



**HAL**  
open science

# Non-Local Methods with Shape-Adaptive Patches (NLM-SAP)

Charles-Alban Deledalle, Vincent Duval, Joseph Salmon

► **To cite this version:**

Charles-Alban Deledalle, Vincent Duval, Joseph Salmon. Non-Local Methods with Shape-Adaptive Patches (NLM-SAP). *Journal of Mathematical Imaging and Vision*, 2012, 43 (2), pp.103-120. 10.1007/s10851-011-0294-y . hal-00536723

**HAL Id: hal-00536723**

**<https://hal.science/hal-00536723>**

Submitted on 16 Nov 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Non-Local Methods with Shape-Adaptive Patches (NLM-SAP)

Charles-Alban Deledalle · Vincent Duval · Joseph Salmon

November 16, 2010

**Abstract** We propose in this paper an extension of the Non-Local Means (NL-Means) denoising algorithm. The idea is to replace the usual square patches used to compare pixel neighborhoods with various shapes that can take advantage of the local geometry of the image. We provide a fast algorithm to compute the NL-Means with arbitrary shapes thanks to the fast Fourier transform. We then consider local combinations of the estimators associated with various shapes by using Stein's Unbiased Risk Estimate (SURE). Experimental results show that this algorithm improve the standard NL-Means performance and is close to *state-of-the-art* methods, both in terms of visual quality and numerical results. Moreover, common visual artifacts usually observed by denoising with NL-Means are reduced or suppressed thanks to our approach.

**Keywords** Image denoising · non-local means · spatial adaptivity · aggregation · risk estimation · SURE

## 1 Introduction

As simple as it may seem, the classical problem of image denoising in the presence of Additive White Gaussian

Noise (AWGN) has drawn a lot of efforts from the image processing community. A wide variety of strategies were proposed during the last decades, from partial differential equations (PDE) to transform-domain methods such as wavelets, approximation theory or stochastic analysis.

In fact, a first major difficulty in image denoising is to handle efficiently regular parts while preventing edges from being blurred. To this end, diffusion methods [35] and variational methods [40] have shown their effectiveness. The latter approaches introduce a prior model of the noise-free image and search for a trade-off between data fitting and regularization. It is crucial for these techniques to define a suitable prior that guarantees both the smoothness of the denoised image and the preservation of its structure. Total variation [40] is an example of a prior that enforces smoothness while preserving edges. However, such priors tend to bias the denoised image [48], especially in texture area or when high noise levels are considered.

Donoho and Johnstone [14] have shown that spatial adaptive methods can be obtained by soft-thresholding wavelet decompositions. Wavelet shrinkage can be interpreted from a Bayesian perspective. It results from an  $\ell_1$  sparsity-promoting prior: the regularity of the signal is enforced by using sparsifying priors on the coefficients in the transform domain. Indeed, wavelet and cosine bases are well known to be able to capture most of a signal/image with very few coefficients, a good property for compression. For denoising applications, such orthogonal transforms lead to a separation of signal and noise. Noise can then be strongly suppressed by canceling the least significant coefficients. These types of methods culminated with the BLS-GSM algorithm [38]. Other improvements of wavelet oriented methods have been proposed such as Shape Adaptive Domains

---

Charles-Alban Deledalle  
Institut Telecom ; Telecom ParisTech ; CNRS LTCI  
46, rue Barrault 75634 Paris cedex 13, FRANCE  
E-mail: charles-alban.deledalle@telecom-paristech.fr

Vincent Duval  
Institut Telecom ; Telecom ParisTech ; CNRS LTCI  
46, rue Barrault 75634 Paris cedex 13, FRANCE  
E-mail: vincent.duval@telecom-paristech.fr

Joseph Salmon  
Université Paris 7 - Diderot ; Laboratoire de Probabilité et  
Modèles Aléatoires ; CNRS-UMR 7599  
175 rue du Chevaleret 75013 Paris, FRANCE  
E-mail: salmon@math.jussieu.fr

[17] and over-complete [45] or geometric [25] decompositions.

Spatial adaptivity can also be reached by considering adaptive neighbor filters, as the well known Sigma Filter [26], Yaroslavsky Filter [56] and Bilateral Filter [49]. Moreover, applications of Lepski’s method [27] (known in signal processing as the ICI rule, Intersection of Confidence Intervals) to kernel smoothing are also known to perform well for denoising. Their theoretical performance is as good as the best wavelet shrinking methods: they reach adaptive rate in the minimax sense [19] for regular functions. In practice, Lepski’s rule has successfully been applied to image denoising with Adaptive Weights Smoothing [36] or Shape Adaptive Windows [22] for instance.

Though very efficient at dealing with edges, such methods cannot proceed efficiently in textured regions. To overcome this drawback, many authors have proposed to work with small sub-images, called patches, to take into account the redundancy in natural images, especially in textured parts. The interest of using patches lies in their robustness to noise. The NL-Means (Non-Local Means) [5] and UINTA [2] algorithms are typical examples of this approach, as is their extension using Lepski’s method [23]. Those algorithms consist in averaging similar pixels by measuring their similarity through patches. Iterative versions have also been proposed [4, 18, 12] as well as mixed methods using Total Variation [30].

Another approach proposed by Aharon *et al.* [1] and extended in [32, 31] also uses patches to denoise the image. Though, for these methods, the key point is to get a good representation for each patch of the image by using dictionary learning techniques, well-suited for overcomplete representation (based for instance on  $\ell_1$  regularization or greedy algorithms). This leads to a *state-of-the-art* denoising method called Learned Simultaneous Sparse Coding (LSSC) [31]. It has to be noted that this algorithm incorporates a prior on images by initializing a dictionary of patches on a large clean dataset. Another *state-of-the-art* method in denoising is BM3D [7]. It also relies on patches and consists of a smart combination of classical filtering techniques, such as wavelet denoising and Wiener’s Filter, applied in the space of patches.

These methods are already quite efficient at dealing with smooth regions and textures. However, since they use patches with a fixed square shape and a fixed scale over the whole image, their performances may be limited for dealing with edges, mostly for high contrasted edges. Indeed, high contrasted edges present few redundancies and their denoising version suffer from a persistence of residual noise: this is named the *noise halo*. In

order to overcome this drawback, more directional priors may be considered, using locally chosen scales and orientations of shapes. Few attempts have been made to use several patch sizes (see [32] for learning with patches or [43] for the NL-Means).

As far as we know, the only work trying to handle variable shapes rather than simple square has recently been proposed by Dabov *et al.* [8] as a way to improve the BM3D algorithm. The authors propose to adapt the shapes used by the algorithm: they locally select a shape by applying Lepski’s method, and then perform the same steps of the BM3D algorithm with these shapes rather than with common square blocks (i.e., patches).

Recently, spatial-adaptive methods for selecting the parameters of NL-means have been proposed in [16, 15]. Both methods propose to locally select the parameters which minimize a local estimate of the risk (i.e., the Mean Square Error, MSE) by considering respectively Stein’s Unbiased Risk Estimate (SURE) [46, 47] or the  $C_p$  criterion [33]. The use of SURE for NL-Means was originally proposed in order to select the best bandwidth parameter [51]. Applications of SURE emerged for choosing the smoothing parameter in families of linear estimates [29] such as for model selection, ridge regression, smoothing splines, etc. It was then widely used in the wavelet community after the introduction of the SURE-Shrink algorithm [14]. Solo [44] gave a general form of SURE for an estimator defined as a minimizer of a regular energy, especially for least square regression regularized by a Sobolev norm or the Total Variation. More recently, linear combinations of estimates based on SURE were considered [3] instead of the selection of a single one. Moreover, Ramani *et al.* [39] have proposed a Monte Carlo approach to evaluate SURE when a closed-form expression is not available or too computer-intensive. Our paper is in the same vein as [51, 16] and proposes to locally select or aggregate the best shapes.

*Our contributions*— The aim of this paper is to investigate the potential benefit of replacing the simple square patches with more general shapes, in the classical NL-Means filter. To this end, we propose a fast algorithm, based on the fast Fourier transform, which allows to compute the solution of the NL-Means for arbitrary patch shapes. Then we explain how to combine the estimators associated with each shape in a suitable way. We select or combine locally the shape-based estimates by measuring the performance of their associated denoisers with SURE. We coin such type of algorithms Non-Local Means with Shape-Adaptive Patches (NLM-SAP).

The main advantage of using adaptive patch shapes in the context of NL-Means is to reduce the *noise halo* produced around high contrasted edges. Such edges can appear anytime in natural images and they appear commonly in high dynamic range images (HDR). It has been shown that high contrasts in synthetic aperture radar images interfere with Non-Local approaches since these kinds of images present very high contrasted features (known as bright scatterers) [13]. Our method is an improvement of the NL-Means taking into account the anisotropy of natural images. It is all the more relevant when the images to denoise present high contrasted edges for which the classical NL-Means fails.

*Organization of the paper* — We give in Section 2 a general overview of the NL-Means method. We introduce in Section 3 a more general framework using general shapes instead of square patches to measure the similarity between pixels. This leads in Section 4 to the natural problem of locally selecting or combining the best shapes in our NLM-SAP algorithm. In Section 5 we illustrate numerically, and above all visually, the gain in aggregating various shape-based estimates in a proper manner.

## 2 An overview of the NL-Means

In this section we define the model we use for the noisy image, and we recall the definition and basic facts about NL-Means. We also present some limitations and justify the need to introduce variable shapes instead of the standard fixed square patches.

### 2.1 The AWGN model

In this paper we are concerned with the problem of denoising. A noisy image  $\mathbf{Y}$  is given and assumed to be a version of an unobserved deterministic image  $\mathbf{f}$  corrupted by an Additive White Gaussian Noise (AWGN). Let  $\Omega \subset \mathbb{Z}^2$  be the (bounded) indexing set of the pixels. For any pixel  $x$  in the grid  $\Omega$ , the model has the following formulation:

$$\mathbf{Y}(x) = \mathbf{f}(x) + \varepsilon(x), \quad (1)$$

where  $\varepsilon$  is a centered Gaussian noise with known variance  $\sigma^2$  and the noise components  $\varepsilon(x)$  are independent. Even if we only focus on the case of AWGN, our methodology could be adapted to other noise models by considering NL-Means extensions such as [24, 12] and statistical results on unbiased risk estimation [21].

### 2.2 Definition and parameters

In this section we present the usual definition of the NL-Means procedure as introduced by Buades *et al.* [5]. It is an averaging filter, i.e., for each pixel the output of the procedure is a weighted average of the whole image. The weights used are selected using a “metric” which determines whether two pixels are similar or not. For instance, with the Bilateral Filter [49], the “metric” consists in comparing both pixels values and pixel locations. The core idea of the NL-Means is to create a metric governed by patches surrounding each pixel, regardless of their position, i.e., non-local in the image space. For any pixel of interest  $x$ , an estimate of the value  $\mathbf{f}(x)$  is given by:

$$\hat{\mathbf{f}}(x) = \frac{\sum_{x' \in \Omega} \omega(x, x') \mathbf{Y}(x')}{\sum_{x' \in \Omega} \omega(x, x')}, \quad (2)$$

where the weights  $\omega(x, x')$  depend on patches around  $x$  and  $x'$ . The denominator is a normalizing factor which ensures the weights sum to one. For a fixed (odd) width  $p$ , a patch  $\mathbf{P}_x$  is a subimage of width  $p$ , centered around the pixel of interest  $x$ :

$$\mathbf{P}_x = \left( \mathbf{Y}(x + \tau), \tau \in \left[ \left[ -\frac{p-1}{2}, \frac{p-1}{2} \right] \right]^2 \right). \quad (3)$$

With this notation, the original weights in the NL-Means are of the following form:

$$\omega(x, x') = \varphi \left( \frac{\|\mathbf{P}_x - \mathbf{P}_{x'}\|_{2,a}^2}{2h^2} \right), \quad (4)$$

where  $h > 0$  is the bandwidth parameter,  $\varphi$  is the kernel used to measure similarity between patches,  $\|\cdot\|_{2,a}$  is a weighted Euclidean norm using a Gaussian kernel, and  $a$  is the bandwidth that controls the concentration of the kernel around the central pixel. Let us now briefly recall the influence of each parameter (see [16] for a more extensive discussion on this subject).

**The bandwidth  $h$ :** this parameter has a smoothing effect and plays the same role as the bandwidth for kernel methods in statistics (cf. [55] for more details). With our parametrization, the larger the bandwidth, the smoother the image becomes. Choosing this parameter is a difficult task and many solutions have been proposed in the literature. The simplest and most common one is to set a single  $h$  for the whole image, whose value is determined by cross validation on a small dataset of images. In [37, 23], the authors set this parameter according to the quantile of a  $\chi^2$  distribution, due to the particular metric they consider to compare pixels (or patches). Van De Ville and



**Fig. 1** Illustration of the *noise halo* appearing around high contrasted edges on images denoised by NL-means. The input noisy images was corrupted version of the noise-free images presented on Fig. 6 damaged by AWGN with standard deviation  $\sigma = 20$ . Noise halo arises from an abrupt lack of redundancy around edges sometimes referred as the *rare patch effect*.

Kocher [51] calculate an unbiased risk estimate of the NL-Means to globally select the bandwidth  $h$ . Pursuing this idea, Duval *et al.* [16] consider a method based on the same approach but to locally select the bandwidth. The main interest of the last two approaches is that they give a closed-form expression of the SURE for NL-Means.

**The search window size  $\ell$ :** the summation defined in Equation (2) is usually restricted to an  $\ell \times \ell$  search window around the pixel of interest. Defining  $\Omega_\ell = \llbracket -\frac{\ell-1}{2}, \frac{\ell-1}{2} \rrbracket^2$ , the search window centered on each pixel  $x$  is then  $x + \Omega_\ell$ . This was proposed in the seminal work [5] for computational acceleration. Though, some authors have also noticed that choosing locally the best search windows [23] or restricting the average over small ones (see [16] or [41] for more details) could benefit to the NL-Means procedure.

**The kernel  $\varphi$ :** function  $\varphi$  was chosen by Buades *et al.* [5] as  $t \mapsto \exp(-t)$ , but other choices may be considered, such as compactly supported smooth functions. It was noticed by several authors [20] that weights with compact support yield better results. In this paper we have restricted our experiments to the kernel given in [5], but any other  $\mathcal{C}^1$  kernel could be handled by our framework.

**The patch size  $p$ :** this parameter is generally chosen equal to 5, 7 or 9. Using a width  $p = 1$  would lead to a method close to the Bilateral Filter [49] or Yaroslavsky Filter [56]. Moreover, in most papers the patch width is a global fixed parameter. Few works have tried to handle the difficult task of using several sizes of patches for a single image. To our knowledge, the first attempt was proposed in the context of learning patches by [32], using Support Vector Machines (SVM).

Our approach extends the notion of patches to the more general concept of shapes. Our goal is to better handle anisotropy in natural images using oriented patches. Indeed, it is well known that edges are not satisfactorily denoised when using square patches with the NL-Means. This leads to a *noise halo* around edges (see Fig. 1). We propose in the next section an algorithm to compute efficiently the solution of NL-means for arbitrary shapes. We also consider in details several natural collections of shapes. Numerical results are given in Section 5.

### 3 From patches to shapes: beyond the *rare patch effect*

In practice, we have seen that the original algorithm suffers from a *noise halo* around edges, due to an abrupt lack of redundancy of the image. This phenomenon is sometimes referred to as the *rare patch effect* (Fig. 1). Statistically, it leads to an NL-Means estimator with large variance around edges. Several solutions have already been proposed to handle this drawback [43, 16, 30] (see Section 3.2 for more details). Here, we generalize these approaches by considering general shapes rather than simple square patches (cf. Fig. 2).

In order to deal with patches of arbitrary shapes, we reformulate the way the distance between two pixels is measured in terms of patches. The weighted Euclidean distance  $\|\cdot\|_{2,a}$  used in Eq. (4) can be generalized using the following expression:

$$d_{\mathcal{S}}^2(x, x') = \sum_{\tau \in \Omega} \mathcal{S}(\tau) (\mathbf{Y}(x + \tau) - \mathbf{Y}(x' + \tau))^2, \quad (5)$$

where  $\mathcal{S}$  encodes the shape we aim at. With this notation we can easily rewrite the original NL-Means with a simple  $\mathcal{S}$ , cf. Eq. (8) and Eq. (9).

One of our contributions is to provide an efficient algorithm, based on the Fast Fourier Transform (FFT), to compute the distances in Eq. (5). Our implementation

is independent of the shape, and can thus be applied with different shapes (see Section 3.1 for details). As soon as we consider the use of anisotropic shapes, and not just squares or disks centered on the pixel of interest, two questions emerge. The first one is how to choose the collection of shapes to consider. The second issue is to propose a way to combine the estimators provided by each shape.

The collection of shapes should be composed of more than one shape to locally take into account the geometrical properties of natural images. Consider for instance the use of a single vertically elongated patch. This could be interesting to handle vertical features, but eventually we would not optimally deal with horizontal details with only one oriented shape. So, the collection should be diversified and numerous enough to identify directional features (see Fig. 8 for a visual illustration). At the same time, it should remain small enough so that the algorithm is not computationally intensive.

### 3.1 Fast algorithm to handle shapes

In this section, we present a fast way to compute the NL-Means weights for general shapes, based on the 2D-FFT. It is inspired from works initiated in [53] and [11] to speed up the NL-Means algorithm. However, contrary to these approaches, ours can deal with non-square and/or non-binary patches, i.e., with general shapes  $\mathcal{S}$ . Like them, our method is independent of the shape size.

In [53] and [11], the authors propose to compute the Euclidean distances using ‘‘Summed Area Tables’’ [6] (also called ‘‘Integral Images’’ [52]). This allows them to reduce the computational cost of the NL-Means from  $O(|W| \cdot |\Omega| \cdot |P|)$  to  $O(|W| \cdot |\Omega|)$ , where  $|W| = \ell^2$  is the number of pixels in the search window,  $|\Omega|$  is the image domain size and  $|P|$  is the patch size (we refer to [11] for more details). To compute these integral images, the authors change the original algorithm by swapping the two ‘‘for’’ loops: instead of considering all the shifts for each pixel, they consider all the pixels for each shift.

We use basically the same swapping trick. Notice that Equation (5) can be reformulated for any translation parameter  $\delta$  (i.e., taking  $x' = x + \delta$ ) as a discrete convolution:

$$\begin{aligned} d_{\mathcal{S}}^2(x, x + \delta) &= \sum_{\tau \in \Omega} \mathcal{S}(\tau) (\mathbf{Y}(x + \tau) - \mathbf{Y}(x + \delta + \tau))^2 \\ &= (\check{\mathcal{S}} \star \Delta_{\delta})(x), \end{aligned} \quad (6)$$

where  $\check{\mathcal{S}}(\tau) = \mathcal{S}(-\tau)$ ,  $\Delta_{\delta}(x) = (\mathbf{Y}(x) - \mathbf{Y}(x + \delta))^2$  and  $\star$  is the convolution operator. This term can be interpreted as the correlation between the shape  $\mathcal{S}$  and

---

#### Algorithm 2D-FFT NL-Means for an arbitrary shape

---

**Inputs:** noisy image  $\mathbf{Y}$ , 2D-FFT of the shape  $\mathcal{F}(\mathcal{S})$

**Parameters:** search window  $W$ , bandwidth  $h$

**Output:** estimated image  $\hat{\mathbf{f}}$

Initialize accumulator images  $\mathbf{A}$  and  $\mathbf{B}$  to zero  
**for all** shift vector  $\delta$  in the search window  $W$  **do**  
  Compute the square difference image  $\Delta_{\delta}$

$$\Delta_{\delta}(x) := (\mathbf{Y}(x) - \mathbf{Y}(x + \delta))^2 \text{ for all pixels } x$$

Compute the 2D-FFT  $\mathcal{F}(\Delta_{\delta})$   
Perform the convolution of  $\Delta_{\delta}$  by the shape  $\check{\mathcal{S}}$

$$d_{\check{\mathcal{S}}}^2(\cdot, \cdot + \delta) \leftarrow \left( \mathcal{F}^{-1} \left( \overline{\mathcal{F}(\check{\mathcal{S})}} \mathcal{F}(\Delta_{\delta}) \right) \right) (\cdot)$$

▷  $O(|\Omega| \cdot \log |\Omega|)$  operations using 2D-FFT

**for all** pixels  $x$  in  $\Omega$  **do**

  Compute the weights

$$\omega(x, x + \delta) = \varphi \left( \frac{d_{\check{\mathcal{S}}}^2(x, x + \delta)}{2h^2} \right)$$

  Update the accumulators

$$\mathbf{A}(x) \leftarrow \mathbf{A}(x) + \omega(x, x + \delta) \mathbf{Y}(x + \delta)$$

$$\mathbf{B}(x) \leftarrow \mathbf{B}(x) + \omega(x, x + \delta)$$

**end for**

**end for**

Final (normalized) estimator  $\hat{\mathbf{f}}(x) = \frac{\mathbf{A}(x)}{\mathbf{B}(x)}$  for all pixel  $x$

Note: the central pixel ( $\delta = 0$ ) is treated as a special case  
  ▷ see Section 5 for details

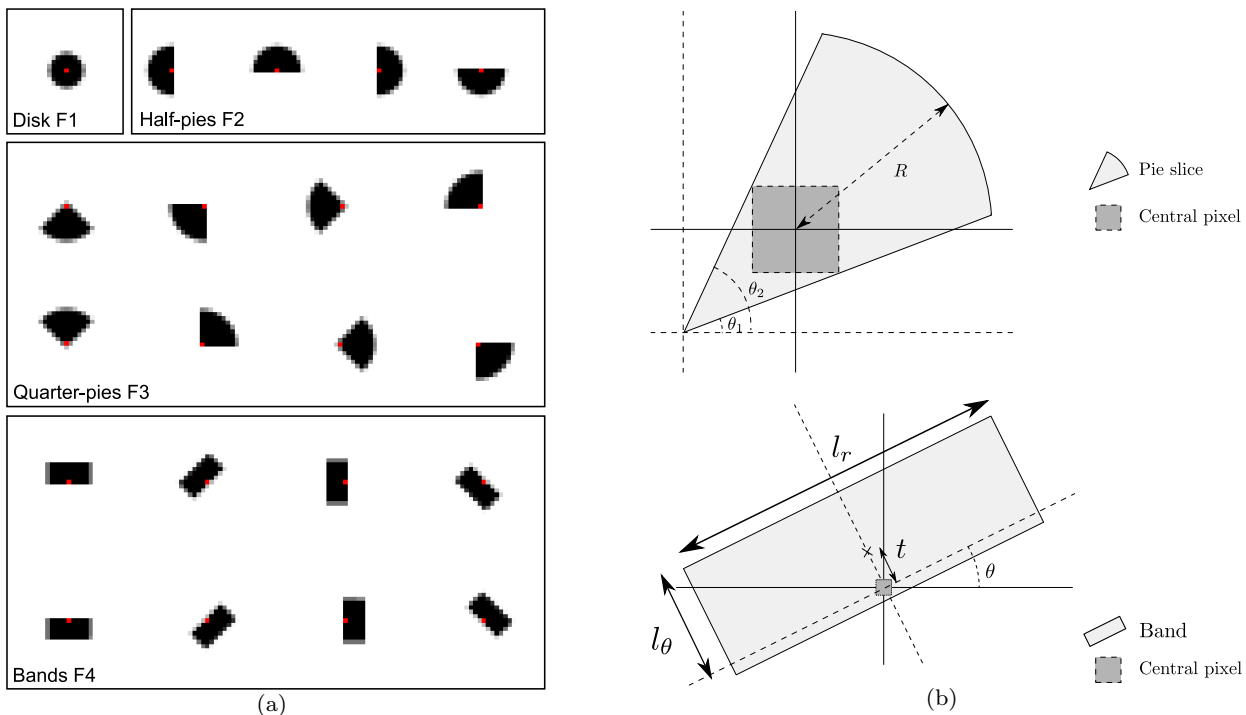
---

**Fig. 3** NL-Means pseudo-code for an arbitrary patch shape  $\mathcal{S}$ . Pre-computations (based on 2D-FFT) of distances between shapes from the noisy image and shapes from its shifted version leads to a smaller complexity of  $O(|W| \cdot |\Omega| \cdot \log |\Omega|)$ , independent of the shape  $\mathcal{S}$ .

the square difference of the observe image and the  $\delta$ -shifted version. The convolution  $\check{\mathcal{S}} \star \Delta_{\delta}$  can be computed quickly thanks to following relation:

$$\check{\mathcal{S}} \star \Delta_{\delta} = \mathcal{F}^{-1}(\mathcal{F}(\check{\mathcal{S}})\mathcal{F}(\Delta_{\delta})) = \mathcal{F}^{-1}(\overline{\mathcal{F}(\mathcal{S})}\mathcal{F}(\Delta_{\delta})), \quad (7)$$

where  $\mathcal{F}$  is the 2D discrete Fourier transform (2D-FFT) and  $\mathcal{F}^{-1}$  is its inverse transform. According to Equation (7), and given a translation  $\delta$ , we only need to perform one term by term multiplication in Fourier domain and two 2D-FFT (note that  $\mathcal{F}(\mathcal{S})$  can be computed off-line). The repetition of this procedure for every translation  $\delta$  covering the search window, leads to an algorithm (whose pseudo-code is detailed in Fig. 3) with a complexity of  $O(|W| \cdot |\Omega| \cdot \log(|\Omega|))$ .



**Fig. 2** (a) Examples of shapes considered. The “central” pixel is shown in red, dark pixels illustrate high weights. Shapes are grouped in four categories: F1. the isotropic *disk* family, F2. the *half-pies* family, F3. the *quarter-pies* family and F4. the *bands* family. (b) Parametrization of the *pie* slices and *bands*.

### 3.2 Families of shapes

The main purpose of this paper is to show that the use of different shapes allows to reduce the *rare patch effect*. This point of view is a generalization of the NL-Means based on square patches with the reprojection studied in [43], since each translated patch can be regarded as a de-centered shape. Here,  $h$  is fixed and the challenge is to find shapes with enough similar candidates in the search window to reduce the noise.

We now present several types of families we have considered. The first collections consist of classical squares and disks shapes. Then, we propose more directional shapes such as *pie slices* and *bands* displayed in Fig. 2.

**Squares:** To begin with, we apply our framework to the most commonly used shapes, i.e., the square shapes of odd length (so the squares have centers we can consider). For instance, choosing:

$$S(\tau) = \begin{cases} 1, & \text{if } \|\tau\|_\infty \leq \frac{p-1}{2}, \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

leads to the classical (simplified) NL-Means definition with square patches of size  $p \times p$  and distance between

patches measured by the Euclidean norm.

**Gaussian:** The original, but less common choice, is to set:

$$S(\tau) = \begin{cases} \exp(-(\tau_1^2 + \tau_2^2)/2a^2), & \text{if } \|\tau\|_\infty \leq \frac{p-1}{2}, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

Equation (9) means that the norm  $\|\cdot\|_{2,a}$  is used to measure the distance between patches. This limits the influence of square patches corners and leads to a more isotropic comparison between patches.

**Disks:** Disk shapes are defined in the same way, using the Euclidean norm instead:

$$S(\tau) = \begin{cases} 1, & \text{if } \|\tau\|_2 \leq \frac{p-1}{2}, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

A non-binary version may also be defined for pixels crossed by the boundary.

**Pie slices:** We study a family of shapes, denoted as “pie”, whose elements are defined with three parameters: two angles and a radius. These shapes represent a portion of a disk delimited by two lines

and surrounding the discrete central pixel.

**Bands:** This family of shapes is simply composed of rectangles, potentially rotated and decentered with respect to the pixel of interest.

### 3.3 Connection with previous work

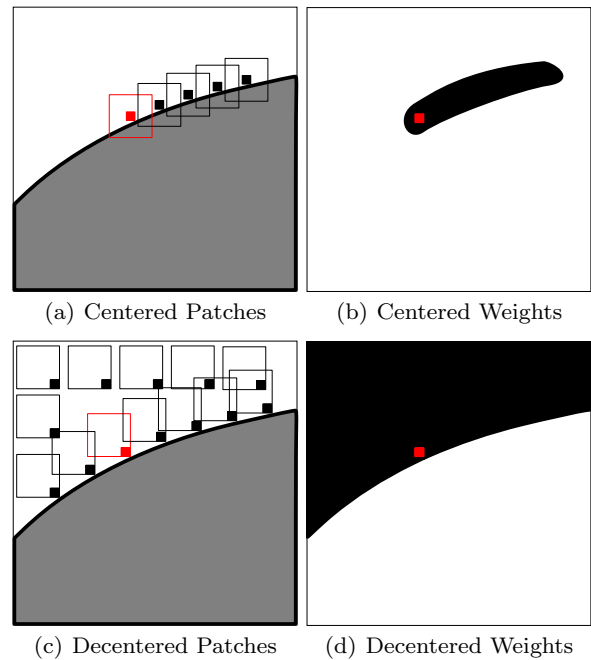
One of our main concern is to address the *rare patch effect* of the NL-Means algorithm. Different methods have been designed to limit this drawback or to improve the NL-Means in terms of quality or speed. The first attempt was proposed to speed-up the algorithm. The idea is to denoise patch by patch rather than pixel by pixel. Taking into account patches overlaps, a fast implementation of the NL-Means is reached by using a sub-sampled grid of pixels [5,24]. Quality improvement can also be obtained by properly using overlapping patches. Indeed, we get  $|\mathcal{P}|$  estimates for each pixel (where  $|\mathcal{P}|$  is the number of elements in a patch). Some authors [5,23] simply propose to uniformly average those  $|\mathcal{P}|$  estimates while a weighted average is performed in [43] (see Section 4.1 and 4.2). In our framework, these blockwise approaches are equivalent to use  $|\mathcal{P}|$ , possibly decentered, square shapes (cf. Fig. 4).

Other methods have been introduced to reduce the *noise halo*. In [30], the authors use a total variation-based pre-filtering of the image and set locally its parameter so that the NL-Means find enough similar patches. In [16], the authors aim to select locally the bandwidth parameter  $h$  using SURE (introduced in details in Section 4.3) to select enough patches according to a bias-variance trade-off.

## 4 Aggregation of shape-based estimates

In this section we investigate several ways to aggregate the NL-Means estimators based on different shapes of “patches”. We have extended the standard square shape to other shapes such as *disks*, *pies* or *bands* (see Section 3.2). Thus, the new goal in this context is to determine how to locally take the most of each proposed denoiser.

Assume that for any pixel  $x$  in the image, we have built a collection of  $K$  pixel estimators  $\hat{\mathbf{f}}_1(x), \dots, \hat{\mathbf{f}}_K(x)$  based on different shapes, and that estimates of their corresponding performances are available. With this information at hand, we can address different aggregation/optimization problems. Indeed, the ways of combining the estimators at hand may depend on the theoretical aggregation problem we aim to solve (as described by [34] and [50]):



**Fig. 4** Examples of decentered patches near edges. If the patch is centered (a) fewer similar patches candidates are found than if the patch is decentered (c). The pixel of interest is in red, and black pixels are some similar pixels. (b) and (d) gives an illustration of the weights obtained (in black) by the two methods to denoise the red pixel.

- (S) The selection problem: finding the best estimator among  $\hat{\mathbf{f}}_1(x), \dots, \hat{\mathbf{f}}_K(x)$ .
- (L) The linear problem: finding the best linear combination of  $\hat{\mathbf{f}}_1(x), \dots, \hat{\mathbf{f}}_K(x)$ .
- (C) The convex problem: finding the best convex combination of  $\hat{\mathbf{f}}_1(x), \dots, \hat{\mathbf{f}}_K(x)$ .

### 4.1 Uniformly weighted aggregation (UWA)

This is the simplest form of aggregation that we consider. The idea is to give the same weight to any shape-based estimator. With few shapes it is already an improvement in practice (both in term of PSNR and SSIM, see Table 3), but as the number of shapes increases, we can take into account positions that are completely irrelevant. It is in essence, the one proposed in [5] and referred as blockwise NL-means in [23], with shapes being decentered square patches. Such an aggregation procedure still suffers from the *rare patch effect* (see Fig. 1, third line, for a visual illustration). We can express for any pixel  $x$  the uniform estimate  $\hat{\mathbf{f}}_{\text{UWA}}(x)$  as the mean of the shape-based estimates:

$$\hat{\mathbf{f}}_{\text{UWA}}(x) := \frac{1}{K} \sum_{k=1}^K \hat{\mathbf{f}}_k(x). \quad (11)$$



In our experiments, this method can be considered as the benchmark to be improved when using more complex aggregating strategies.

#### 4.2 Variance-based decision, Weighted Average (WAV)

A possible way to limit the halo effect is to adapt WAV-reprojection [43] to general shapes. The idea, also proposed by Dabov *et al.* [7] in a different context, is to perform a weighted average of the estimates  $\hat{\mathbf{f}}_1(x), \dots, \hat{\mathbf{f}}_K(x)$ . To limit the *noise halo*, each weight should be chosen inversely proportional to the (approximate) variance of the corresponding estimator. In the context of NL-means, this approximate variance can be obtained in closed-form in the same way as in [23], assuming that the coefficients  $\omega(x, x')$  in Equation (2) can be treated as deterministic. Measuring the performances of the estimators in term of variance is well justified since the halo effect results to the high variance of our estimators around the edges (see [43]). However, it tends to over-smooth the edges and the thin details since it does not consider the bias of each estimator.

#### 4.3 SURE-based decisions

A way to take the bias into account is to consider the risk estimate rather than the variance to locally attribute more weight to the estimators with small risks. In [51], Van De Ville and Kocher give a closed-form expression of Stein's Unbiased Estimator of the Risk (SURE) for the NL-Means. They aim at selecting globally the best bandwidth for a given image. Here, our approach is different, despite the use of the same tool. Indeed, our choice of shape is done locally (i.e., for each pixel), since it is very unlikely that a single shape should be optimal for a whole natural image. Our method is closer to the one proposed by Duval *et al.* [16]. The authors rely on SURE to locally determine the parameters: the bandwidth  $h$  and the patch size  $p$ . They have shown that a local choice of  $h$  reduces the visual artifacts, especially the *rare patch effect*. Let us now rephrase Stein's Lemma [46, 47] in our NLM-SAP framework.

**Proposition 1 (Stein)** *Let  $\mathbf{f}(x) \in \mathbb{R}$ ,  $\varepsilon(x) \sim \mathcal{N}(0, \sigma^2)$ , and  $\mathbf{Y}(x) = \mathbf{f}(x) + \varepsilon(x)$ . Denote by  $\gamma(\mathbf{Y}(x))$  an estimate of  $\mathbf{f}(x)$  and assume  $\gamma: \mathbb{R} \rightarrow \mathbb{R}$  is absolutely continuous, and*

- i)  $\lim_{|z| \rightarrow \infty} \gamma(\mathbf{f}(x) + z) e^{-\frac{z^2}{2\sigma^2}} = 0$ ,
- ii)  $\mathbb{E}(\gamma(\mathbf{f}(x) + \varepsilon(x)))^2 < +\infty$ , and
- iii)  $\mathbb{E}|\gamma'(\mathbf{f}(x) + \varepsilon(x))| < +\infty$ .

*Then the following relation holds:*

$$\begin{aligned} \mathbb{E}|\gamma(\mathbf{Y}(x)) - \mathbf{f}(x)|^2 = \\ \mathbb{E} \left( (\gamma(\mathbf{Y}(x)) - \mathbf{Y}(x))^2 + 2\sigma^2 \frac{\partial \gamma(\mathbf{Y}(x))}{\partial \varepsilon(x)} \right) - \sigma^2. \end{aligned} \quad (12)$$

Proposition 1 provides an unbiased estimate of the risk for the  $K$  shape-based denoised values  $\hat{\mathbf{f}}_k(x)$  at each pixel  $x$ . Indeed, if  $\gamma$  is one of the shape-based estimators and if the kernel  $\varphi$  is such that  $\gamma$  is absolutely continuous, then

$$\mathbf{r}_k(x) = (\hat{\mathbf{f}}_k(x) - \mathbf{Y}(x))^2 + 2\sigma \frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \varepsilon(x)} - \sigma^2, \quad (13)$$

is an unbiased estimate of the risk at pixel  $x$  for the  $k$ -th shape-based estimate, i.e.,  $\mathbb{E}(\mathbf{r}_k(x)) = \mathbb{E}|\hat{\mathbf{f}}_k(x) - \mathbf{f}(x)|^2$ .

The main contribution of [51, 16] is that they give a closed-form expression of  $\frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \varepsilon(x)}$  for NL-Means. Indeed, thanks to Eq. (2) and Eq. (4), its expression can be recast in the following form:

$$\begin{aligned} \frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \varepsilon(x)} = \frac{\varphi(0)}{C_x} + \frac{1}{C_x} \sum_{x'} \mathbf{Y}(x') \frac{\partial \omega(x, x')}{\partial \varepsilon(x')} \\ - \left( \frac{1}{C_x} \sum_{x'} \mathbf{Y}(x') \omega(x, x') \right) \left( \sum_{x''} \frac{\partial \omega(x, x'')}{\partial \varepsilon(x)} \right). \end{aligned} \quad (14)$$

where  $C_x = \sum_{x'} \omega(x, x')$  is a normalization constant. In our NLM-SAP framework, our shape-based norm defined in Eq. (5) leads to the following expression of the derivative of the weights  $\omega(x, x')$ :

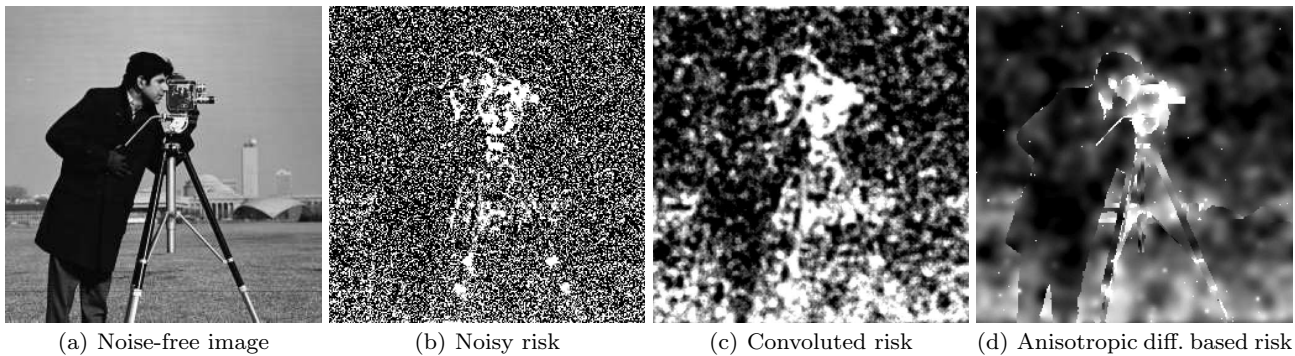
$$\begin{aligned} \frac{\partial \omega(x, x')}{\partial \varepsilon(x')} = \frac{1}{h^2} (\mathbf{S}(x) [\mathbf{Y}(x) - \mathbf{Y}(x')] \\ + \mathbf{S}(2x - x') [\mathbf{Y}(x) - \mathbf{Y}(2x - x')]). \end{aligned} \quad (15)$$

where  $\mathbf{S}$  encodes the shape of our  $k$ -th shape-based estimator. Finally, combining equations (13), (14) and (15) leads to an unbiased estimate of the risk of our NLM-SAP denoiser.

Now, since we are using  $K$  shapes, we get  $\mathbf{r}_1(x), \dots, \mathbf{r}_K(x)$ , unbiased risk estimates respectively for the shape-based estimators  $\hat{\mathbf{f}}_1(x), \dots, \hat{\mathbf{f}}_K(x)$ .

##### 4.3.1 Minimizing the risk of linear combinations

Suppose we want to solve the linear problem of aggregation. A natural way to find a good linear combination



**Fig. 5** Maps of the estimated risk associated to patches with a small circular shape. From left to right, the underlying noise-free image, the map of the risk without regularization, with convolution (isotropic regularization) and with regularization based on anisotropic diffusion. Low risks are black, high ones are white.

is to solve the following problem:

$$\hat{\mathbf{f}}_{\text{LIN}}(x) := \sum_{k=1}^K \alpha_k^* \hat{\mathbf{f}}_k(x), \quad (16)$$

$$\text{where } \alpha^* := \arg \min_{\alpha \in \mathbb{R}^K} \mathbb{E} \left( \sum_{k=1}^K \alpha_k \hat{\mathbf{f}}_k(x) - \mathbf{f}(x) \right)^2.$$

Here, the linearity of the combination allows to use Stein's Lemma (Proposition 1). It provides an unbiased estimate of the risk of the weighted average estimate:

$$\begin{aligned} & \mathbb{E} \left( \sum_{k=1}^K \alpha_k \hat{\mathbf{f}}_k(x) - \mathbf{f}(x) \right)^2 = \\ & \mathbb{E} \left( \left( \sum_{k=1}^K \alpha_k \hat{\mathbf{f}}_k(x) - \mathbf{Y}(x) \right)^2 + 2\sigma^2 \sum_{k=1}^K \alpha_k \frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \boldsymbol{\varepsilon}(x)} \right) - \sigma^2. \end{aligned}$$

Therefore, neglecting the term that does not depend on  $\alpha$  in the last equation, the optimal weights  $\alpha^*$  can be obtained by solving the following:

$$\begin{aligned} \alpha^*(x) = \arg \min_{\alpha \in \mathbb{R}^K} & \sum_{k,k'} \alpha_k \alpha_{k'} \mathbb{E}(\hat{\mathbf{f}}_k(x) \hat{\mathbf{f}}_{k'}(x)) + \\ & \sum_{k=1}^K \alpha_k \left( 2\sigma^2 \mathbb{E} \left( \frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \boldsymbol{\varepsilon}(x)} \right) - 2\mathbb{E}(\mathbf{Y}(x) \hat{\mathbf{f}}_k(x)) \right). \end{aligned}$$

Using the first order optimality conditions, the last problem amounts to solving a linear system in  $\alpha$ . This type of estimator is known to perform quite well for wavelet thresholding estimation where it is referred to as the SURE-LET (for SURE-Linear Expansion of Thresholds, see [3]). However, in our framework, this method is not applicable. Indeed, since we work pixelwise, we have only one sample of the SURE for each pixel. This leads to an ill-conditioned system to solve, and the calculated  $\alpha^*$  behaves poorly. Algebraically, this is due to the fact that the matrix

$[\hat{\mathbf{f}}_k(x) \hat{\mathbf{f}}_{k'}(x)]_{k,k'=1,\dots,K}$  is of rank one, and thus the system to solve is ill-conditioned. Our attempts to regularize the problem (e.g. with Tikhonov regularization known as ‘‘Ridge Regression’’ in statistics) or to solve the problem blockwise instead of pixelwise did not yield satisfying results. More generally, the problem of using SURE to take a local decision for each pixel  $x$  is difficult since this estimator has large oscillations (see for instance Fig. 5). In the next paragraph, we present how to regularize the risk maps, i.e., the ‘‘images’’  $\mathbf{r}_1, \dots, \mathbf{r}_K$ .

#### 4.3.2 Regularizing the risk maps

To make the risk estimates more robust, it is necessary to locally regularize the risk maps, in order to approximate at each pixel the expectations used in Eq. (13).

In [16], the convolution of the risk map is an efficient way to estimate the local risk in view of setting  $h$  since on both sides of an edge a large value of  $h$  should be used. Here, the anisotropy of the shapes implies that on one side of an edge the risk may be low whereas it may be very high on the other side. Since the convolution diffuses the risks on both sides of the edges, any comparison of the risks associated with each shape becomes unstable.

In order to diffuse the risks on each side of edges, we have adopted a heat equation with spatially and timely dependent coefficients (inspired by the Perona-Malik equation [35]). So, we let the risk maps  $\mathbf{r}_1, \dots, \mathbf{r}_K$  evolve according to:

$$\begin{cases} \frac{\partial \mathbf{r}_k}{\partial t}(x, t) = \text{div} (g(|\nabla u(x, t)|) \nabla \mathbf{r}_k(x, t)), \\ \mathbf{r}_k(x, 0) = (\hat{\mathbf{f}}_k(x) - \mathbf{Y}(x))^2 + 2\sigma^2 \frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \boldsymbol{\varepsilon}(x)} - \sigma^2, \end{cases} \quad (17)$$

where  $g(x) = \exp(-x^2/\kappa^2)$ , the parameter  $\kappa$  controls the anisotropy of the diffusion (the larger  $\kappa$ , the more isotropic the diffusion), and  $u$  is the smoothed noisy



**Fig. 6** Chosen  $256 \times 256$  noise-free images for our experiments, from left to right: *cameraman*, *city*, *windmill* and *lake*. These images present high contrasted edges for which the classical NL-Means suffer from the *rare patch effect*.

image which evolves at the same time using the Perona-Malik equation:

$$\begin{cases} \frac{\partial u}{\partial t}(x, t) = \operatorname{div}(g(|\nabla u(x, t)|) \nabla u(x, t)), \\ u(x, 0) = \mathbf{Y}(x). \end{cases} \quad (18)$$

Curiously, we have noticed that we obtain better risk maps by diffusing  $\sqrt{\mathbf{r}_k}$  instead of  $\mathbf{r}_k$  itself. Fig. 5 shows that this regularization procedure provides smooth risk maps, following edges of the underlying noise-free image, and finer than without regularization or with convolution.

#### 4.3.3 Minimizer of the risk estimates (MRE)

With our measure of performance, the most natural way to address the selection problem (S) is to select the shape that minimizes the local risk estimate we have at hand:

$$\hat{\mathbf{f}}_{\text{MRE}}(x) := \hat{\mathbf{f}}_{k^*}(x) \quad \text{where} \quad k^* = \arg \min_k \mathbf{r}_k(x). \quad (19)$$

Such a selection rule is all the more relevant as the shapes are really different. In most cases, selecting the locally optimal shape yields satisfying results, but it is sometimes more appropriate to combine some of the best performing estimators as in the next paragraph.

#### 4.3.4 Exponentially weighted aggregation (EWA)

In many cases, it might be better to combine several estimators rather than just selecting one. Especially, it happens to be relevant if the best estimators (in term of evaluated risk) are diversified enough or if the risk of the MRE was wrongly under-estimated. Thus, we have also used the statistical method of Exponentially Weighted Aggregation as introduced by Leung and Barron [28]. This method has been theoretically studied in [9, 10] and adapted for patch-based denoising in [42]. It

consists in aggregating the estimators by performing a weighted average with weights based on the confidence attributed to each estimator, measured in term of the risk. More precisely:

$$\hat{\mathbf{f}}_{\text{EWA}}(x) := \sum_{k=1}^K \alpha_k \hat{\mathbf{f}}_k(x), \quad (20)$$

with 
$$\alpha_k = \frac{\exp(-\mathbf{r}_k(x)/T)}{\sum_{k'=1}^K \exp(-\mathbf{r}_{k'}(x)/T)}.$$

The temperature parameter  $T > 0$  is a smoothing parameter, that controls the confidence attributed to the risk estimates. If  $T \rightarrow \infty$ , then the EWA is simply the uniform aggregate  $\hat{\mathbf{f}}_{\text{UWA}}$  defined before. Conversely, when  $T \rightarrow 0$ , then  $\hat{\mathbf{f}}_{\text{EWA}} \rightarrow \hat{\mathbf{f}}_{\text{MRE}}$ . Most theoretical works about EWA (see [28, 9, 10]) recommend a large temperature parameter  $T = 4\sigma^2$  under few assumptions (like independence) on the estimators  $\hat{\mathbf{f}}_1(x), \dots, \hat{\mathbf{f}}_K(x)$ . In practice, since assumptions on the estimators family may not be satisfied, we have used a smaller value,  $T = 0.02\sigma^2$ .

## 5 Numerical and Visual Results

This section presents quantitative and qualitative results obtained on four images synthetically corrupted by AWGN. The corrupted images are obtained from four  $256 \times 256$  noise-free images presented on Fig. 6: the famous *cameraman* image and *city*, *windmill* and *lake*<sup>1</sup>. These images are particularly interesting in the study of our proposed Non-Local Means with Adaptive Patch Shapes (NLM-SAP) since they present high contrasted edges for which the classical NL-Means suffer from the *rare patch effect* (see Fig. 1).

<sup>1</sup> three sub-images extracted from the Laurent Condat's database: <http://www.greyc.ensicaen.fr/~lcondat/>

**Table 1** Gain by using multi-scale isotropic shapes in terms of PSNR and SSIM values (PSNR/SSIM). Circular patch shapes respectively with area of 12.5, 25 and 50 px<sup>2</sup> are used and their results are compared to the one obtained when using a combination of these three scales of patches.

	Cameraman	City	Windmill	Lake
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.385	22.13/0.456
Patch shape area of 12.5 px <sup>2</sup>	<b>29.56</b> /0.822	28.11/0.873	30.91/0.880	28.67/0.849
Patch shape area of 25 px <sup>2</sup>	29.33/0.829	27.93/0.880	30.72/0.896	28.44/0.855
Patch shape area of 50 px <sup>2</sup>	28.99/0.825	27.56/0.879	30.35/0.900	28.30/0.857
Combination of these three scales	29.45/ <b>0.832</b>	<b>28.16/0.885</b>	<b>30.97/0.904</b>	<b>28.68/0.863</b>

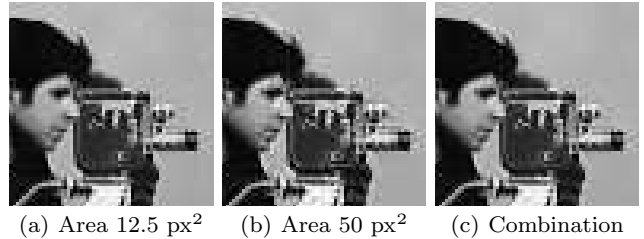
In all the experiments, unless otherwise specified, the NLM-SAP is used with the following default parameters:

- search window: width  $\ell = 11$  px,
- shape family: 15 shapes from families F1 and F2 on Fig. 2.a with shape areas of 12.5, 25 and 50 px<sup>2</sup>,
- aggregation: EWA with  $T = 0.02\sigma^2$ , and
- risk regularization: 50 iterations with timestep  $dt = 1/8$  and  $\kappa = 30$ .

We set the bandwidth parameter to  $h = 0.7\sigma$  for patches of area 50 px<sup>2</sup> as suggested in [5]. However, as soon as we consider shapes of different areas, the parameter  $h$  has to adapt to the sizes of the shapes. To adapt the bandwidth to other sizes, we thus have extended a rule proposed by [23,37], based on quantiles of a  $\chi^2$  distribution, to the norm defined in Eq. (5). The quantile is selected such that  $h = 0.7\sigma$  for patches of area 50 px<sup>2</sup>. For the central pixel, we set its central weight as recommended in [41].

The main limitation in computing time is due to the number  $K$  of shapes required by our NLM-SAP algorithm. We need to perform  $K$  times an NL-Means like algorithm, i.e., one for each shape. Thanks to our FFT acceleration, the computing time required for one shape, whatever the shape, is of about 2s for a  $256 \times 256$  image with a Matlab implementation on an Intel Pentium 64-bit, 3.00 GHz. By comparison, the naive Matlab implementation of NL-Means takes about 100s, for square patches of area  $7 \times 7$  px<sup>2</sup>. The computation of one local SURE map, using Perona-Malik diffusion, takes about 1s per shape. Finally, NLM-SAP<sup>2</sup> using 15 shapes leads to a computing time of about 45s which is less than the naive Matlab implementation of NL-Means.

Visual results are given to assess the denoising qualities relative to the different settings of NLM-SAP and to compare NLM-SAP with other denoising approaches (see Fig. 12). Numerical criteria support our claims: the



**Fig. 7** Circular patch shapes: visual results obtained with NLM-SAP when using only circular shapes of different scales on a noisy realization of the *cameraman* image. The patch shape areas are respectively, from left to right, 12.5 px<sup>2</sup>, 50 px<sup>2</sup> and a combination of 12.5, 25 and 50 px<sup>2</sup>.

Peak Signal to Noise Ratio (PSNR) defined by

$$\text{PSNR}(\hat{f}, f) = 10 \log_{10} \frac{255^2}{\frac{1}{|\Omega|} \sum_{x \in \Omega} (\mathbf{f}(x) - \hat{\mathbf{f}}(x))^2}, \quad (21)$$

and the Structural Similarity (SSIM) defined in [54]. The SSIM is between 0 and 1 and a value closed to 1 means that the estimated image has a similar structure to the noise-free image.

### 5.1 Behavior of NLM-SAP

In this section, we will study the behavior of NLM-SAP according to some parameters such as the type of family, the type of aggregation and the type of risk regularization. Each noisy image is corrupted by AWGN with standard deviation  $\sigma = 20$ .

Table 1 and Figure 7 illustrate the gain of performance to use multi-scale patch shapes instead of using only one fixed size. In this experiment, we consider three circular shapes of areas: 12.5, 25 and 50 px<sup>2</sup>. Comparatively, for the original version of NL-means, the authors suggest to use square patches of fixed size  $7 \times 7 = 49$  px<sup>2</sup>. Surprisingly, using the smallest shapes provides always the best PSNR. It means that the bias and the *noise halos* introduced by using large patches are actually more penalizing than the remaining noise left by the use of small patches. The aggregation of these three scales of shapes with our NLM-SAP methods improves

<sup>2</sup> our Matlab implementation will be available on-line when this paper is published

**Table 2** Gain by using anisotropic or mixture of isotropic and anisotropic shapes in terms of PSNR and SSIM values (PSNR/SSIM). The studied patch shapes are the isotropic *disks*, the *half-pies*, the *quarter-pies*, the *bands* and some combination of them (see Fig. 2.a).

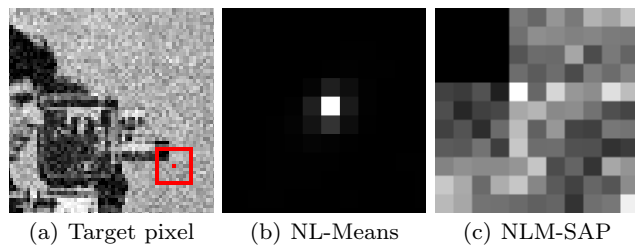
	Cameraman	City	Windmill	Lake
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.385	22.13/0.456
Patches with <i>disk</i> shapes (family F1)	29.45/0.832	28.16/0.885	30.97/0.904	28.68/0.863
Patches with <i>half-pie</i> shapes (family F2)	29.43/0.832	28.08/0.886	30.97/0.906	28.60/0.863
Patches with <i>quarter-pie</i> shapes (family F3)	29.31/0.831	27.87/0.883	30.95/0.909	28.49/0.862
Patches with <i>band</i> shapes (family F4)	29.46/0.832	28.05/0.885	31.05/0.906	28.61/0.862
Combination of shape families F1 and F2	<b>29.50/0.833</b>	<b>28.21/0.887</b>	31.11/0.907	<b>28.73/0.865</b>
Combination of shape families F1, F2, F3 and F4	<b>29.50/0.833</b>	28.20/0.887	<b>31.19/0.909</b>	28.72/0.865



**Fig. 8** Eight denoised images obtained for different oriented patch shapes. The proposed final aggregate is in the center. Each denoiser provides good performances in a specific target direction but suffers from *noise halos* in the other directions. The final (central) aggregate takes advantage of every oriented-denoiser to provide high quality restored edges. The patch shape used is indicated in white.

slightly the PSNR. Visually speaking, using only small isotropic patches already decreases the *rare patch effect* while using too large isotropic shapes produces a strong *noise halo* effect. However, the level of noise is much more decreased by using large shapes than small ones in homogeneous areas. Finally, combining different scales of isotropic patch shapes leads to a diminution of both the level of noise and the halo effect. This is well reflected in Tab. 1 by the gain in term of the SSIM criterion which provides quality measurements closer to our perception system. We will see in the following that the results can still be improved by considering both multi-scale and anisotropic patch shapes.

Table 2 gives numerical results obtained by using different families. The compared families are the



**Fig. 9** Illustration of the diminution of the *rare patch effect*. (a) The noisy image with an highlighted target pixel and its neighborhood. (b) The associated maps of weights obtained by using only square patches of fixed size (i.e., NL-Means). (c) The associated maps of weights aggregating multi-scale and anisotropic patch shapes (i.e., NLM-SAP). Low weights are black, high ones are white.

ones presented on Fig. 2.a, i.e., the *disks*, the *half-pies*, the *quarter-pies* and the *bands*. Combination of these families are also studied. Our experiments show that most suitable shape families, both in terms of PSNR and SSIM, have to contain isotropic shapes, directional shapes and various scales of shapes. Increasing the number of shapes does not necessarily improve the denoising quality. Using 15 shapes from families F1 and F2 with the three different scales, seems to be a good trade-off between computing time and denoising quality. Figure 8 illustrates why using directional shapes is important to reduce the *rare patch effect*. Indeed, each oriented patch shape enables the restoration of edges in the target direction but is inappropriate in the other directions. Then a fine aggregation of them leads to high quality restoration of edges in all directions. Figure 9 displays weight maps induced by using patches with only one fixed square shape (i.e., NL-Means) compared to patches with adaptive scales and orientations (i.e., NLM-SAP). For NL-Means, all the weights are concentrated around the target pixel: this is the *rare patch effect*. For NLM-SAP, the weights are more spread, and other similar pixels are detected thanks to multi-scale and anisotropic patch shapes. It is clear that the limitation of the *rare patch effect* around high contrasted edges leads to a good reduction of the *noise halo*.

**Table 3** Comparisons of different aggregation procedures in terms of PSNR and SSIM values (PSNR/SSIM). The compared aggregation types are UWA, WAV, MRE and EWA.

	Cameraman	City	Windmill	Lake
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.385	22.13/0.456
Uniform Weighted Average (UWA)	29.40/0.830	27.99/0.880	30.76/0.897	28.53/0.858
Weighted Average based on Variance (WAV)	29.46/0.830	27.98/0.879	30.82/0.898	28.48/0.856
Minimizer of the Risk Estimates (MRE)	29.33/0.829	28.02/0.885	30.88/0.905	28.58/0.862
Exponentially Weighted Average (EWA)	<b>29.50/0.833</b>	<b>28.21/0.887</b>	<b>31.11/0.907</b>	<b>28.73/0.865</b>

**Table 4** Comparisons of regularization procedures of the risk maps in terms of PSNR and SSIM values (PSNR/SSIM). The compared regularization procedures are the ones using the noisy risk maps directly (i.e., SURE maps), the convoluted risk maps and the risk maps obtained by anisotropic diffusions.

	Cameraman	City	Windmill	Lake
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.385	22.13/0.456
Noisy risk maps (SURE maps)	29.04/0.816	27.37/0.865	30.13/0.873	28.33/0.846
Convoluted risk maps	<b>29.59/0.836</b>	<b>28.36/0.891</b>	<b>31.23/0.909</b>	<b>28.85/0.869</b>
Anisotropic diffusions based risk maps	29.50/0.833	28.21/0.887	31.11/0.907	28.73/0.865
True risk maps (MSE maps provided by an oracle)	32.09/0.881	32.32/0.938	34.44/0.936	32.28/0.923

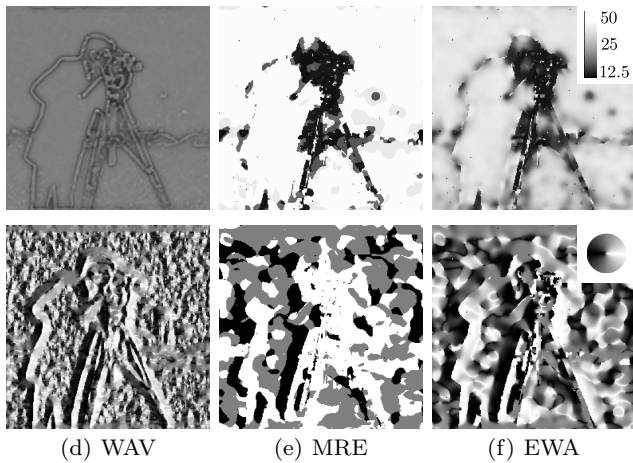
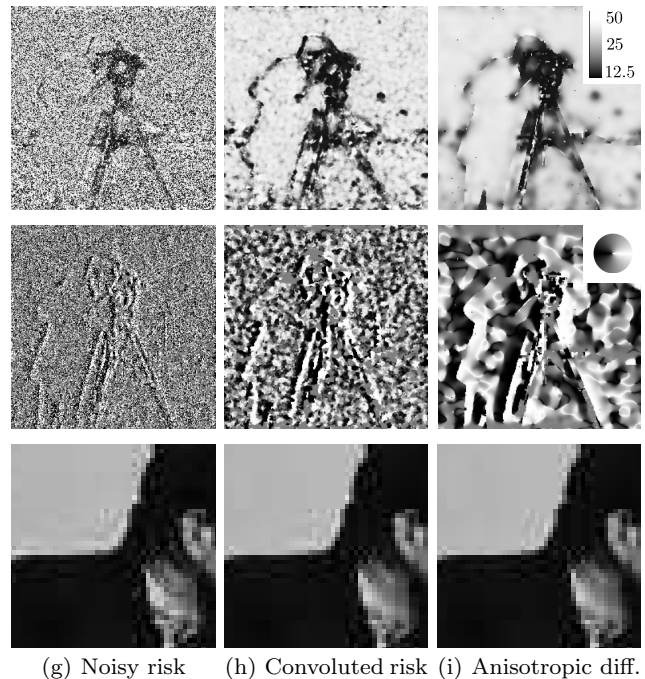
**Fig. 10** (top) Average areas and (bottom) average orientations of selected shapes for different aggregation procedures on a noisy realization of the *cameraman* image. From left to right: WAV, MRE and EWA aggregations. The average areas and the average orientations are represented using gray level colors whose legends are given on the top right corners.

Table 3 presents the numerical performances associated with four aggregation procedures: UWA, WAV, MRE and EWA. As expected, EWA provides best results, in terms of PSNR and SSIM, since compared to the other three it combines estimates with the best bias-variance trade-off. The local behaviors of NLM-SAP for WAV, MRE and EWA are presented on Fig. 10. The average areas and the average orientations of the selected shapes are given for the *cameraman* image. The selected patch shapes clearly adapt to the local scale and orientation of the image geometry. The chosen sizes of the shapes are smaller around edges and textured areas than in homogeneous areas. The chosen orientation follows the orthogonal orientation of the shapes which

**Fig. 11** (top) Average areas and (middle line) average orientations of selected shapes for different risk maps on a noisy realization of the *cameraman* image. (bottom) Corresponding results focused on the cameraman's neck. From left to right, results using the noisy risk maps (i.e., SURE maps), the convoluted risk maps and the risk maps obtained by anisotropic diffusions. The average areas and the average orientations are represented using gray level colors whose legends are given on the top right corners.

is consistent with the remarks given in Section 3. Compared to EWA, MRE suffers from brutal transitions, since it selects only one shape per pixel, while EWA evolves in a smoother way due to the weighted combination of shapes for each pixel.

Finally, we have studied the influence of the regularization of the risk maps on the aggregation results. Three methodologies are compared: aggregation using the noisy risk maps (i.e., SURE maps), the convoluted risk maps (using a disk kernel of radius 4) and the risk maps obtained by anisotropic diffusions. Table 4 gives the corresponding numerical performances and Fig. 11 illustrates the behavior of each type of risk map regularization. Surprisingly, the risk maps based on isotropic diffusions of SURE maps provide the best results in terms of PSNR and SSIM. However, the choice of the local sizes and orientations of the patch shapes is more relevant in the maps obtained by anisotropic diffusions, in terms of scale adaptivity, feature directions and spatial coherency. Using anisotropic diffusion, the NLM-SAP acts as expected by selecting big sizes of shapes, even around edges, since the shape orientations have been chosen properly to reduce the *rare patch effect*. By comparison, isotropic diffusion forces the size of shapes to be small around edges since it cannot select properly the suitable orientations. This difference of behaviors can be noticed around the cameraman’s neck. In the case of the convolution, a small halo is present due to wrong decisions while for anisotropic diffusion this halo is strongly reduced. Then, even if isotropic diffusion provides the best numerical results in practice, anisotropic regularizations seems better theoretically founded to resolve this issue (see also Section 4.3). Other regularization strategies could then be investigated. To support this idea, Tab. 4 shows that there is still a gap of numerical performances between regularizations of the risk maps and the true risk maps (i.e., the local mean square errors provided by an oracle).

## 5.2 Comparisons with *state-of-the-art* methods

In this section, the proposed NLM-SAP approach is compared to *state-of-the-art* denoising methods. Comparisons have been performed with the classical (pixelwise) NL-Means [5], the blockwise NL-Means using UWA reprojection [5], the blockwise NL-Means using WAV reprojection [43], the pixelwise NL-means using SURE-based adaptive bandwidth selection [16], the Block-Matching and 3D filtering (BM3D) denoiser [7], and our proposed NLM-SAP approach.

Table 5 shows that NLM-SAP out-performs all other NL-Means improvements. NLM-SAP brings a gain of PSNR of about 1 dB compared to the classical NL-Means. The SSIM is also usually increased. The BM3D approach leads to better numerical results than all Non Local based approaches. Figure 12 gives the visual results. While the presence of the *rare patch*

*effect* is very well illustrated by the *noise halos* for NL-means and blockwise NL-Means using UWA aggregation, BM3D and NLM-SAP have reduced a lot this phenomenon. Our NLM-SAP provides smooth results with accurate details, such as the cameraman’s head, the house windows, the windmill blades, the tree-trunk and the car. Visually, the quality of images obtained with NLM-SAP challenges those obtained with BM3D.

## 6 Conclusion

In this paper, we have addressed the problem of the *rare patch effect* arising in the NL-Means procedure and responsible of the noisy halos created around high contrasted edges. The proposed solution consists in substituting the square patches of fixed size by spatially adaptive patch shapes. A fast implementation of NL-Means, based on FFT calculations, has been proposed in this context to handle any kind of patch shape with arbitrary scale. Thanks to this acceleration, different estimates are obtained by using different patch shapes, typically one isotropic patch shape and four edge oriented patch shapes, all of them with three different scales. We have extended SURE-based approaches to aggregate properly these different shape-based estimates in a spatially adaptive way. To get an efficient locally adaptive filter, we have shown that the SURE-based risk maps require to be regularized and that anisotropic diffusion can be used to this purpose. Simulations have shown that exponentially weighted aggregation based on the regularized risk maps of the different shape-based estimates could lead to both numerical and visual improvements (the *noise halo* is suppressed around edges). Our method out-performs all the NL-Means improvements we have considered in our comparisons but is still out-performed by BM3D in terms of PSNR and SSIM.

Future work could be to reduce computation time of the algorithm by choosing more convenient shapes or parallel implementations. Other method to regularize the risk maps could also be considered. Another extension might be to apply our framework to more general kinds of noise, by adapting results given by Hudson [21] on unbiased risk estimates for general exponential families.

**Acknowledgements** The authors would like to thanks Jean-François Aujol, Loïc Denis, Yann Gousseau, Erwan Le Pennec and Florence Tupin for their interesting comments and criticisms.

**Table 5** Comparisons of denoising approaches for different degradation levels in terms of PSNR and SSIM values (PSNR/SSIM). The compared methods are the classical (pixelwise) NL-Means [5], the blockwise NL-Means using UWA reprojection [5], the blockwise NL-Means using WAV reprojection [43], the pixelwise NL-means using SURE-based adaptive bandwidth selection [16], the BM3D denoiser [7], and our proposed NLM-SAP approach.

	Camerman	City	Windmill	Lake
$\sigma = 5$				
Noisy input image	34.17/0.845	34.17/0.901	34.17/0.827	34.17/0.866
NL-Means [5]	36.92/0.951	35.87/0.965	38.10/0.972	36.76/0.964
UWA Blockwise NL-Means [5]	37.19/0.953	36.19/0.967	38.50/0.974	37.14/0.967
WAV Blockwise NL-Means [43]	37.27/0.954	36.37/0.970	38.75/0.977	37.03/0.969
SURE-based adaptive NL-Means [16]	37.46/0.956	36.76/0.975	39.14/0.978	37.28/0.970
BM3D [7]	<b>38.17/0.962</b>	<b>37.48/0.978</b>	<b>39.91/0.983</b>	<b>38.15/0.977</b>
NLM-SAP	37.80/0.957	37.26/0.975	39.60/0.979	37.92/0.974
$\sigma = 10$				
Noisy input image	28.15/0.633	28.15/0.757	28.15/0.607	28.15/0.678
NL-Means [5]	32.46/0.905	31.11/0.932	33.62/0.945	32.07/0.926
UWA Blockwise NL-Means [5]	32.80/0.908	31.50/0.934	34.06/0.946	32.47/0.930
WAV Blockwise NL-Means [43]	32.70/0.908	31.37/0.936	33.91/0.950	31.90/0.926
SURE-based adaptive NL-Means [16]	33.11/0.918	32.11/0.948	34.78/0.954	32.61/0.935
BM3D [7]	<b>34.06/0.931</b>	<b>33.15/0.956</b>	<b>35.84/0.966</b>	<b>33.63/0.950</b>
NLM-SAP	33.44/0.914	32.84/0.950	35.28/0.955	33.27/0.940
$\sigma = 20$				
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.385	22.13/0.456
NL-Means [5]	28.72/0.820	27.11/0.870	30.04/0.897	28.12/0.855
UWA Blockwise NL-Means [5]	28.89/0.822	27.34/0.872	30.17/0.899	28.21/0.858
WAV Blockwise NL-Means [43]	28.49/0.820	26.58/0.855	29.48/0.898	27.31/0.842
SURE-based adaptive NL-Means [16]	29.49/0.845	27.85/0.889	30.96/0.906	28.46/0.867
BM3D [7]	<b>30.35/0.871</b>	<b>29.07/0.912</b>	<b>32.07/0.936</b>	<b>29.38/0.895</b>
NLM-SAP	29.50/0.833	28.21/0.887	31.11/0.907	28.73/0.865
$\sigma = 40$				
Noisy input image	16.11/0.222	16.11/0.364	16.11/0.222	16.11/0.259
NL-Means [5]	25.26/0.724	22.37/0.692	25.52/0.793	23.81/0.706
UWA Blockwise NL-Means [5]	25.05/0.726	22.29/0.693	25.29/0.795	23.70/0.708
WAV Blockwise NL-Means [43]	24.66/0.720	21.95/0.680	25.05/0.792	23.51/0.703
SURE-based adaptive NL-Means [16]	26.20/0.741	23.77/0.759	26.95/0.808	24.94/0.741
BM3D [7]	<b>27.26/0.802</b>	<b>25.25/0.830</b>	<b>28.05/0.879</b>	<b>25.86/0.805</b>
NLM-SAP	25.92/0.718	23.30/0.720	26.26/0.776	24.58/0.714

## References

- Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006)
- Awate, S.P., Whitaker, R.T.: Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(3), 364–376 (2006)
- Blu, T., Luisier, F.: The SURE-LET approach to image denoising. *IEEE Trans. Image Process.* **16**(11), 2778–2786 (2007)
- Brox, T., Kleinschmidt, O., Cremers, D.: Efficient nonlocal means for denoising of textural patterns. *IEEE Trans. Image Process.* **17**(7), 1083–1092 (2008)
- Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* **4**(2), 490–530 (2005)
- Crow, F.C.: Summed-area tables for texture mapping. In: *SIGGRAPH Comput. Graph.*, pp. 207–212 (1984)
- Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.O.: Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007)
- Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.O.: BM3D image denoising with shape-adaptive principal component analysis. In: *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS’09)* (2009)
- Dalalyan, A.S., Tsybakov, A.B.: Aggregation by exponential weighting, sharp oracle inequalities and sparsity. In: *COLT*, pp. 97–111 (2007)
- Dalalyan, A.S., Tsybakov, A.B.: Aggregation by exponential weighting, sharp pac-bayesian bounds and sparsity. *Mach. Learn.* **72**(1-2), 39–61 (2008)
- Darbon, J., Cunha, A., Chan, T.F., Osher, S., Jensen, G.J.: Fast nonlocal filtering applied to electron cryomicroscopy. In: *ISBI*, pp. 1331–1334 (2008)
- Deledalle, C.A., Denis, L., Tupin, F.: Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Trans. Image Process.* **18**(12), 2661–2672 (2009)
- Deledalle, C.A., Denis, L., Tupin, F.: NL-InSAR: Nonlocal interferogram estimation. *IEEE Trans. Geosci. Remote Sens.* **49**(4) (2011)
- Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90**(432), 1200–1224 (1995)
- Doré, V., Cheriet, M.: Robust NL-Means Filter With Optimal Pixel-Wise Smoothing Parameter for Statisti-





**Fig. 12** Comparisons of the visual denoising performances of the proposed NLM-SAP approach and other *state-of-the-art* methodologies. From top to bottom, the input noisy images, the results obtained by the classical (pixelwise) NL-Means [5], the blockwise NL-Means using UWA reprojection [5], the BM3D denoiser [7], and our proposed NLM-SAP approach.

- cal Image Denoising. *IEEE Trans. Signal Process.* **57**, 1703–1716 (2009)
16. Duval, V., Aujol, J.F., Gousseau, Y.: On the parameter choice for the non-local means. Tech. Rep. hal-00468856, HAL (2010)
  17. Foi, A., Katkovnik, V., Egiazarian, K.: Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images. *IEEE Trans. Image Process.* **16**(5), 1395–1411 (2007)
  18. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. *Multiscale Model. Simul.* **7**(3), 1005–1028 (2008)
  19. Goldenshluger, A., Nemirovski, A.S.: On spatially adaptive estimation of nonparametric regression. *Math. Methods Statist.* **6**(2), 135–170 (1997)
  20. Goossens, B., Luong, H.Q., Pizurica, A., Philips, W.: An improved non-local denoising algorithm. In: LNLA, pp. 143–156 (2008)
  21. Hudson, H.M.: A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6**(3), 473–484 (1978)
  22. Katkovnik, V., Foi, A., Egiazarian, K.O., Astola, J.T.: Directional varying scale approximations for anisotropic signal processing. In: EUSIPCO, pp. 101–104 (2004)
  23. Kervrann, C., Boulanger, J.: Optimal spatial adaptation for patch-based image denoising. *IEEE Trans. Image Process.* **15**(10), 2866–2878 (2006)
  24. Kervrann, C., Boulanger, J., Coupé, P.: Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In: SSVM, vol. 4485, pp. 520–532 (2007)
  25. Le Pennec, E., Mallat, S.: Sparse geometric image representations with bandelets. *IEEE Trans. Image Process.* **14**(4), 423–438 (2005)
  26. Lee, J.S.: Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing* **24**(2), 255–269 (1983)
  27. Lepski, O.V., Mammen, E., Spokoiny, V.G.: Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25**(3), 929–947 (1997)
  28. Leung, G., Barron, A.R.: Information theory and mixing least-squares regressions. *IEEE Trans. Inf. Theory* **52**(8), 3396–3410 (2006)
  29. Li, K.C.: From Stein’s unbiased risk estimates to the method of generalized cross validation. *Ann. Statist.* **13**(4), 1352–1377 (1985)
  30. Louchet, C., Moisan, L.: Total variation as a local filter. to appear (2010)
  31. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. *ICCV* (2009)
  32. Mairal, J., Sapiro, G., Elad, M.: Learning multiscale sparse representations for image and video restoration. *Multiscale Model. Simul.* **7**(1), 214–241 (2008)
  33. Mallows, C.L.: Some comments on  $c_p$ . *Technometrics* **15**(4), 661–675 (1973)
  34. Nemirovski, A.S.: Topics in non-parametric statistics, *Lecture Notes in Math.*, vol. 1738. Springer, Berlin (2000)
  35. Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 629–639 (1990)
  36. Polzehl, J., Spokoiny, V.G.: Adaptive weights smoothing with applications to image restoration. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62**(2), 335–354 (2000)
  37. Polzehl, J., Spokoiny, V.G.: Propagation-separation approach for local likelihood estimation. *Probab. Theory Related Fields* **135**(3), 335–362 (2006)
  38. Portilla, J., Strela, V., Wainwright, M., Simoncelli, E.P.: Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Image Process.* **12**(11), 1338–1351 (2003)
  39. Ramani, S., Blu, T., Unser, M.: Monte-Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms. *IEEE Trans. Image Process.* **17**(9), 1540–1554 (2008)
  40. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D* **60**(1-4), 259–268 (1992)
  41. Salmon, J.: On two parameters for denoising with Non-Local Means. *IEEE Signal Process. Lett.* **17**, 269–272 (2010)
  42. Salmon, J., Le Pennec, E.: NL-Means and aggregation procedures. In: ICIP, pp. 2977–2980 (2009)
  43. Salmon, J., Strozeccki, Y.: From patches to pixels in non-local methods: Weighted-Average reprojection. In: ICIP (2010)
  44. Solo, V.: A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems. In: ICIP, vol. 3, pp. 89–92 (1996)
  45. Starck, J.L., Candès, E.J., Donoho, D.L.: The curvelet transform for image denoising. *IEEE Trans. Image Process.* **11**(6), 670–684 (2002)
  46. Stein, C.M.: Estimation of the mean of a multivariate distribution. In: Proc. Prague Symp. Asymptotic Statist. (1973)
  47. Stein, C.M.: Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9**(6), 1135–1151 (1981)
  48. Strong, D., Chan, T.: Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems* **19**(6), S165–S187 (2003). Special section on imaging
  49. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: ICCV, pp. 839–846 (1998)
  50. Tsybakov, A.B.: Optimal rates of aggregation. In: COLT, pp. 303–313 (2003)
  51. Van De Ville, D., Kocher, M.: SURE-based Non-Local Means. *IEEE Signal Process. Lett.* **16**, 973–976 (2009)
  52. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. *CVPR* **1**, 511–518 (2001)
  53. Wang, J., Guo, Y.W., Ying, Y., Liu, Y.L., Peng, Q.S.: Fast non-local algorithm for image denoising. In: ICIP, pp. 1429–1432 (2006)
  54. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Signal Process.* **13**(4), 600–612 (2004)
  55. Wasserman, L.: All of Nonparametric Statistics (Springer Texts in Statistics). Springer (2007)
  56. Yaroslavsky, L.P.: Digital picture processing, *Springer Series in Information Sciences*, vol. 9. Springer-Verlag, Berlin (1985)