averaging. In *Proc. Comp. Vis. Patt. Recogn. (CVPR'05)*, San Diego, Ca, USA, 2005.

[14] F. Catte, P.-L. Lions, J.-M. Morel, T. Coll. Image selective smoothing and edge-detection by nonlinear diffusion. *SIAM J. Numerical Analysis*, 29(1): 182-193, 1992.

[15] C. Chefd'hotel, D. Tschumperlé, R. Deriche, O. Faugeras. Regularizing flows for constrained matrix-valued images. *J. Math. Imag. Vis.*, 20(1-2): 147-162, 2004.

[16] I. Cheng. Mean-shift, mode seeking, and clustering. *IEEE Trans. Patt. Anal. Mach. Intel.*, 17(8): 790-7999, 1995.

[17] C.K. Chu, K. Glad, F. Godtliebsen, J.S. Marron. Edge-preserving smoothers for image processing. *J. Am. Stat. Ass.*, 93(442): 526-555., 1998.

[18] D. Comaniciu, V. Ramesh, P. Meer. The variable bandwidth mean-shift and data-driven scale selection. In *Proc. Int Conf. Comp. Vis. (ICCV'01)*, pp. 438-445, Vol 1, Vancouver, Canada, 2001.

[19] D. Comaniciu, P. Meer. Mean-shift: a robust approach toward feature space analysis. *IEEE Trans. Patt. Anal. Mach. Intel.*, 24(5): 603-619, 2002.

[20] A. Criminisi, P. Pérez, K. Toyama. Region filling and object removal by exemplar-based inpainting. *IEEE Trans. Image Process.*, 13(9): 1200-1212, 2004.

[21] I. Csiszár. Why least squares and maximum entropy ? An axiomatic approach to inference for linear inverse problems. *Ann. Statist.*, 19: 2032-2066, 1991.

[22] M.N. Do, M. Vetterli. The Contourlet transform: an efficient directional multiresolution image representation. *IEEE Trans. Image Process.*, 2004. in press.

[23] D.L. Donoho, I.M. Johnston. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81: 425-455, 1994.

[24] D.L. Donoho, I.M. Johnston. Denoising by soft-thresholding. *IEEE Transactions on Information Theory*, 41: 613-627, 1995.

[25] A. Efros, T. Leung. Texture synthesis by non-parametric sampling. In *Proc. ICCV'99*, pp. 1033-1038, Kerkyra, Greece, 1999.

[26] M. Elad. On the bilateral filter and ways to improve it. *IEEE Trans. Image Process.*, 11(10): 1141-1151, 2002.

[27] B. Fischl, E.L. Schwartz. Adaptive nonlocal filtering: a fast alternative to anisotropic diffusion for image enhancement. *IEEE Trans. Patt. Anal. Mach. Intell.*, 21(1): 42-48, 1999.

[28] A. Fitzgibbon, Y. Wexler, A. Zisserman. Image-based rendering using image-based priors. In *Proc. ICCV'03*, Nice, France, 2003.

[29] W. T. Freeman, E. C. Pasztor, O. T. Carmichael. Learning low-level vision. *Int. J. Comp. Vis.*, 40(1): 25-47, 2000.

[30] T. Gasser, L. Sroka, C. Jennen Steinmetz. Residual variance and residual pattern in nonlinear regression. *Biometrika*, 73: 625-633, 1986.

[31] D. Geman, D. Geman. Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intell.*, 6(???): 721-741, 1984.

[32] M. Ghazel, G.H. Freeman, E.R. Vrscay. Fractal image denoising. *IEEE Trans. Image Process.*, 12(12): 1560-1578, 2003.

[33] G. Gilboa, N. Sochen, Y.Y. Zeevi. Texture preserving variational denoising using an adaptive fidelity term. In *Proc. VLSM'03*, Nice, France, 2003.

[34] G. Gilboa, N. Sochen, Y.Y. Zeevi. Estimation of the optimal variational parameter via SNR analysis. In *Proc. Int. Conf. on Scale-Space and PDE methods in Computer Vision*, Hofgeismar, Germany, April 2005.

[35] F. Godtliebsen, E. Spjotvoll, J.S. Marron. A nonlinear Gaussian filter applied to images with discontinuities. *J. Nonparametric Statistics*, 8: 21-43, 1997.

[36] A. Goldenshluger, A. Nemirovsky. On spatial adaptive estimation of nonparametric regression. *Math. Meth. of Statist.*, 6(2): 135-170, 1997.

[37] G. Gomez, J.L. Marroquin, L.E. Sucar. Probabilistic estimation of local scale. In *Proc. Int. Conf. Patt. Recog. (ICPR'00)*, vol. 3, pp. 798-801, Barcelona, Spain, 2000.

[38] M.L. Green. Statistics of images, the TV algorithm of Rudin-Osher-Fatemi for image denoising and an improved denoising algorithm. *CAM Report*, 02(55), 2002.

[39] W. Hardle, O. Linton. Applied nonparametric methods. *In R.F. Engle, D.L. McFadden ed., Handbook of Econometrics*, Vol. IV, Amsterdam, North Holland, pp. 2295-2381, 1994.

[40] C. Harris, M. Stephens. A combined corner and edge detector. In *Alvey Vision Conf.*, pp. 147-151, 1988.

[41] N. Jojic, B. Frey, A. Kannan. Epitomic analysis of appearance and shape. *Proc. Int Conf. Comp. Vis. (ICCV'03)*, pp. 34-41, Vol. 1, Nice, France, 2003.

[42] A. Juditsky. Wavelet estimators: adapting to unknown smoothness. *Math. Meth. S-tatist.*, 1:1-20, 1997.

[43] V. Katkovnik, K. Egiazarian, J Astola. Adaptive window size image denoising based on intersection of confidence intervals (ICI) rule. *J. Math. Imag. Vis.*, 16(3): 223-235, 2002.

[44] C. Kervrann. An adaptive window approach for image smoothing and structures p-reserving. *Proc. Eur. Conf. Comp. Vis. (ECCV'04)*, pp. 132-144, Vol. 3, Prague, Czech Republic, 2004.

[45] J.S. Lee. Digital image smoothing and the sigma filter. *Comp. Vis. Graph. Image Process.*, 24: 255-269, 1983.

[46] A.B. Lee, K.S. Pedersen, D. Mumford. The nonlinear statistics of high-contrast patches in natural images. *Int. J. Comp. Vis.*, 54(1-3): 83-103, 2003.

[47] E. Le Pennec, S. Mallat. Sparse geometric image representation with Bandelets. *IEEE Trans. on Image Process.*, April 2005.

[48] O. Lepskii. On a problem of adaptive estimation on white Gaussian noise. *Theory Probab. Appl.*, 35: 454-466, 1990.

[49] O. Lepskii. Asymptotically minimax adaptive estimation 1: uppers bounds. *SIAM J. Theory of Prob. and Appl.*, 36(4): 654-659, 1991.

[50] O. V. Lepski, E. Mammen, V. G. Spokoiny. Lepskii. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.*, 25(3): 929-947, 1997.

[51] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *Int. J. Comp. Vis.*, 30(2): 117-154, 1998.

[52] M. Maurizot, P. Bouthemy, B. Delyon, A. Juditski, J.-M. Odobez. Determination of singular points in 2D deformable flow fields. In *IEEE Int. Conf. Image Processing*, Washington DC, 1995.

[53] Y. Meyer. Oscillating patterns in image processing and nonlinear evolution equations. *University Lecture Series*, 22, AMS 2002.

[54] P. Mrazek. Selection of optimal stopping time for nonlinear diffusion filtering. *Int. J. Comp. Vis.*, 52(2/3): 189-203, 2003.

[55] P. Mrazek, J. Weickert, A. Bruhn. On robust estimation and smoothing with spatial and tonal kernels. *Preprint no 51*, University of Bremen, Germany, june 2004.

[56] D. Mumford, J.Shah. Optimal approximations by piecewise smooth functions and variational problems. *Comm. Pure and Appl. Math.*, 42(5): 577-685, 1989.

[57] M. Nitzberg, T. Shiota. Nonlinear image filtering with edge and corner enhancement. *IEEE Trans. Patt. Anal. Mach. Intel.*, 14(8): 826-833, 1992.

[58] S. Osher, A. Solé, L. Vese. Image decomposition and restoration using total variation minimization and the H^{-1} norm. *SIAM J. Multiscale Model. Simul.*, 1(3): 349-370, 2003.

[59] P. Perona, J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Trans. Patt. Anal. Mach. Intell.*, 12(7): 629-239, 1990.

[60] A. Pizurica, W. Philips. Estimating probability of presence of a signal of interest in multiresolution single- and multiband image denoising. *IEEE Trans. Image Process.* , 2004.

[61] J. Polzehl, V. Spokoiny. Adaptive weights smoothing with application to image restoration. *J. R. Stat. Soc. B.*, 62(2): 335-354, 2000.

[62] J. Portilla, V. Strela, M. Wainwright, E. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Process.*, 12(11): 1338-1351, 2003.

[63] S. Roth, M.J. Black. Fields of experts: a framework for learning image priors with applications. *Proc. IEEE Conf. on Comp. Vis. Patt. Recog. (CVPR'05)*, San Diego, USA, 2005.

[64] L. Rudin, S. Osher, E. Fatemi. Nonlinear Total Variation based noise removal algorithms. *Physica D*, 60: 259-268, 1992.

[65] P. Saint-Marc, J.S. Chen, G. Médioni. Adaptive smoothing: a general tool for early vision. *IEEE Trans. Patt. Anal. Mach. Intel.*, 13(6): 514-529, 1991.

[66] C. Schmid, R. Mohr, C. Bauckhage. Evaluation of interest point detectors. *Int. J. Comp. Vis.*, 37(2): 151-172, 2000.

[67] D. W. Scott. *Multivariate Density Estimation*. Wiley, 1992.

[68] M. Singh, N. Ahuja. Regression based bandwidth selection for segmentation using Parzen windows. In *Proc. Int Conf. Comp. Vis. (ICCV'03)*, pp. 2-9, Vol. 1, Nice, France, 2003.

[69] S.M. Smith, M. Brady. SUSAN - a new approach to low-level image processing. *Int. J. Comp. Vis.*, 23(1): 45-78, 1997.

[70] V. G. Spokoiny. Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Ann. Statist.*, 26(4): 141-170, 1998.

[71] L. Stankovic. Performance analysis of the adaptive algorithm for bias-to-variance tradeoff. *IEEE Trans. Signal Process.*, 52(5): 1228-1234, 2004.

[72] N. Sochen, R. Kimmel, A. M. Bruckstein. Diffusions and Confusions in Signal and Image Processing. *J. Math. Imag. Vis.*, 14(3): 237-244, 2001.

[73] A. Spira, R. Kimmel, N. Sochen. Efficient Beltrami flow using a short-time kernel. In *Proc. Int. Conf. Scale-Space Methods in Computer Vision*, pp. 551-522, Isle of Skye, Scotland, 2003.

[74] J.L. Starck, E. Candes, D.L. Donoho. The Curvelet transform for image denoising. *IEEE Trans. on Image Process.* , 11(6): pp 670-684, 2002.

[75] C. Tomasi, R. Manduchi. Bilateral filtering for gray and color images. In *Proc. Int Conf. Comp. Vis. (ICCV'98)*, pp. 839-846, Bombay, India, 1998.

[76] R. van den Boomgaard, J. van de Weijer. On the equivalence of local-mode finding, robust estimation and mean-shift analysis as used in early vision tasks. In *Proc. Int. Conf. Patt. Recogn. (ICPR'02)*, vol. III, pp. 927-930, Quebec City, Canada, 2002.

[77] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.

[78] J. Weickert. Coherence-enhancing diffusion filtering. *Int. J. Comp. Vis.*, 31(2/3): 111-127, 1999.

[79] J. van de Weijer, R. van den Boomgaard. Local mode filtering. In *Proc. Comp. Vis. Patt. Recogn. (CVPR'01)*, vol. II, pp. 428-433, Kauai, Hawaii, USA, 2001.

[80] G.Z. Yang, P. Burger, D.N. Firmin, S.R. Underwood. Structure adaptive anisotropic image filtering. *Image Vis. Comput.*, 14: 135-145, 1996.

[81] L.P. Yaroslavsky, M. Eden *Fundamentals of digital optics*, Birkhaüser, Boston, 1996.

[82] S.C. Zhu, Y. Wu, D. Mumford. Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. *Int. J. Comp. Vis.* 27(2): 107-126, 1998.

A Appendix

A.1 Proof of the inequality: $\mathrm{Var}(\widehat{u}_{i,n} - \widehat{u}_{i,n'}) \leq \widehat{\upsilon}^2_{i,n'},\ \forall\ 1 \leq n' < n.$

We have:

$$\widehat{u}_{i,n} - \widehat{u}_{i,n'} \sim \mathcal{N}(0, \mathrm{Var}(\widehat{u}_{i,n} - \widehat{u}_{i,n'}))$$

since both biases are negligible ($\mathbb{E}\,[\widehat{u}_{i,n} - \widehat{u}_{i,n'}] \approx 0$) and then

$$\mathrm{Var}(\widehat{u}_{i,n} - \widehat{u}_{i,n'})) = \mathbb{E}\left[|\widehat{u}_{i,n} - \widehat{u}_{i,n'}|^2\right].$$

We recall that $Y_i = u(\mathbf{x}_i) + \epsilon_i$ and

$$\widehat{u}_{i,n} = \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n} Y_j \quad \text{and} \quad \widehat{u}_{i,n'} = \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'} Y_j.$$

Let us assume $u(\mathbf{x}_i)$ is constant in $\Delta_{i,n'} \subset \Delta_{i,n} \subset \widehat{\Delta}(\mathbf{x}_i)$, then we write

$$\begin{aligned}
&\mathrm{Var}(\widehat{u}_{i,n} - \widehat{u}_{i,n'})) \\
&= \mathbb{E}\left[\left(\sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n}\,\epsilon_j + \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n}\,u(\mathbf{x}_i) - \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'}\,\epsilon_j - \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'}\,u(\mathbf{x}_i)\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n}\,\epsilon_j + u(\mathbf{x}_i) \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n} - \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'}\,\epsilon_j - u(\mathbf{x}_i) \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'}\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n}\,\epsilon_j - \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'}\,\epsilon_j\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n}\,\epsilon_j - \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'}\,\epsilon_j\right)\left(\sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n}\,\epsilon_j - \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'}\,\epsilon_j\right)\right] \\
&= \widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi^2_{i\sim j,n} + \widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi^2_{i\sim j,n'} - 2\,\mathbb{E}\left[\sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n}\,\epsilon_j \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'}\,\epsilon_j\right] \\
&= \widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi^2_{i\sim j,n} + \widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi^2_{i\sim j,n'} - 2\widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'}\pi_{i\sim j,n} \\
&= \widehat{\upsilon}^2_{i,n} + \widehat{\upsilon}^2_{i,n'} - 2\widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'}\pi_{i\sim j,n}\,.
\end{aligned}$$

In addition, if $\pi_{i\sim j,n'} = 0$ when $\mathbf{x}_j \notin \Delta_{i,n'}$, it follows that

$$\sum_{\mathbf{x}_j \in \Delta_{i,n'}} \pi_{i\sim j,n'} \pi_{i\sim j,n} \;\geq\; \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n'} \pi_{i\sim j,n} \,.$$

By definition, $\pi_{i\sim j,n'} \geq \pi_{i\sim j,n}$ for $\mathbf{x}_j \in (\Delta_{i,n} \cap \Delta_{i,n'})$, hence

$$\widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi_{i\sim j,n'} \pi_{i\sim j,n} \;\geq\; \widehat{\sigma}^2 \sum_{\mathbf{x}_j \in \Delta_{i,n}} \pi^2_{i\sim j,n} \;=\; \widehat{\upsilon}_{i,n}.$$

Finally, we obtain the inequality:

$$\mathrm{Var}(\widehat{u}_{i,n} - \widehat{u}_{i,n'})) \;\leq\; \widehat{\upsilon}^2_{i,n'} + \widehat{\upsilon}^2_{i,n} - 2\widehat{\upsilon}^2_{i,n} \;\leq\; \widehat{\upsilon}^2_{i,n'} - \widehat{\upsilon}^2_{i,n} \;\leq\; \widehat{\upsilon}^2_{i,n'}.$$

□

A.2 Proof of Proposition 1.

Let the condition of the theorem be satisfied, that is $n^\star(\mathbf{x}_i) \leq \widehat{n}(\mathbf{x}_i)$, then, from the inequality (20), we have

$$|u^\star(\mathbf{x}_i) - \widehat{u}(\mathbf{x}_i)| \;\leq\; (2\gamma + \varkappa)\, \upsilon^\star(\mathbf{x}_i).$$

By taking the expectation of this expression, we obtain:

$$\begin{aligned}
\left[\mathbb{E}_{n^\star(\mathbf{x}_i) \leq \widehat{n}(\mathbf{x}_i)} |u^\star(\mathbf{x}_i) - \widehat{u}(\mathbf{x}_i)|^2\right]^{1/2} &\leq (2\gamma + \varkappa)\, \upsilon^\star_{(}\mathbf{x}_i) \\
&\leq \frac{2\gamma + \varkappa}{\sqrt{1+\gamma^2}} \left[\mathbb{E}|u^\star(\mathbf{x}_i) - u(\mathbf{x}_i)|^2\right]^{1/2}.
\end{aligned}$$

Finally, by applying the triangular inequality, we get

$$\begin{aligned}
\left[\mathbb{E}_{n^\star(\mathbf{x}_i) \leq \widehat{n}(\mathbf{x}_i)} |u^\star(\mathbf{x}_i) - \widehat{u}(\mathbf{x}_i)|^2\right]^{1/2} &\leq \left[\mathbb{E}_{n^\star(\mathbf{x}_i) \leq \widehat{n}(\mathbf{x}_i)} |u^\star(\mathbf{x}_i) - \widehat{u}(\mathbf{x}_i)|^2\right]^{1/2} \\
&\quad + \left[\mathbb{E}|u^\star(\mathbf{x}_i) - u(\mathbf{x}_i)|^2\right]^{1/2} \\
&\leq \frac{2\gamma + \varkappa}{\sqrt{1+\gamma^2}} \left[\mathbb{E}|u^\star(\mathbf{x}_i) - u(\mathbf{x}_i)|^2\right]^{1/2} \\
&\quad + \left[\mathbb{E}|u^\star(\mathbf{x}_i) - u(\mathbf{x}_i)|^2\right]^{1/2} \\
&\leq \left[\frac{2\gamma + \varkappa}{\sqrt{1+\gamma^2}} + 1\right] \left[\mathbb{E}|u^\star(\mathbf{x}_i) - u(\mathbf{x}_i)|^2\right]^{1/2},
\end{aligned}$$

what yields the desired bound.

□

A.3 Proof of Proposition 2.

If $X = \dfrac{\widehat{u}_{i,n} - \widehat{u}_{i,n'}}{\widehat{v}_{i,n'}}$ is $\mathcal{N}(0,1)$, then

$$\mathbb{E}\exp(\varrho X) = \exp\left(\frac{\varrho^2}{2}\right).$$

From this inequality and the exponential Chebychev's inequality stated as follows:

$$\mathbb{P}(X \geq a) \ \leq \ \frac{\mathbb{E}[g(X)]}{g(a)}$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ is a positive monotonous function and $a \in \mathbb{R}$, we have

$$\mathbb{P}\{(\widehat{u}_{i,n} - \widehat{u}_{i,n'}) > \varrho\widehat{v}_{i,n'})\} \leq \frac{\mathbb{E}\exp\left(\frac{\varrho(\widehat{u}_{i,n} - \widehat{u}_{i,n'})}{\widehat{v}_{i,n'}}\right)}{\exp\varrho^2}.$$

by taking $g(x) = \exp(\varrho x)$ and $a = \varrho$. Hence

$$\mathbb{P}\{(\widehat{u}_{i,n} - \widehat{u}_{i,n'}) > \varrho\widehat{v}_{i,n'}\} \leq \exp\left(-\frac{\varrho^2}{2}\right)$$

and the following result comes from the symmetry of the normal distribution.

$$\mathbb{P}\{|\widehat{u}_{i,n} - \widehat{u}_{i,n'}| > \varrho\widehat{v}_{i,n'}\} \leq 2\exp\left(-\frac{\varrho^2}{2}\right).$$

To prove Proposition 2., we note also that

$$\begin{aligned} \{\widehat{n}(\mathbf{x}_i) = n\} &= \{\exists n' \in \{1, \ldots, n-1\} : |\widehat{u}_{i,n} - \widehat{u}_{i,n'}| > \varrho\widehat{v}_{i,n'}\} \\ &\subseteq \bigcup_{n'<n} \{|\widehat{u}_{i,n} - \widehat{u}_{i,n'}| > \varrho\widehat{v}_{i,n'}\}. \end{aligned}$$

Using this definition, we get

$$\begin{aligned} \mathbb{P}\{\widehat{n}(\mathbf{x}_i) = n\} &\leq \sum_{n'<n} \mathbb{P}\{|\widehat{u}_{i,n} - \widehat{u}_{i,n'}| > \varrho\widehat{v}_{i,n'}\} \\ &\leq \sum_{n'<n} 2\exp\left(-\frac{\varrho^2}{2}\right) \ \leq \ 2N_\Delta \exp\left(-\frac{\varrho^2}{2}\right) \end{aligned}$$

This proves the assertion.

□