# An Optimized Blockwise Non Local Means Denoising Filter for 3D Magnetic Resonance Images

Pierrick Coupé[1,2,4], Pierre Yger[1,2,4,6], Sylvain Prima[1,2,4], Pierre Hellier[1,2,4], Charles Kervrann[3,5] and Christian Barillot[1,2,4]

[1] University of Rennes I - CNRS UMR 6074, IRISA, Campus de Beaulieu, F-35042 Rennes,France
[2] INRIA, VisAGeS U746 Unit/Project, IRISA, Campus de Beaulieu, F-35042 Rennes, France
[3] INRA, UR341 Mathematiques et Informatique Appliquées, F-78352 Jouy en Josas, France
[4] INSERM, VisAGeS U746 Unit/Project, IRISA, Campus de Beaulieu, F-35042 Rennes, France
[5] INRIA, VISTA Project-team, IRISA, Campus de Beaulieu, F-35042 Rennes, France
[6] ENS, 61 avenue du Président Wilson 94235 Cachan, France

*Abstract*— A critical issue in image restoration is the problem of noise removal while keeping the integrity of relevant image information. Denoising is a crucial step to increase image quality and to improve the performance of all the tasks needed for quantitative imaging analysis. The method proposed in this paper is based on a 3D optimized blockwise version of the Non Local (NL) means filter [1]. The NL-means filter uses the redundancy of information in the image under study to remove the noise. The performance of the NL-means filter has been already demonstrated for 2D images, but reducing the computational burden is a critical aspect to extend the method to 3D images. To overcome this problem, we propose improvements to reduce the computational complexity. These different improvements allow to drastically divide the computational time while preserving the performances of the NL-means filter. A fully-automated and optimized version of the NL-means filter is then presented. Our contributions to the NL-means filter are: (a) an automatic tuning of the smoothing parameter, (b) a selection of the most relevant voxels, (c) a blockwise implementation and (d) a parallelized computation. Quantitative validation was carried out on synthetic datasets generated with BrainWeb [2]. The results show that our optimized NL-means filter outperforms the classical implementation of the NL-means filter, as well as two other classical denoising methods (Anisotropic Diffusion [3] and Total Variation minimization process [4]) in terms of accuracy (measured by the Peak Signal to Noise Ratio) with low computation time. Finally, qualitative results on real data are presented.

## I. INTRODUCTION

Quantitative imaging involves image processing workflows (registration, segmentation, visualization, *etc.*) with increasing complexity and sensitivity to possible image artifacts. As a consequence, image processing procedures often require to remove image artifacts beforehand in order to make quantitative post-processing more robust and efficient. A critical issue concerns the problem of noise removal while keeping the integrity of relevant image information. This is particularly true for ultrasound images or magnetic resonance images (MRI) in presence of small structures with signals barely detectable above the noise level. In addition, a constant evolution of quantitative medical imaging is to process always larger cohorts of 3D data in order to find significant discriminants for a given pathology (e.g. see [5]). In this context, complex automatic image processing workflows are required [6] since human interpretation of images is no longer feasible. For effectiveness, these workflows have to be robust to a wide range of image qualities and parameter-free (or at least using auto-tuned parameters). This paper focuses on these critical aspects by introducing a new restoration scheme in the context of 3D medical imaging. The Non Local (NL-) means filter was originally introduced by Buades *et al.* [1] for 2D image denoising. The adaptation of this filter we propose for 3D images is based on (a) an automatic tuning of the smoothing parameter, (b) a selection of the most relevant voxels for the NL-means computation, (c) a blockwise implementation and (d) a parallelized computation. These different contributions allow to make the adapted filter fully-automated and above all to overcome the main limitation of the classical NL-means: the computational burden.

Section II gives a short overview of the literature on image restoration. Section III presents the proposed method with details about our contributions. Sections IV, V and VI show (a) the impact of our adaptations compared to the classical NL-means implementation and (b) a comparison with respect to other well established denoising methods on Gaussian and Rician noise. Both validation experiments are performed on a phantom data set in a quantitative way. Section VII shows results on real data such as 3 Tesla T1-weighted (T1-w) MRI and T2-weighted (T2-w) MRI of a patient with Multiple Sclerosis (MS) lesions. In section VIII we propose a discussion on the applicability and the further improvements of the NL-means filter in the context of 3D medical imaging.

## II. STATE-OF-THE-ART

### A. General overview

Many methods have been proposed for edge-preserving image denoising. Some popular approaches include Bayesian approaches [7], PDE-based approaches [3], [4], [8], [9], robust and regression estimation [10], adaptive smoothing [11],

wavelet-based methods [12]–[14], bilateral filtering [15]–[17], local mode filtering [18], hybrid approaches [19]–[21].

Strong theoretical links exist between most of these techniques, as recently shown for local mode filtering [18], bilateral filtering, anisotropic diffusion and robust estimation [17], [22] and adaptive smoothing [23], anisotropic diffusion and total variation minimization scheme [24].

More recently, some promising methods have been proposed for improved image denoising, based on statistical averaging schemes enhanced via incorporating a variable spatial neighborhood scheme [25]–[29]. Other approaches consist in modeling non-local pairwise interactions from training data [30] or a library of natural image patches [31], [32]. 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 [30]–[32]. Awate and Whitaker proposed another non-parametric patch-based method relying on comparisons between probability density functions [33].

Some of these techniques, generally developed for 2D images, have often been extended to 3D medical data, especially to MR images: anisotropic diffusion [34], [35], total variation [36], bilateral filtering and variants [37], wavelet-based filtering [38]–[42], hybrid approaches [43], [44].

### B. Introduction of the NL-means filter

Most of the denoising methods restore the intensity value of each image voxel by averaging in some way the intensities of its (spatially) neighboring voxels. The basic and intuitive approach is to replace the value of the voxel by the average of the voxels in its neigbourhood (so-called box filtering [45]). In practice, this filter has been shown to be outperformed by the Gaussian filter, which consists in weighting each voxel in the neighborhood according to its distance to the voxel under study. Both filters can be iterated until the desired amount of smoothing is reached. Such *data-independent* approaches can be implemented very efficiently. Their major drawback is that they blur the structures of interest in the image (e.g. edges or small structures and textures).

This has naturally led to *data-dependent* approaches, which aim at eliminating (or reducing the influence of) the neighboring voxels dissimilar to the voxel under study. Simple order statistic operators can be used for this purpose, such as the median filter, leading to a simple generalization of the box filter. More sophisticated approaches, based on image derivatives have been successfully proposed for many applications, such as adaptive smoothing [11] and anisotropic diffusion [3]. Neighborhood filters [46], [47] and variants [15], [16], have been also proposed and consist in averaging input data over the image voxels that are spatially close to the voxel under study and with similar gray-level values.

All these techniques rely on the idea that the restored value of a voxel should only depend on the voxels in its spatial neighborhood that belong to the same *population*, that is the same image context. This has been termed by Michael Elad as the *locally adaptive recovery paradigm* [17]. Another approach has been recently proposed, that has shown very promising

results. It is based on the idea that any natural image has redundancy, and that any voxel of the image has similar voxels that are not necessarily located in a spatial neighborhood. First introduced by Buades *et al.* in [1], the NL-means filter is based on this redundancy property of periodic images, textured images or natural images to remove noise. In this approach, the weight involving voxels in the average, does not depend on their spatial proximity to the current voxel but is based on the intensity similarity of their neighborhoods with the neighborhood of the voxel under study, as in patched-based approaches. In other words, the NL-means filter can be viewed as an extreme case of neighborhood filters with infinite spatial kernel and where the similarity of the neighborhood intensities is substituted to the point-wise similarity of gray levels as in commonly-used bilateral filtering. This new *non-local recovery paradigm* allows to combine the two most important attributes of a denoising algorithm: edge preservation and noise removal.

## III. METHODS

In the following, we introduce the notations:

- $u : \Omega^3 \longmapsto \mathbb{R}$ is the image, where $\Omega^3$ represents the grid of the image, considered as cubic for the sake of simplicity and without loss of generality ($|\Omega^3| = N^3$).
- for the original voxelwise NL-means approach
  - $u(x_i)$ is the intensity observed at voxel $x_i$.
  - $V_i$ is the cubic search volume centered on voxel $x_i$ of size $|V_i| = (2M + 1)^3$, $M \in \mathbb{N}$.
  - $N_i$ is the cubic local neighborhood of $x_i$ of size $|N_i| = (2d + 1)^3$, $d \in \mathbb{N}$.
  - $\mathbf{u}(N_i) = (u^{(1)}(N_i), ..., u^{(|N_i|)}(N_i))^T$ is the vector containing the intensities of $N_i$.
  - $NL(u)(x_i)$ is the restored value of voxel $x_i$.
  - $w(x_i, x_j)$ is the weight of voxel $x_j$ when restoring $u(x_i)$.
- for the blockwise NL-means approach
  - $B_i$ is the block centered on $x_i$ of size $|B_i| = (2\alpha + 1)^3$, $\alpha \in \mathbb{N}$.
  - $\mathbf{u}(B_i)$ is the vector containing the intensities of the block $B_i$.
  - $\mathbf{NL}(u)(B_i)$ is the vector containing the restored value of $B_i$.
  - $w(B_i, B_j)$ is the weight of block $B_j$ when restoring the block $\mathbf{u}(B_i)$.
  - the blocks $B_{i_k}$ are centered on voxels $x_{i_k}$ with $i_k = (k_1 n, k_2 n, k_3 n), (k_1, k_2, k_3) \in \mathbb{N}^3$ and $n$ represents the distance between the centers of the blocks $B_{i_k}$.

### A. The Non Local means filter

In the classical formulation of the NL-means filter, the restored intensity $NL(u)(x_i)$ of the voxel $x_i$, is the weighted average of all the voxel intensities in the image $u$ defined as:

$$NL(u)(x_i) = \sum_{x_j \in \Omega^3} w(x_i, x_j) u(x_j) \qquad (1)$$

where $u(x_j)$ is the intensity of voxel $x_j$ and $w(x_i, x_j)$ is the weight assigned to $u(x_j)$ in the restoration of voxel $x_i$.

More precisely, the weight quantifies the similarity of the local neighborhoods $N_i$ and $N_j$ of the voxels $x_i$ and $x_j$ under the assumptions that $w(x_i, x_j) \in [0,1]$ and $\sum_{x_j \in \Omega^3} w(x_i, x_j) = 1$ (cf Fig. 1 left). The classical definition of the NL-means filter considers that each voxel can be linked to all the others, but for practical computational reasons the number of voxels taken into account in the weighted average can be limited to the so-called "search volume" $V_i$ of size $(2M+1)^3$, centered at the current voxel $x_i$.

For each voxel $x_j$ in $V_i$, the Gaussian-weighted Euclidean distance $\|.\|_{2,a}^2$ defined in [1], is computed between $\mathbf{u}(N_j)$ and $\mathbf{u}(N_i)$. This distance is a classical $L_2$ norm convolved with a Gaussian kernel of standard deviation $a$, and measures the distance between neighborhood intensities. Given this distance, $w(x_i, x_j)$ is computed as follows:

$$w(x_i, x_j) = \frac{1}{Z_i} e^{-\frac{\|\mathbf{u}(N_i) - \mathbf{u}(N_j)\|_{2,a}^2}{h^2}} \qquad (2)$$

where $Z_i$ is a normalization constant ensuring that $\sum_j w(x_i, x_j) = 1$, and $h$ acts as a smoothing parameter controlling the decay of the exponential function. When $h$ is very high, all the voxels $x_j$ in $V_i$ will have the same weight $w(x_i, x_j)$ with respect to the voxel $x_i$. The restored value $NL(u)(x_i)$ will be then approximately the average of the intensity values of the voxels in $V_i$ leading to strong smoothing of the image. When $h$ is very low, the decay of the exponential function will be strong, thus only few voxels $x_j$ in $V_i$ with $u(N_j)$ very similar to $u(N_j)$ will have a significant weight. The restored value $NL(u)(x_i)$ will tend to be the weighted average of some voxels with a similar neighborhood to current voxel $x_i$ leading to a weak smoothing of the image. In Section III-B.1, a trade-off has then to be found, and we propose a method to automatically estimate the optimal value of $h$.

In [1], Buades *et al.* show that, for 2D natural images, the NL-means filter outperforms state-of-the-art denoising methods such as the Rudin-Osher-Fatemi Total Variation minimization scheme [4], the Perona-Malik Anisotropic diffusion [3] or translation invariant wavelet thresholding [48]. Nevertheless, the main drawback of the NL-means filter is the computational burden due to its complexity, especially for 3D images. Indeed, for each voxel of the volume, distances between intensity neighborhoods $\mathbf{u}(N_i)$ and $\mathbf{u}(N_j)$ for all the voxels $x_j$ contained in $V_i$ need to be computed. Let $N^3$ denote the size of the 3D image, then the complexity of the filter is in the order of $\mathcal{O}((N(2M+1)(2d+1))^3)$. For a 3D MRI of size $181 \times 217 \times 181$ with the smallest possible value for $d$ and $M$ ($d = 1$ and $M = 5$), the computational time reaches up to 6 hours on 3 GHz CPU. This time is far beyond a reasonable duration expected for a denoising filter in a medical practice. For this reason, we propose several adaptations to reduce the computational burden which are detailed in Section III-B. We also show that these adaptations improve the quality of the denoising compared to the classical implementation.

### B. Improvements of the NL-means filter

*1) Automatic tuning of the Smoothing parameter h:* According to [1], the smoothing parameter $h$ depends on the standard deviation of the noise $\sigma$, and typically a good choice for 2D images is $h \approx 10\sigma$. Equation 2 shows that $h$ also needs to take into account $|N_i|$, if we want a filter independent of the neighborhood size. Indeed, the $L_2$ norm increasing with $|N_i|$, $h$ needs also to be increased to obtain an equivalent filter. The automatic tuning of the smoothing parameter $h$ comes to determine the relationship $h^2 = f(\sigma^2, |N_i|, \beta)$ where $\beta$ is a constant. Let us show how we can estimate this relationship:

(a) In case of an additive white Gaussian noise, the standard deviation of noise can be estimated *via* pseudo-residuals $\epsilon_i$ as defined in [49], [50]. For each voxel $x_i$ of the volume $\Omega^3$, let us define:

$$\epsilon_i = \sqrt{\frac{6}{7}} \left( u(x_i) - \frac{1}{6} \sum_{x_j \in P_i} u(x_j) \right) \qquad (3)$$

$P_i$ being the 6-neighborhood at voxel $x_i$ and the constant $\sqrt{6/7}$ is used to ensure that $\mathbb{E}[\epsilon_i^2] = \hat{\sigma}^2$ in homogeneous areas. Thus, the standard deviation of noise $\hat{\sigma}$ is computed as the least square estimator:

$$\hat{\sigma}^2 = \frac{1}{|\Omega^3|} \sum_{i \in \Omega^3} \epsilon_i^2 \qquad (4)$$

(b) Initially, the NL-means filter was defined with a Gaussian-weighted Euclidean distance, $\|.\|_{2,a}^2$ defined in [1]. However, in order to make the filter independent of $|N_i|$, to simplify the complexity of the problem, and to reduce the computational time, we used the classical Euclidean distance $\|.\|_2^2$ normalized by the number of elements:

$$\frac{1}{|N_i|} \|\mathbf{u}(N_i) - \mathbf{u}(N_j)\|_2^2 = \frac{1}{|N_i|} \sum_{p=1}^{|N_i|} (u^{(p)}(N_i) - u^{(p)}(N_j))^2. \qquad (5)$$

Finally, Equation 2 becomes:

$$w(x_i, x_j) = \frac{1}{Z_i} e^{-\frac{\|\mathbf{u}(N_i) - \mathbf{u}(N_j)\|_2^2}{2\beta\hat{\sigma}^2 |N_i|}} \qquad (6)$$

where only the adjusting constant $\beta$ needs to be manually tuned. In the case of Gaussian noise, $\beta$ is theoretically be close to 1 (see [51] p. 21 for theoretical justification) if the estimation $\hat{\sigma}$ of the standard deviation of the noise is correct. The adjustment of $\beta$ will be discussed in section V-A.

*2) Voxel selection in the search volume:* To deal with computational burden, Mahmoudi and Sapiro [52] recently proposed a method to preselect a subset of the most relevant voxels $x_j$ in $V_i$ to avoid useless weight computations. In other words, the main idea is to select only the voxels $x_j$ in $V_i$ that will have the highest weights $w(x_i, x_j)$ in Equation 1 without having to compute all the Euclidean distances between $\mathbf{u}(N_i)$ and $\mathbf{u}(N_j)$. A priori neglecting the voxels which are expected to have small weights tends to speed up the filter and to improve the results (see Table II). In [52], this selection is based on the similarity of the mean value of $\mathbf{u}(N_i)$ and $\mathbf{u}(N_j)$, and on the similarity of the average over the neighborhoods $N_i$ and $N_j$ of the gradient orientation at pixel $x_i$ and $x_j$. Intuitively, similar neighborhoods have the same mean and the same gradient orientation. The computation of the gradient orientation is very sensitive to noise and thus requires robust
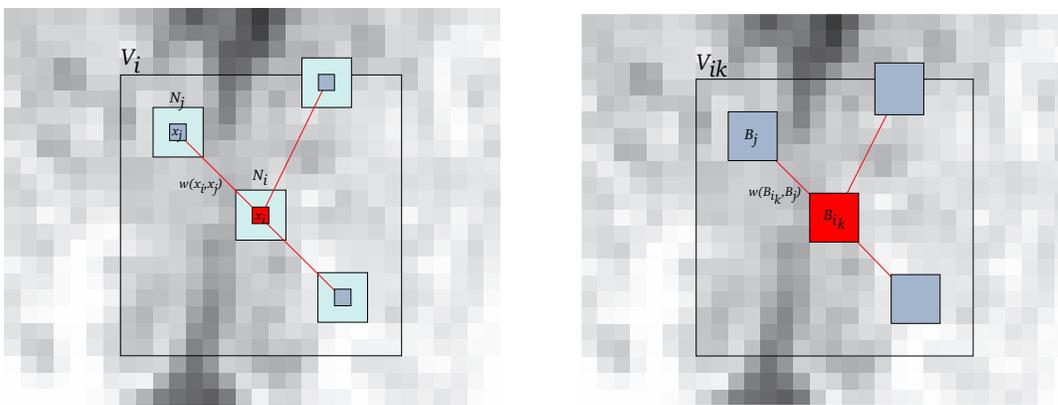
Fig. 1.   Left: **Classical voxelwise NL-means filter:** 2D illustration of the NL-means principle. The restored value of voxel $x_i$ (in red) is the weighted average of all intensities of voxels $x_j$ in the search volume $V_i$, based on the similarity of their intensity neighborhoods $\mathbf{u}(N_i)$ and $\mathbf{u}(N_j)$. In this example, we set $d = 1$ and $M = 8$. Right: **Blockwise NL-means filter:** 2D illustration of the blockwise NL-means principle. The restored value of the block $B_{i_k}$ is the weighted average of all the blocks $B_j$ in the search volume $V_{i_k}$. In this example, we set $\alpha = 1$ and $M = 8$.



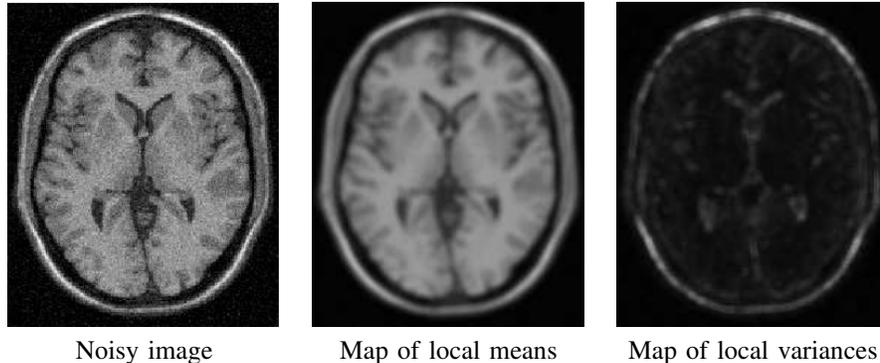|        Noisy image        |        Map of local means        |        Map of local variances        |

Fig. 2.   Left: noisy image with 9 % of Gaussian noise (see Section IV). Center: map of the mean of $\mathbf{u}(N_i)$ denoted $\overline{\mathbf{u}(N_i)}$. Right map of the variance of $\mathbf{u}(N_i)$ denoted $\mathrm{Var}(\mathbf{u}(N_i))$. In these examples, we set $N_i = 5 \times 5 \times 5$ voxels.

estimation techniques. This is too computationally expensive for medical applications. For this reason, in our implementation, the preselection of voxels in $V_i$ is based on the mean and the variance of $\mathbf{u}(N_i)$ and $\mathbf{u}(N_j)$ which allows to decrease the computational burden. Figure 2 shows that the maps of local means and local variances are simple estimators allowing to discriminate different tissue classes and edges in images. In this way, the maps of local means and local variances are precomputed in order to avoid repetitive calculations for the same neighborhood. The selection tests can be expressed as follows:

$$
w(x_i, x_j) = \begin{cases} \dfrac{1}{Z_i} e^{-\frac{\|\mathbf{u}(N_i)-\mathbf{u}(N_j)\|_2^2}{2\beta\hat{\sigma}^2|N_i|}} & \text{if } \mu_1 < \dfrac{\overline{\mathbf{u}(N_i)}}{\overline{\mathbf{u}(N_j)}} < \dfrac{1}{\mu_1} \\ & \text{and } \sigma_1^2 < \dfrac{\mathrm{Var}(\mathbf{u}(N_i))}{\mathrm{Var}(\mathbf{u}(N_j))} < \dfrac{1}{\sigma_1^2} \\ 0 & \text{otherwise.} \end{cases}
$$
(7)

where $\overline{\mathbf{u}(N_i)}$ and $\mathrm{Var}(\mathbf{u}(N_i))$ represents respectively the mean and the variance of the local neighborhood $N_i$ of voxel $x_i$. As suggested in [52], with this kind of selection, the NL-means filter tends to better preserve the detailed regions while slightly spoiling the denoising of the flat regions. Indeed, in flat regions increasing the number of voxels tends to improve denoising because there are a large number of similar voxels. In more cluttered regions, increasing the number of voxels tends to remove the details during smoothing because there are very few similar voxels.

*3) Blockwise implementation:* A blockwise implementation of the NL-means is developed as suggested in [1]. This approach consists in a) dividing the volume into blocks with overlapping supports, b) performing NL-means-like restoration of these blocks and c) restoring the voxels values based on the restored values of the blocks they belong to.

a) A partition of the volume $\Omega^3$ into overlapping blocks $B_{i_k}$ of size $(2\alpha + 1)^3$ is performed, such as $\Omega^3 = \bigcup_k B_{i_k}$, under the constraint that the intersections between the $B_{i_k}$ are non-empty (see Fig 3). These blocks are centered on voxels $x_{i_k}$ which constitute a subset of $\Omega^3$. The $x_{i_k}$ are equally distributed at positions $i_k = (k_1 n, k_2 n, k_3 n), (k_1, k_2, k_3) \in \mathbb{N}^3$ where $n$ represents the distance between the centers of $B_{i_k}$. To ensure a global continuity in the denoised image, the support overlapping of blocks has to be non empty: $2\alpha > n$.

b) For each block $B_{i_k}$, a NL-means-like restoration is performed as follows:

$$
\mathbf{NL}(u)(B_{i_k}) = \sum_{B_j \in V_{i_k}} w(B_{i_k}, B_j)\mathbf{u}(B_j)
$$
(8)

with

$$
w(B_{i_k}, B_j) = \frac{1}{Z_{i_k}} e^{-\frac{\|\mathbf{u}(B_{i_k})-\mathbf{u}(B_j)\|_2^2}{2\beta\hat{\sigma}^2|N_i|}}
$$
(9)

4

where $Z_{i_k}$ is a normalization constant ensuring that $\sum_{B_j \in V_{i_k}} w(B_{i_k}, B_j) = 1$ (see Fig. 1 (right)).

c) For a voxel $x_i$ included in several blocks $B_{i_k}$, several estimations of the restored intensity $NL(u)(x_i)$ are obtained in different $\mathbf{NL}(u)(B_{i_k})$ (see Fig 3). The estimations given by different $\mathbf{NL}(u)(B_{i_k})$ for a voxel $x_i$ are stored in a vector $\mathbf{A}_i$. The final restored intensity of voxel $x_i$ is then defined as:

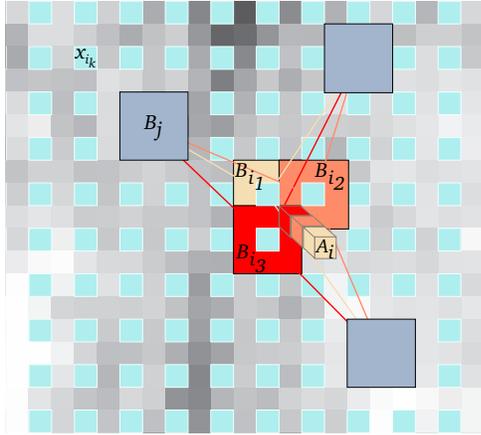$$NL(u)(x_i) = \frac{1}{|\mathbf{A}_i|} \sum_{p \in \mathbf{A}_i} \mathbf{A}_i(p). \tag{10}$$



Fig. 3. **Blockwise NL-means Filter**. For each block $B_{i_k}$ centered on voxel $x_{i_k}$, a NL-means like restoration is performed from blocks $B_j$. In this way, for a voxel $x_i$ included in several blocks, several estimations are obtained. The restored value of voxel $x_i$ is the average of the different estimations stored in vector $\mathbf{A}_i$. In this example $\alpha = 1$, $n = 2$ and $|\mathbf{A}_i| = 3$.

The main advantage of this approach is to significantly reduce the complexity of the algorithm. Indeed, for a volume $\Omega^3$ of size $N^3$, the global complexity is $\mathcal{O}((2\alpha + 1)^3 (2M + 1)^3 (\frac{N-n}{n})^3)$. For instance, with $n = 2$, the complexity is divided by a factor 8. The voxels selection principle can also be applied to the blockwise implementation:

$$w(B_{i_k}, B_j) = \begin{cases} \frac{1}{Z_{i_k}} e^{-\frac{\|\mathbf{u}(B_{i_k}) - \mathbf{u}(B_j)\|_2^2}{2\beta \hat{\sigma}^2 |N_i|}} & \text{if } \mu_1 < \frac{\overline{\mathbf{u}(B_{i_k})}}{\overline{\mathbf{u}(B_j)}} < \frac{1}{\mu_1} \\ & \text{and } \sigma_1^2 < \frac{\text{Var}(\mathbf{u}(B_{i_k}))}{\text{Var}(\mathbf{u}(B_j))} < \frac{1}{\sigma_1^2} \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

where $\overline{\mathbf{u}(B_{i_k})}$ and $\text{Var}(\mathbf{u}(B_{i_k}))$ represent respectively the mean and the variance of the intensity function, for the block $B_{i_k}$ centered on the voxel $x_{i_k}$.

*4) Parallel computation:* Another way to reduce the computational time is to distribute the operations on several processors via a cluster or a grid. The intrinsic nature of the NL-means filter makes it perfectly suited for parallelization and multithreading implementation. One of the main advantage of this filter, when compared to others method such as Anisotropic Diffusion or Total Variation minimization, is that the operations are performed without any iterative schemes. Thus, the parallelization of the NL-means filter is straightforward to perform and very efficient. We divide the volume into sub-volumes, each of them being treated separately by one processor. A server with 8 Xeon processors at 3 GHz and a Intel(R) Pentium(R) D CPU 3.40GHz were used in our experiments.

## IV. MATERIALS

### A. The BrainWeb Database

In order to evaluate the performances of the NL-means filter on 3D MR images, tests were performed on the BrainWeb database[1] [2]. Two images were simulated: T1-w MR image using SFLASH sequence (volume size $= 181 \times 217 \times 181$) and T2-w MR image with MS from SFLASH sequence (volume size $= 181 \times 217 \times 181$). As reported previously, it is a known fact that the MR images are corrupted by a Rician noise [53], [54], which can be well approximated by a white Gaussian noise in high intensity areas, typically in brain tissues [38]. In order to verify if this approximation can be used for a NL-means based denoising, experiments are performed on phantom images with Gaussian and Rician noise.

*1) Gaussian Noise:* A white Gaussian noise was added on the "ground truth", and the notations of BrainWeb are used: a noise of 3% is equivalent to $\mathcal{N}(0, \nu \frac{3}{100})$, where $\nu$ is the value of the brightest tissue in the image (150 for T1-w and 250 for T2-w). Several images were simulated to validate the performances of the denoising on various images (see Fig. 4):

- T1-w MR images for 4 levels of noise 3%, 9%, 15% and 21%.
- T2-w MR images with Multiple Sclerosis (MS) lesions for 4 levels of noise 3%, 9%, 15% and 21%.

T2-w images were used in order to show that our approach and its calibration are not specific to T1-w MRI sequences. Moreover, the tests on T2-w MRI with MS show how the NL-means filter could be useful in a pathological context due to its preservation of anatomic and pathologic structures.

*2) Rician Noise:* The Rician noise was build from white Gaussian noise in the complex domain. Firstly, two images are computed:

- $I_r(x_i) = I_0(x_i) + \eta_1(x_i)$, $\eta_1(x_i) \frown \mathcal{N}(0, \sigma)$
- $I_i(x_i) = \eta_2(x_i)$, $\eta_2(x_i) \frown \mathcal{N}(0, \sigma)$

where $I_0$ is the "ground truth" and $\sigma$ is the standard deviation of the added white Gaussian noise. Then, the noisy image is computed as:

$$I_N(x_i) = \sqrt{I_r(x_i)^2 + I_i(x_i)^2} \tag{12}$$

The notation 3% for the Rician noise means that the Gaussian noise used in complex domain is equivalent to $\mathcal{N}(0, \nu \frac{3}{100})$, where $\nu$ is the value of the brightest tissue in the image (150 for T1-w). According to the Peak Signal to Noise Ratio (PSNR) (see Eq. 13) between "ground truth" and noisy images, for a same level of noise, the Rician noise is stronger than the Gaussian noise (see Tab. I). Several images were simulated (see Fig. 5):

- T1-w MR images for 4 levels of noise 3%, 9%, 15% and 21%.

### B. Real Data

*1) T1-w high field MRI Data:* To show the efficiency of the NL-means filter on real data, tests were performed on image acquired with a high field MR system (Bruker 3 Tesla). The data used was a $256 \times 256 \times 256$ T1-w image.

---

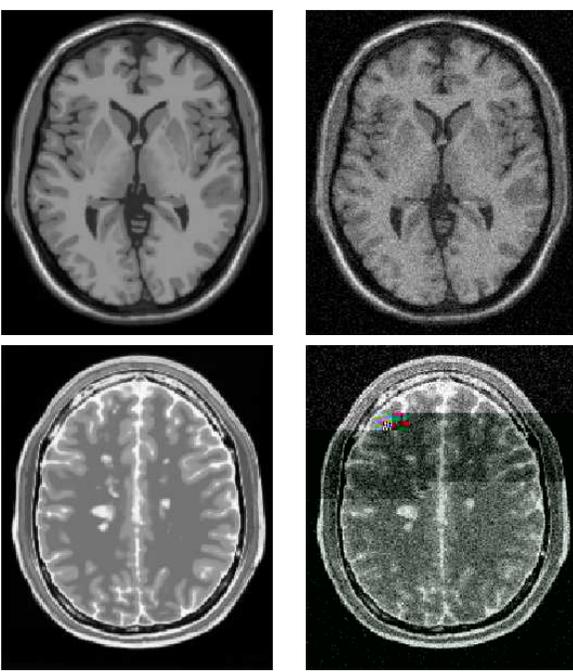[1]http://www.bic.mni.mcgill.ca/brainweb/

5

Fig. 4. **Synthetic data used for validation with Gaussian Nnoise.** Example of the Brainweb Database. Top: T1-w images without any noise (left), and corrupted with a white Gaussian noise at $9\%$ (right). Bottom: T2-w images with MS lesions without noise (left), and corrupted with a white Gaussian noise at $9\%$ (right).

| Noise level | PSNR with Gaussian noise in dB | PSNR with Rician noise in dB |
|---|---|---|
| $3\%$ | 35.09 | 35.05 |
| $9\%$ | 25.64 | 25.57 |
| $15\%$ | 21.30 | 21.17 |
| $21\%$ | 18.49 | 18.29 |

TABLE I

PEAK SIGNAL TO NOISE RATIO (PSNR) BETWEEN "GROUND TRUTH" AND NOISY IMAGES FOR GAUSSIAN AND RICIAN NOISES. FOR A SAME LEVEL OF NOISE, THE RICIAN NOISE IS STRONGER THAN THE GAUSSIAN NOISE.

*2) T2-w with Multiple Sclerosis lesions:* In a pathological context, the denoising step is crucial especially when the structures of interest have a small size: the integrity of pathological structures must be preserved by the denoising method. As said earlier, one objective of denoising is to include such processing in complex medical imaging workflows. This kind of workflows is widely used to process large cohort of subjects in many neurological diseases such as MS lesions. The data used for MS lesions qualitative validation was a T2-w MR image from an axial dual-echo, turbo spin-echo sequence (Philips 1.5 Tesla).

## V. VALIDATION ON A PHANTOM DATA SET WITH ADDED GAUSSIAN NOISE

In the following, let us define:

- **NL-means** is the standard voxelwise implementation with automatic tuning of the smoothing parameter.
- **Optimized NL-means** is a voxelwise implementation with automatic tuning of the smoothing parameter, voxel
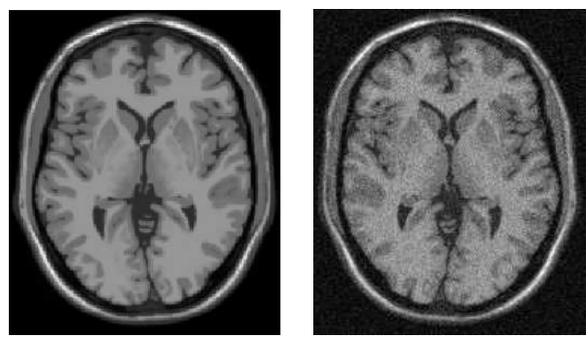


Fig. 5. **Synthetic data used for validation with Rician noise.** Example of the Brainweb Database. T1-w images without any noise (left), and corrupted with a Rician noise at $9\%$ (right).

selection and multithreading.

- **Blockwise NL-means** is the standard blockwise implementation with automatic tuning of the smoothing parameter.
- **Optimized Blockwise NL-means** is a blockwise implementation with automatic tuning of the smoothing parameter, block selection and multithreading.

In this section different aspects of NL-means filter implementation were investigated. First, the impact of the automatic tuning of the filtering parameter (Section V-A) and the influence of the size of the search volume and the neighborhood were studied (Section V-B). Then, the impact of voxels selection and blockwise implementation is analyzed *via* the comparison of the **NL-means**, **Optimized NL-means**, **Blockwise NL-means** and **Optimized Blockwise NL-means** filters (Sections V-C and V-D). Finally, we compare the proposed **Optimized Blockwise NL-means** filter with other well-established denoising methods: Anisotropic Diffusion filter [3] (implemented in VTK[2]) and Rudin-Osher-Fatemi Total Variation (TV) minimization process [4] (3D extension of the Megawave2 implementation[3]) (Section V-G). The different variants of the NL-means filter can be freely tested at: `http://www.irisa.fr/visages/benchmarks`

In the following, several criteria are used to quantify the performances of each method: the PSNR obtained for different noise levels, histogram comparisons between the denoised images and the "ground truth", and finally visual assessment. For the sake of clarity, the PSNR and the histograms are estimated only in a region of interest obtained by removing the background. For 8-bit encoded images, the PSNR is defined as follows:

$$PSNR = 20 \log_{10} \frac{255}{RMSE} \qquad (13)$$

where the RMSE is the root mean square error estimated between the ground truth and the denoised image.

In this section, the central parameters of interest are:

- $\beta$ defining the smoothing parameter $h$: $h^2 = 2\beta\hat{\sigma}^2|N_i|$ (see section III-B.1).
- $M$ related to $|V_i|$: $|V_i| = (2M + 1)^3$.
- $d$ related to $|N_i|$: $|N_i| = (2d + 1)^3$.

[2]www.vtk.org

[3]http://www.cmla.ens-cachan.fr/Cmla/Megawave/index.html

- $\mu_1$ and $\sigma_1^2$ corresponding to the thresholds involving in the voxel selection.

In each experiment (Sections V-A, V-B and V-C), we let one parameter vary while remaining the others constant, with default values: $\beta = 1$, $M = 5$, $d = 1$ and $\mu_1 = 0.95$, $\sigma_1^2 = 0.5$. Concerning the blockwise implementation the default parameters are $n = 2$ and $\alpha = 1$.

Our experiments have shown that all the versions of the NL-means filter (**NL-means**, **Optimized NL-means**, **Blockwise NL-means** and **Optimized Blockwise NL-means**) tend to have a similar behavior with respect to the variation of the parameters. In that context, all the results are displayed with the proposed **Optimized Blockwise NL-means** filter, even if equivalent conclusions can be drawn with the **NL-means**, **Optimized NL-means** and **Blockwise NL-means** filters.

Validation was performed on T1-w and T2-w MRI, but the results concerning the study of the parameter influences are shown for T1-w MRI only. The results on T2-w MRI are shown in section V-F in order to underline that the parameters calibrated for T1-w MRI work fine on T2-w MRI.

### A. Influence of the automatic tuning of smoothing parameter h

Figure 6 shows the influence of the automatic determination of the smoothing parameter $h^2 = 2\beta\hat{\sigma}^2|N_i|$. As described in III-B.1, $h$ is a function of the global standard deviation of the noise $\hat{\sigma}$ in the volume esimated from pseudo-residuals (see 3 and Eq. 4). Here, $\beta$ allows to adjust the automatic estimation of $h$ in order to determine the optimal smoothing parameter $2\beta\hat{\sigma}^2|N_i|$ for each level of noise (see 6). For low
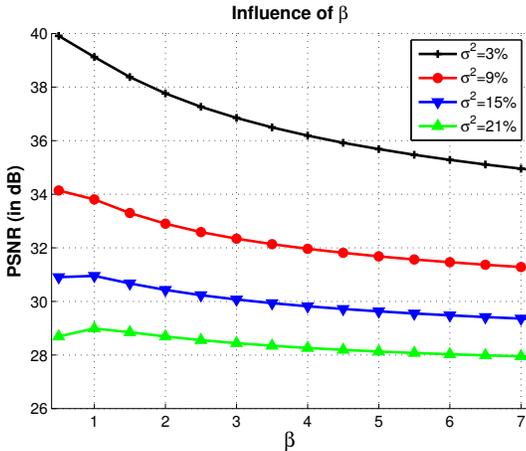


Fig. 6. **Calibration of the smoothing parameter** $h$**:** Influence of the smoothing parameter $2\beta\hat{\sigma}^2|N_i|$ on the PSNR, according to $\beta$ and for several levels of noise. For low levels of noise the best value of $\beta$ is close to 0.5. For high levels of noise this value is 1. The default value of $\beta$ is set to 1, thus the estimation of $h$ is $h^2 = 2\hat{\sigma}^2$. These results are obtained with $\hat{\sigma}^2 = 3.42\%$ at 3%, $\hat{\sigma}^2 = 7.93\%$ at 9%, $\hat{\sigma}^2 = 12.72\%$ at 15% and $\hat{\sigma}^2 = 17.44\%$ at 21%.

levels of noise the best value of $\beta$ is close to 0.5. For high levels of noise this value is 1. These results show that the estimation of the standard deviation of the noise is correctly performed by pseudo-residuals. These observations underline (a) how efficient the automatic estimation of the smoothing

parameter $h$ is, and (b) how the NL-means can be used without manual parameter tuning.

### B. Influence of the size of the search volume and the neighborhood

Figure 7 shows the influence of the size of the search volume and the local neighborhood. Increasing the number of voxels in the search volume $V_i$ does not seem to affect the PSNR when $M$ is greater than 5. Indeed, the theoretic definition of the NL-means filter states that the weighted average (see Eq. 1) computed for the restoration of voxel $x_i$ should be performed on all voxels $x_j \in \Omega^3$. Practically, the limit $M = 5$ prevents useless computations. Moreover, increasing $d$ degrades the denoising process. When $d$ increases the NL-means filter drastically slows down. That is why we have not investigated the impact of $d$ for $d > 3$.

### C. Influence of the voxel selection

The selection of the voxels in the search volume $V_i$ is achieved by supposing that only the voxels whose the neighborhood is similar to the neighborhood of the voxel under study could be considered (see Eq. 1). To do so, as defined in III-B.2, the weight $w(x_i, x_j)$ is calculated only for voxels such that: $\mu_1 < \frac{\overline{\mathbf{u}(N_i)}}{\overline{\mathbf{u}(N_j)}} < \frac{1}{\mu_1}$ and $\sigma_1^2 < \frac{\mathrm{Var}(\mathbf{u}(N_i))}{\mathrm{Var}(\mathbf{u}(N_j))} < \frac{1}{\sigma_1^2}$. The influence of the limits $\mu_1$ and $\sigma_1^2$ is studied in Figure 8. In a first experiment $\mu_1$ varies according to $\gamma$ such as $\mu_1 = 1 - \gamma$ while $\sigma_1^2 = 0.5$. In a second experiment $\sigma_1^2$ varies according to $\gamma$ following $\sigma_1^2 = 1 - \gamma$ while $\mu_1 = 0.95$.

Figure 8 (left) shows that a restrictive selection based on the mean (low values of $\gamma$) increases the PSNR. In other words, the number of voxels taken into account in the weighted average is drastically reduced, as well as the computational time (also see Tab. II). The optimal limits were obtained for $\mu_1 = 0.95$ while $\sigma_1^2 = 0.5$. Concerning the variance (Figure 8, right), we observe that a too restrictive selection degrades the PSNR. In addition, a too permissive selection does not increase the PSNR while increasing uselessly the computational burden. A compromise was found by fixing $\sigma_1^2 = 0.5$. There is a clear dependency between the bounds for the mean and the variance. An optimal trade-off was determined experimentally.

### D. Influence of the blockwise implementation

Tab. II shows that the blockwise approach of the NL-means filter, with and without voxels selection (see Eq. 11), allows to drastically reduce the computational time. With a distance between the block centers $n = 2$, the blockwise approach divides this time by a factor $2^3 = 8$ (see Tab. II). However, computational time reduction needs to be balanced with a slight decrease of the PSNR (see Fig. 9, left). For the optimized versions, the voxels/blocks selection in the search volume has several impacts. First, by reducing the average number of voxels/blocks used in the weighted averages, this decreases the computational time compared to the non-optimized versions (see Tab. II). Second, the selection of the most relevant voxels/blocks increases the quality of denoising for all the noise levels (see Fig. 9 (left) and Tab. II).
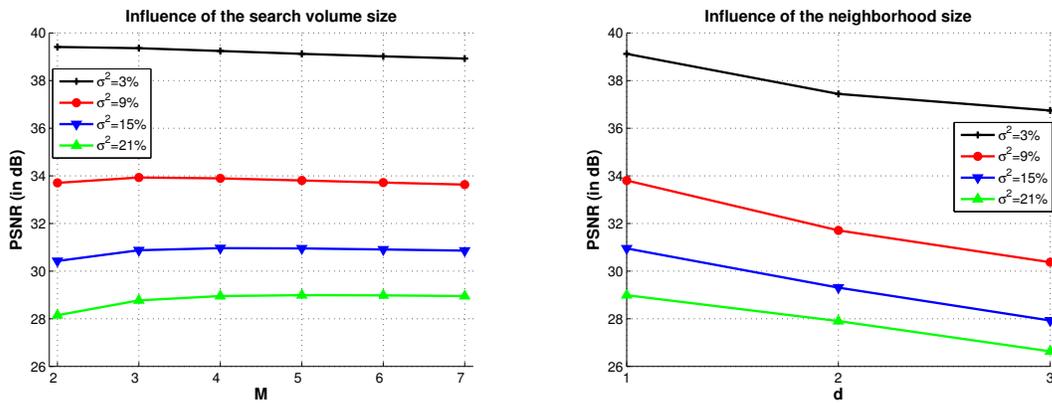
Fig. 7. **Influence of the size** $|V_i| = (2M + 1)^3$ **and** $|N_i| = (2d + 1)^3$ **for denoising:** Influence of the size of the search volume and the size of the neighborhood on the PSNR, for several levels of noise. Left: Variation of the size $M$ of the search volume $V_i$ for $d = 1$. Right: Variation of the size $d$ of the neighborhood $N_i$ for $M = 5$. These results show that the limit $M = 5$ prevents useless computation. Moreover, increasing $d$ degrades and drastically slows down the algorithm.
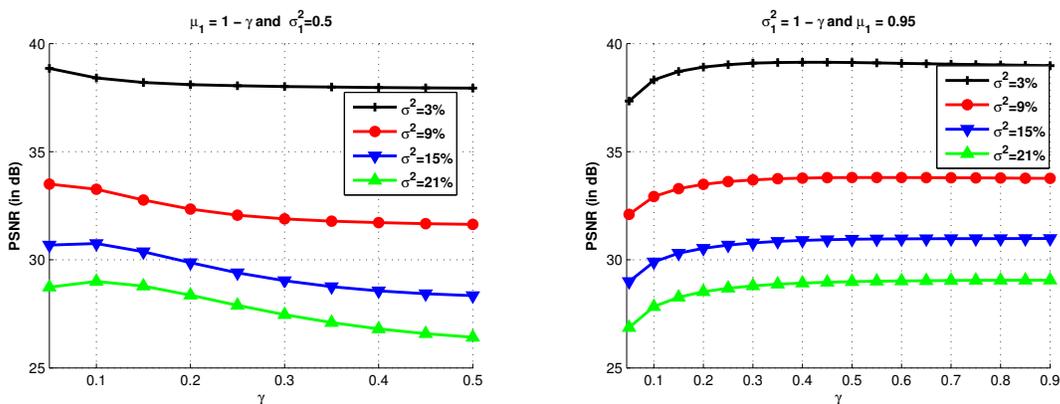


Fig. 8. **Influence of the limits of the voxels selection.** Influence of the limits $\mu_1$ and $\sigma_1^2$ on the PSNR, for several level of noise. Left: $\sigma_1^2 = 0.5$, while $\mu_1$ varies with $\gamma$. A restrictive selection based on the mean (low values of $\gamma$) increases the PSNR. The optimal limits are obtained for $\mu_1 = 0.95$. Right: $\mu_1 = 0.95$ and $\sigma_1^2$ varies accordingly to $\gamma$. A too restrictive selection (low values of $\gamma$) degrades the PSNR. In addition, a too permissive selection (high values of $\gamma$) does not increase the PSNR while concurrently increasing uselessly the computational burden. A good compromise is found by fixing $\sigma_1^2 = 0.5$.
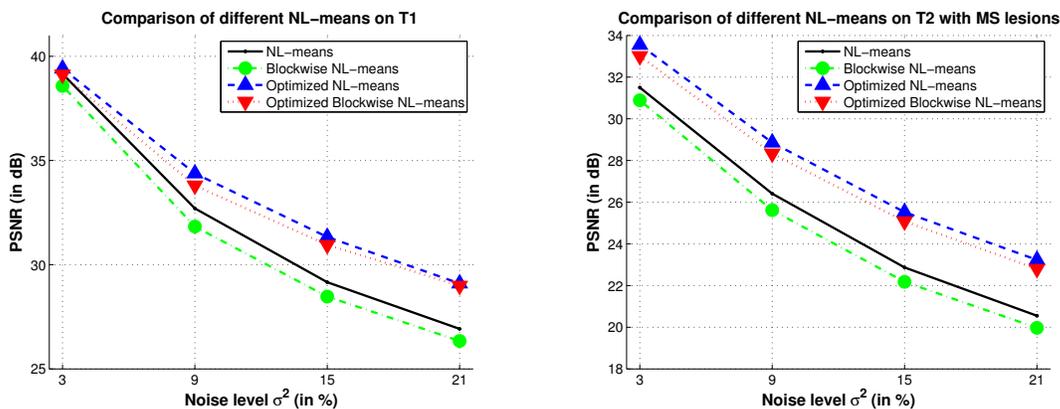


Fig. 9. **Impact of the blockwise implementation and voxels selection.** Comparison of the different implementations of the NL-means filter, with $\alpha = 1$. Left: on T1-w images. For the **Optimized Blockwise NL-means** filter, as for the **Optimized NL-means** filter, the selection of voxels/blocks in the search volume improves the quality of denoising and decreases the computational burden (see Tab. II). The reduction of computational time brought by the blockwise approach needs to be balanced with a slight decrease in quality of denoising. Right: on T2-w images with MS lesions. The same conclusions can be drawn for this kind of images. These results suggest that the parameters tuning determined experimentally on T1-w images are not T1-specific.

## E. Multithreading

As described in section III-B.4, the multithreading in the case of the NL-means filter is particularly adapted due its non iterative nature. For the classical pixelwise NL-means implementation, the parallelization allows to divide the computational time by a factor close to the numbers of CPU. As eight processors were used for our experiments, the computational time with multithreading is about 8 times smaller (see Tab. II, $\frac{21790}{2780} = 7.84$ and $\frac{3169}{436} = 7.27$). For the blockwise implementations the speedup is less ($\frac{1800}{251} = 7.17$ and $\frac{328}{63} = 5.37$). The difference of speedup between the classical NL-means and the blockwise NL-means filters have two origins.

- First, in blockwise version several threads could write at the same memory location (i.e. vector $A_i$) at the same time. In multi-treading programming this kind of possibilities requires a lock which protects the memory location during the writing. Unfortunately, to lock a memory location speeds down the computational process.
- Second, as the required computation time is shorter for the blockwise than for the voxelwise implementation, the relative contribution of the non-multithreaded operations in the overall computation time (opening and closing of file, computation of the local maps, etc.) is much higher in the blockwise compared to the voxelwise implementation. As a consequence, the speed-up factor will be higher in the latter

In order to underline that the utilization of 8-CPUs is not required by our filter, the denoising have been also performed on a more common architecture: a DualCore Intel(R) Pentium(R) D CPU 3.40GHz. The results show that our filter takes less than 3 minutes to denoise a volume $181 \times 217 \times 181$ voxels on this architecture.

To conclude, the different improvements included in the proposed **Optimized Blockwise NL-means** filter (i.e., blockwise approach and blocks selection) allow to speed up the denoising procedure, compared to **NL-means** filter, by a factor of 66 on 1 Xeon at 3GHz, 44 on $8\times$ Xeon at 3GHz and 31 on a DualCore at 3.40GHz.

## F. *Optimized Blockwise NL-means* filter on T2-w MRI with MS

Figure 9 (right) presents the results obtained by the different NL-means filter versions on T2-w MRI with MS lesions. The optimal parameters (i.e. the default parameters described in section V), experimentally determined on T1-w MRI, and the automatic tuning of $h$ were used on T2-w MRI. The **Optimized NL-means** and **Optimized Blockwise NL-means** filters outperform the **NL-means** and **Blockwise NL-means** filters also on T2-w MRI. The most important difference between the optimized and non-optimized versions are observed on T2-w MRI, which could be explained by the higher level of noise in the simulated T2-w MRI compared to T1-w MRI. Actually, the variance of noise varies with respect to the highest intensity tissues which is 150 in T1-w and 250 in T2-w. For $9\%$ the variance of noise is 13.5 in T1-w images and is 22.5 in T2-w images because the highest tissue intensity is superior in T2-w images. These results suggest that the parameters experimentally tuned on T1-w images can be used for T2-w images. Figure 10 shows an example of denoising obtained by the optimized blockwise NL-means and the blockwise NL-means filters. The MS lesions are visually more preserved with the optimized version; this was confirmed by an experienced MRI reader.

## G. Comparison with other denoising methods

*1) Focus on two classical denoising approaches:*

*a) Anisotropic Diffusion filtering:* As reported in Section II, the Anisotropic Diffusion filter (AD) was introduced to overcome the blurring effect of the Gaussian smoothing approach. First introduced by Perona and Malik [3], in this approach the image $u$ is only convolved in the direction orthogonal to the gradient of the image which ensures the preservation of edges. The iterative denoising process of initial image $u_0$ can be expressed as:

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = \text{div}(c(x,t)\nabla u(x,t)) \\ u(x,0) = u_0(x) \end{cases} \quad (14)$$

where $\nabla u(x,t)$ is the image gradient at voxel $x$ and iteration $t$, $\frac{\partial u(x,t)}{\partial t}$ is the partial temporal deviation of $u(x,t)$ and

$$c(x,t) = g(\|\nabla u(x,t)\|) = e^{-\frac{\|\nabla u(x,t)\|}{K^2}} \quad (15)$$

where $K$ is the diffusivity parameter. The AD filter method produces a good preservation of edges [34], [35]. Nonetheless, the main disadvantage of Ad filter is to poorly denoise the constant regions (see Fig. 13).

*b) Total Variation minimization scheme:* The difficult task to preserve edges while correctly denoising constant areas has been addressed also by Rudin, Osher and Fatemi. They proposed to minimize the Total Variation (TV) norm subject to noise constraints [4], that is:

$$\hat{u} = \arg \min_{u \in \Omega^3} \int |\nabla u(x)| dx \quad (16)$$

subject to

$$\int_{\Omega^3} (u(x) - u_0(x)) dx = 0 \text{ and } \int_{\Omega^3} |u(x) - u_0(x)|^2 dx = \sigma^2 \quad (17)$$

where $u_0$ is the original noisy image, $u$ is the restored image and $\sigma$ the standard deviation of the noise. In this model, the TV minimization tends to smooth inside the image structures while keeping the integrity of boundaries. The TV minimization scheme can be expressed as an unconstrained problem:

$$\hat{u} = \arg \min_{u \in \Omega^3} \left[ \int_{\Omega^3} |\nabla u(x)| dx + \lambda \int_{\Omega^3} |u(x) - u_0(x)|^2 dx \right] \quad (18)$$

where $\lambda$ is a Lagrange multiplier which controls the balance between the TV norm and the fidelity term. Thus, $\lambda$ acts as the filtering parameter. Indeed, for high values for $\lambda$ the fidelity term is encouraged. For small values for $\lambda$ the regularity term is desired. In practice, the TV minimization scheme tends to remove texture and small image structures as seen in Fig. 13 [36]. To solve this problem, iterative total variation schemes have been recently developed [55], [56].

| Gaussian Noise | Computational time (in s) | | | PSNR (in dB) | Average number of voxels/blocks used in $V_i$ to denoise $x_i$ |
| | Xeon 3GHz | 8 × Xeon 3GHz | DualCore 3.40 GHz | | |
|---|---|---|---|---|---|
| NL-means | 21790 | 2780 | 4208 | 32.59 | $11^3 = 1331$ voxels |
| Blockwise NL-means | 1800 | 251 | 734 | 31.73 | $11^3 = 1331$ blocks |
| Optimized NL-means | 3169 | 436 | 778 | 34.44 | 174.8 voxels |
| Optimized Blockwise NL-means | 328 | 63 | 135 | 33.75 | 174.8 blocks |

TABLE II

**COMPARISON OF DIFFERENT IMPLEMENTATIONS OF THE NL-MEANS FILTERS IN TERMS OF COMPUTATIONAL TIME AND DENOISING QUALITY.** THE TIME IS OBTAINED WITH MULTITHREADING ON 8 CPUs AT 3GHz AND INTEL(R) PENTIUM(R) D CPU 3.40GHz AND WITHOUT MULTITHREADING ON 1 CPU AT 3GHz. THESE RESULTS ARE OBTAINED ON A T1-W BRAINWEB IMAGE WITH 9% OF GAUSSIAN NOISE ($\sigma = 13.5$). THE PARAMETERS USED ARE THE DEFAULT PARAMETERS. THE AVERAGE NUMBER OF VOXELS/BLOCKS USED IN $V_i$ TO DENOISE $u(x_i)$ SHOWS THE IMPACT OF VOXELS/BLOCKS SELECTION. FOR THE NON-OPTIMIZED IMPLEMENTATIONS ALL THE VOXELS/BLOCKS IN $V_i$ ARE TAKEN INTO ACCOUNT TO DENOISE $u(x_i)$. THUS, THE NUMBER OF VOXELS/BLOCKS USED ARE $|V_i| = (2M + 1)^3 = 11^3$. FOR THE OPTIMIZED IMPLEMENTATIONS THE VOXEL SELECTION ALLOWS TO DRASTICALLY REDUCE THIS NUMBER.



"Ground truth"     Noisy image 9%     **Optimized Blockwise NL-means**     **Blockwise NL-means**
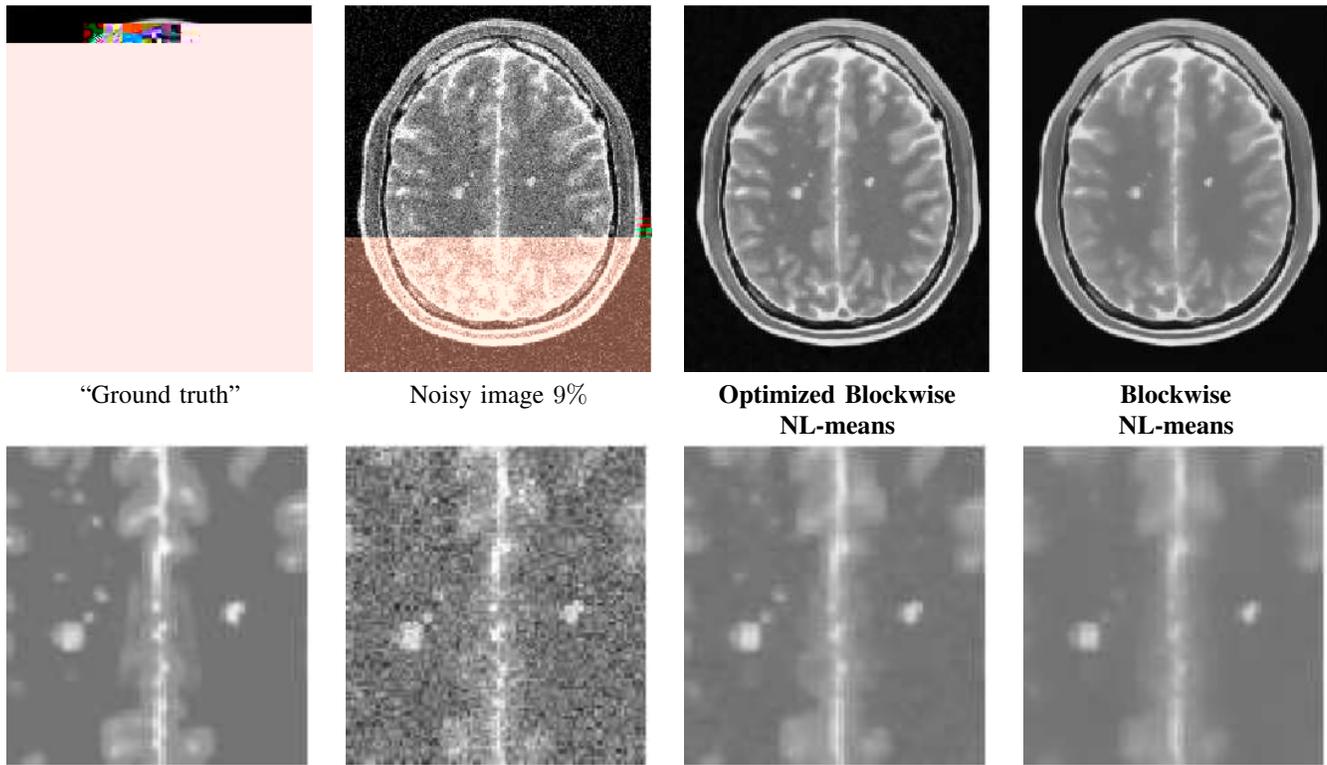
Fig. 10. **Comparison of the optimized and non-optimized blockwise NL-means on T2-w images:** NL-means restoration of T2-w Brainweb data with MS lesions. From left to right: "Ground truth", noised image at 9% of Gaussian noise, restored images by the **Optimized Blockwise NL-means** filter and by the **Blockwise NL-means** filter. The **Optimized Blockwise NL-means** filter preserves efficiently the contours of the MS lesions.

*2) Quantitative and qualitative comparison:* The main difficulty to achieve this comparison is related to the tuning of smoothing parameters in order to obtain the best results for AD filter and TV minimization scheme. In order not to penalize AD filter and TV minimization scheme, an exhaustive search for all parameters into a certain range. Then, the best results obtained with AD filter and TV minimization have been selected, whereas the fully-automatic results have been mentioned for the NL-means filters. The results of the NL-means filters are not "optimal" due to the non perfect estimation of the noise standard deviation. For AD filter, the parameter K varies from 0.05 to 1 with a step of 0.05 and the number of iterations varies from 1 to 10. For TV minimization,

the parameter $\lambda$ varies from 0.01 to 1 with step of 0.01 and the number of iterations varies from 1 to 10. The results obtained for 9% of Gaussian noise are presented Fig. 11, but this screening have been performed for the four levels of noise. It is important to underline that the results giving the best PSNR are used, but these results do not necessary give the best visual output. Indeed, the best PSNR for AD filter is obtained for a visually under-smoothed image since this method tends to spoil the edges. To obtain a high PSNR, the denoised image needs to balance edge preserving and noise removing. For AD filter, this trade-off leads to inhomogeneities in flat areas in denoised image (see Fig. 13). For TV minimization, the best PSNR is obtained with a visually under-smoothed image since
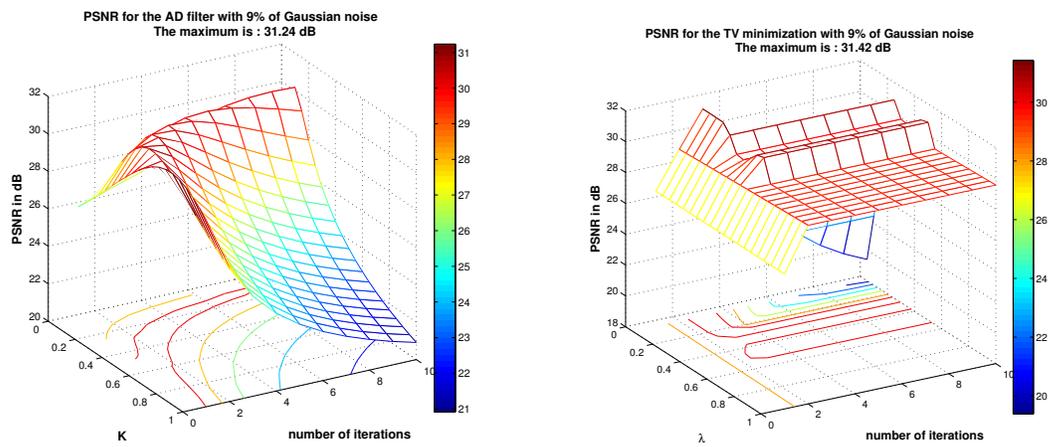
Fig. 11. Result for AD filter and TV minimization on phantom images with Gaussian noise at 9%. For AD filter K varies from 0.05 to 1 with step of 0.05 and the number of iterations varies from 1 to 10. For TV minimization $\lambda$ varies from 0.01 to 1 with step of 0.01 and the number of iterations varies from 1 to 10

noise is present in denoised image (see Fig. 13).

*a) PSNR comparison:* As presented in Fig. 12 (left), our block optimized NL-means filter produces the best PSNR values whatever the noise level. On average, a gain of $1.85$ dB is measured compared to TV and Anisotropic Diffusion methods. The PSNR value between the noisy image and the ground truth is called "No processing" and is used as reference.

*b) Histogram comparison:* To better understand how these differences in the PSNR values between the three methods can be explained, the histograms of the denoised images were compared to the histogram of the ground truth. Figure 12 (right) shows that the **Optimized Blockwise NL-means** filter is the only method able to retrieve a histogram similar to the ground truth. The NL-means-based restoration schemes clearly distinguish the three main peaks representing the white matter, the gray matter and the cerebrospinal fluid. The sharpness of the peaks shows how the **Optimized Blockwise NL-means** filter increases the contrast between denoised biological structures (see also Fig. 13). The distances between these histograms are estimated with the Bhattacharyya coefficient ($BC$) defined as:

$$BC(p,q) = \sum_{b=0}^{255} \sqrt{p(b)q(b)} \qquad (19)$$

where $p$ and $q$ are the two histograms be to compared and $b$ is the bin index. A $BC$ close to 1 means $p$ and $q$ are very similar. Each histogram of denoised images is compared to the "ground truth" one (see Tab. III). The distance between the histogram of the noisy image and the histogram of the "ground truth" is used as a reference. The BC distance shows that the restored histogram obtained with the **Optimized Blockwise NL-means** filter is the closest to the "ground truth", as visually assessed in Figure 12 (right). Finally, Table III suggests that the NL-means-based approach could improve the registration of images, since the Mutual Information ($MI$) computed between the restored image and the "ground truth" is higher in comparison with AD filter and TV minimization. The $MI$ is a similarity measure commonly used in image registration.

*c) Visual Assessment:* Figures 13 and 14 show the restored images and the removed noise obtained with the three compared methods. As shown in the previous analysis, we observe that the homogeneity of white matter is higher when the image is denoised with the **Optimized Blockwise NL-means** filter. Moreover, focusing on the structure of the removed noise, it clearly appears that NL-means-based restoration schemes better preserves the high frequency components of the image corresponding to anatomical structures while removing efficiently the high frequencies due to noise. According to the "method noise" introduced in [57], the NL-means is a better denoising method since the removed noise is the most similar to a white Gaussian noise. Finally, the difference between the "ground truth" and the denoised image is presented in order to show which structures are removed during the denoising process. In Fig. 13, this difference shows that (a) the AD filter tends to spoil the edges especially on the skull, (b) the TV minimization slightly better preserves the edges but does not remove all the noise, and (c) the **Optimized Blockwise NL-means** filter visually better preserves the edges while efficiently removing the noise (especially for white matter).

## VI. Validation on a Phantom data set with added Rician Noise

In this section, the same experiments are performed on phantom data set corrupted by Rician noise in order to study the impact of the Gaussian assumption. Table IV shows the computation time and the denoising performance of the different compared NL-means filters. These results show that the optimized NL-means versions outperform the classical ones also for Rician noise. Figure 15 presents the comparison with AD filter and TV minimization in terms of PSNR values and histogram analysis. As for the AD filter and TV minimization, the NL-means-based denoising is able to correctly restore an image corrupted by Rician noise using a Gaussian approximation. When the histograms are compared, low values of intensity ($< 20$) are incorrectly restored for all the filters; the Gaussian approximation is not appropriate in that case.
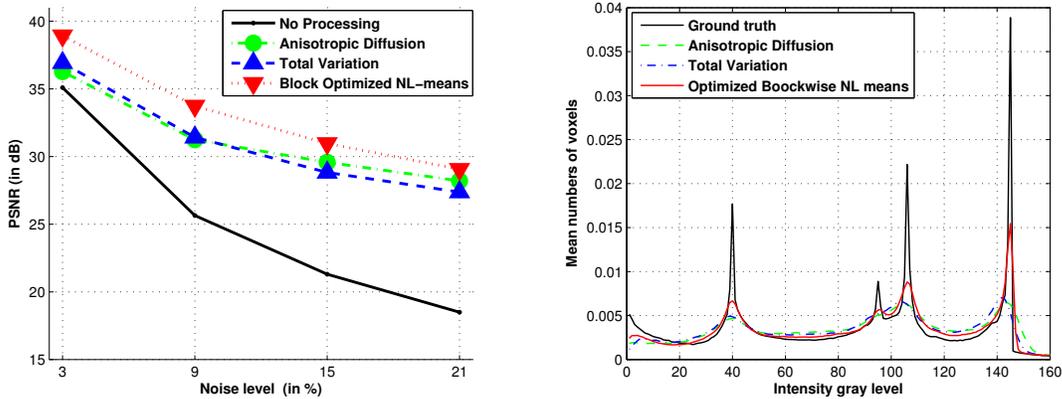
11

Fig. 12. **Comparison between Anisotropic Diffusion, Total Variation and Optimized Blockwise NL-means denoising:** the PSNR values and histograms for different denoising methods on BrainWeb at 9% of Gaussian noise. Left: the PSNR experiment shows that the **Optimized Blockwise NL-means** filter outperforms the well-established Total Variation minimization process and the Anisotropic Diffusion approach. Right: Contrary to others methods, the NL-means based restoration clearly distinguishes the three main peaks representing the white matter, the gray matter and the cerebrospinal fluid. The sharpness of the peaks shows how the **Optimized Blockwise NL-means** filter increases the contrast between denoised biological structures.

|  | Bhattacharyya Coefficient | Mutual Information |
|---|---|---|
| Noisy image | 0.9388 | 1.282 |
| Anisotropic diffusion | 0.9608 | 2.024 |
| Total Variation | 0.9639 | 1.974 |
| **Optimized Blockwise NL-means** | 0.9756 | 2.214 |

TABLE III

COMPARISON OF HISTOGRAMS OBTAINED WITH THE THREE DIFFERENT METHODS AT 9% OF GAUSSIAN NOISE. THIS TABLE PRESENTS (A) THE BHATTACHARYYA COEFFICIENT COMPUTED BETWEEN THE HISTOGRAMS OF DENOISED IMAGES AND THE "GROUND TRUTH" ONE AND (B) THE MUTUAL INFORMATION COMPUTED BETWEEN THE DENOISED IMAGES AND THE "GROUND TRUTH". THE DISTANCE BETWEEN THE NOISY IMAGE AND THE "GROUND TRUTH" IS USED AS A REFERENCE. COMPARED TO AD FILTER AND TV MINIMIZATION, THE **OPTIMIZED BLOCKWISE NL-MEANS** FILTER ALLOWS TO OBTAIN A DENOISED IMAGE WHOSE HISTOGRAM IS MORE CLOSER TO "GROUND TRUTH" HISTOGRAM.

Nevertheless, it seems the underlying assumption is well suited to high values ($> 60$).

As for Gaussian noise, the NL-means-based restoration clearly emphasizes the three main peaks corresponding to the white matter, the gray matter and the cerebrospinal fluid. Figure 16 shows the visual results obtained when the methods are compared on phantom data with Rician noise. Compared to Fig. 13, the denoising of background is worse in the Rician case, but the cerebral structures are correctly restored with the NL-means filter especially the white matter (see Fig. 16). Finally, Figure 17 shows the PSNR results of the parameter screening for the AD filter and the TV minimization at 9% of Rician noise. All these results on Rician noise show that the PNSR values slightly decrease due to more pronounced noise compared to Gaussian case for a same level (see IV-A.2 for explanation), but the general performance of the filters is preserved.

## VII. EXPERIMENTS ON CLINICAL DATA

### A. High field MRI

The restoration results presented in Fig. 18 show good preservation of the cerebellum contours. Fully automatic segmentation and quantitative analysis of such structures are still a challenge, and improved restoration schemes could greatly improve these processings.

### B. MS pathological context

Figure 19 shows that the optimized blockwise NL-means filter preserves the lesions while removing the noise. The impact on further processing is not the scope of this paper and is not studied here. Nevertheless, visually the lesions appears more contrasted and as seen on the difference image the proposed NL-means approach does not include any structure of lesion in the estimated noise image. This was confirmed by an experienced neurologist.

## VIII. DISCUSSION AND CONCLUSION

This paper presents an optimized blockwise version of the Non Local (NL-) means filter, applied to 3D medical data. Validation was performed on the BrainWeb dataset [2] and showed that the proposed **Optimized Blockwise NL-means** filter outperforms the classical implementation of the NL-means filter and some state-of-the-art techniques, such as the Anisotropic Diffusion approach [3] and the Total Variation minimization process [4] on both Gaussian and Rician noise. These first results show that the image-redundancy assumption required for NL-means based restoration holds for 3D MRI. Compared to the classical NL-means filter, our implementation (with voxel preselection, multithreading and blockwise implementation) considerably decreases the required computational time (up to a factor of 60 on a Xeon at 3GHz) and increases the PSNR value of the denoised image. Nevertheless, the problem
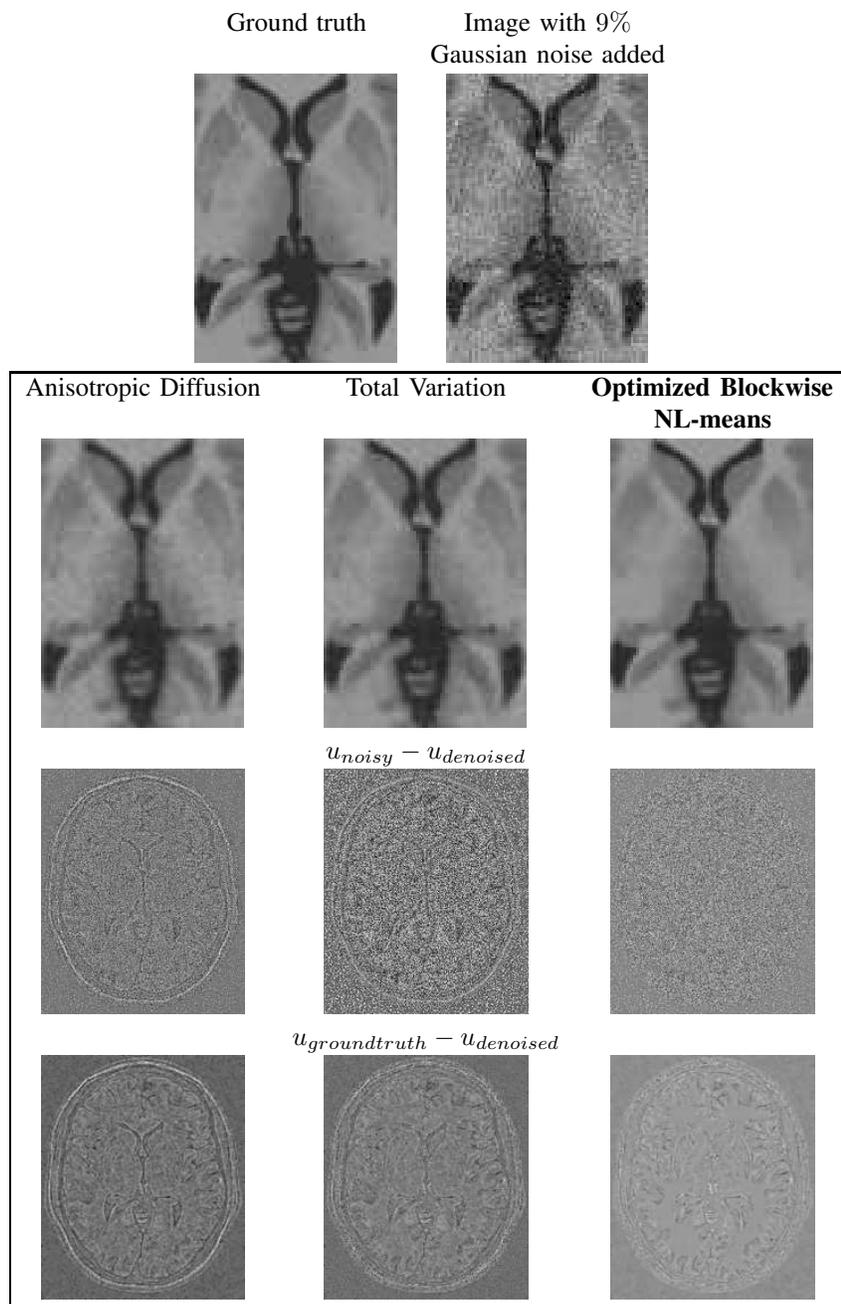
Fig. 13. **Comparison with Anisotropic Diffusion, Total Variation and NL-means denoising on synthetic T1-w images.** Top: zooms on T1-w BrainWeb images. Left: the "ground truth". Right: the noisy images with 9% of Gaussian noise. Middle: the results of restoration obtained with the different methods and the images of the removed noise (i.e. the difference (centered on 128) between the noisy image and the denoised image. Bottom: the difference (centered on 128) between the denoised image and the ground truth. Left: Anisotropic Diffusion denoising. Left: Anisotropic Diffusion denoising. Middle: Total Variation minimization process. Right: **Optimized Blockwise NL-means** filter. The NL-means based restoration better preserves the anatomical structure in the image while efficiently removing the noise as it can be seen in the image of removed noise.

of the computational burden can still be investigated with other faster implementations such as the "plain multiscale" scheme also suggested in [1]. Further works should be pursued for comparing NL-means based restoration with recent promising denoising methods, such as Total Variation in wavelet domain [43] or adaptive estimation method [28], [50]. Moreover, the efficiency of the technique limiting the staircasing effect proposed in [58] needs to be studied for MRI.

We show on sample pathological cases (patients with MS lesions) that the filter preserves the major visual signature of the given pathology. However, the impact on specific pathologies needs to be further investigated.

Finally, the impact of the NL-means-based denoising on the performances of post-processing algorithms, like segmentation and registration schemes also should be studied. Nonetheless, the first results presented on the Mutual Information ($MI$) suggest that the proposed **Optimized Blockwise NL-means** filter could improve the image registration process. Indeed, the $MI$ computed between the restored image and the "ground truth" is higher with the **Optimized Blockwise NL-means**

13

Ground truth | Image with 21% Gaussian noise added

Anisotropic Diffusion | Total Variation | **Optimized Blockwise NL-means**

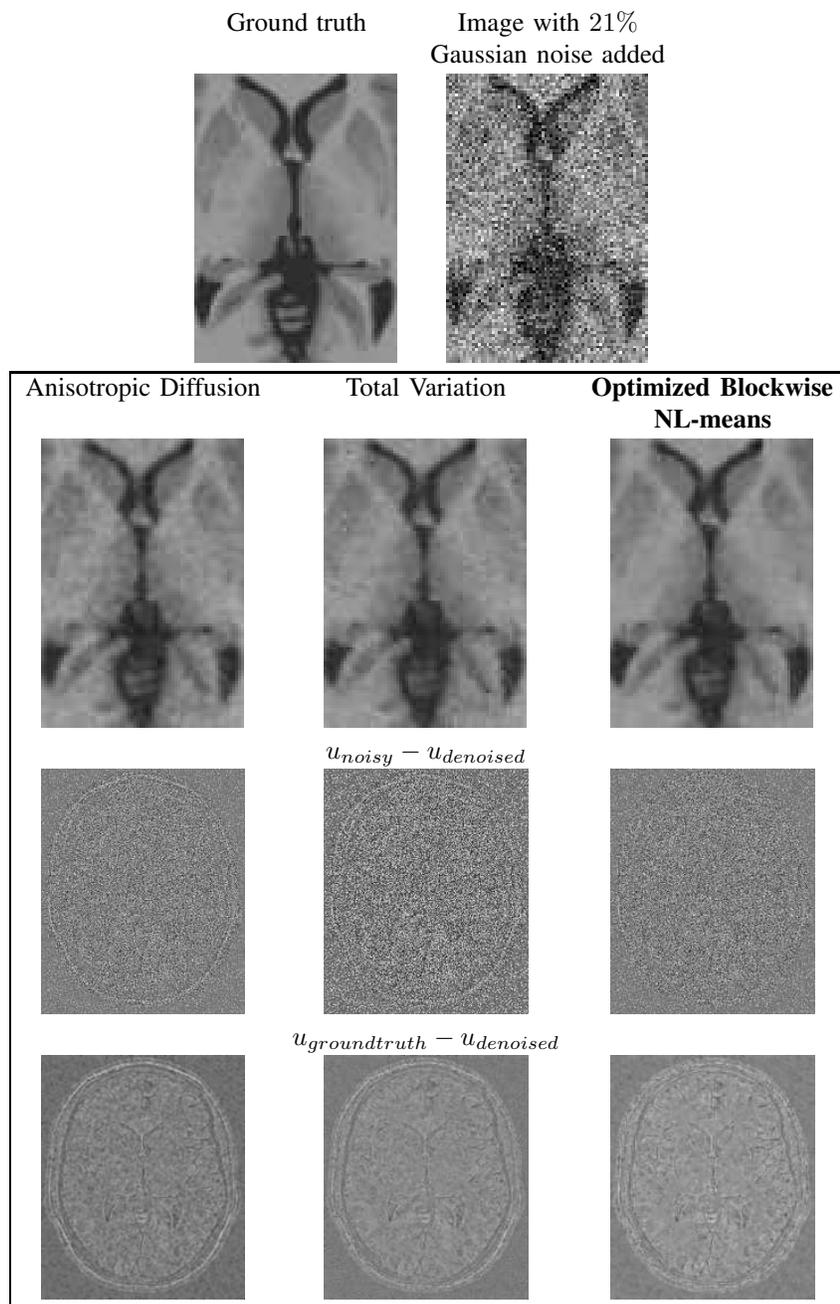$u_{noisy} - u_{denoised}$

$u_{groundtruth} - u_{denoised}$

Fig. 14. **Comparison with Anisotropic Diffusion, Total Variation and NL-means denoising on synthetic T1-w images.** Top: zooms on T1-w BrainWeb images. Left: the "ground truth". Right: the noisy images with 21% of Gaussian noise. Middle: the results of restoration obtained with the different methods and the images of the removed noise (i.e. the difference (centered on 128) between the noisy image and the denoised image. Bottom: the difference (centered on 128) between the denoised image and the ground truth. Left: Anisotropic Diffusion denoising. Middle: Total Variation minimization process. Right: **Optimized Blockwise NL-means** filter.

filter than with the Anisotropic Diffusion approach and the Total Variation minimization process.

REFERENCES

[1] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[2] D. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes, and A. Evans, "Design and construction of a realistic digital brain phantom." *IEEE Trans. Med. Imaging*, vol. 17, no. 3, pp. 463–468, 1998.

[3] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, 1990.

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

[5] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima, and B. Mazoyer, "A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm)." *Philos Trans R Soc Lond B Biol Sci*, vol. 356, no. 1412, pp. 1293–1322, August 2001.

[6] A. Zijdenbos, R. Forghani, and A. Evans, "Automatic "pipeline" analysis of 3D MRI data for clinical trials: application to multiple sclerosis." *IEEE Trans Med Imaging*, vol. 21, no. 10, pp. 1280–1291, October 2002.

| Rician Noise | Computational time (in s) | PSNR (in dB) | Average number of voxels/blocks used in $V_i$ to denoise $x_i$ |
|---|---|---|---|
| NL-means | 4190 | 31.35 | $11^3 = 1331$ voxels |
| Blockwise NL-means | 759 | 30.90 | $11^3 = 1331$ blocks |
| Optimized NL-means | 1045 | 33.40 | 251.1 voxels |
| Optimized Blockwise NL-means | 169 | 32.64 | 251.1 blocks |

TABLE IV

COMPARISON OF DIFFERENT IMPLEMENTATIONS OF THE NL-MEANS FILTER IN TERMS OF COMPUTATIONAL TIME AND DENOISING QUALITY. THESE RESULTS ARE OBTAINED ON A T1-W BRAINWEB IMAGE WITH 9% OF RICIAN NOISE ON A INTEL(R) PENTIUM(R) D CPU 3.40GHZ WITH 2GO OF RAM. THE PARAMETERS USED ARE THE DEFAULT PARAMETERS. THE AVERAGE NUMBER OF VOXELS/BLOCKS USED IN $V_i$ TO DENOISE $x_i$ SHOWS THE IMPACT OF VOXELS/BLOCKS SELECTION. FOR NON-OPTIMIZED IMPLEMENTATIONS ALL THE VOXELS/BLOCKS IN $V_i$ ARE TAKEN INTO ACCOUNT TO DENOISE $x_i$. THUS, THE NUMBER OF VOXELS/BLOCKS USED ARE $|V_i| = (2M+1)^3 = 11^3$. FOR OPTIMIZED IMPLEMENTATIONS THE VOXEL SELECTION ALLOWS TO DRASTICALLY REDUCE THIS NUMBER.
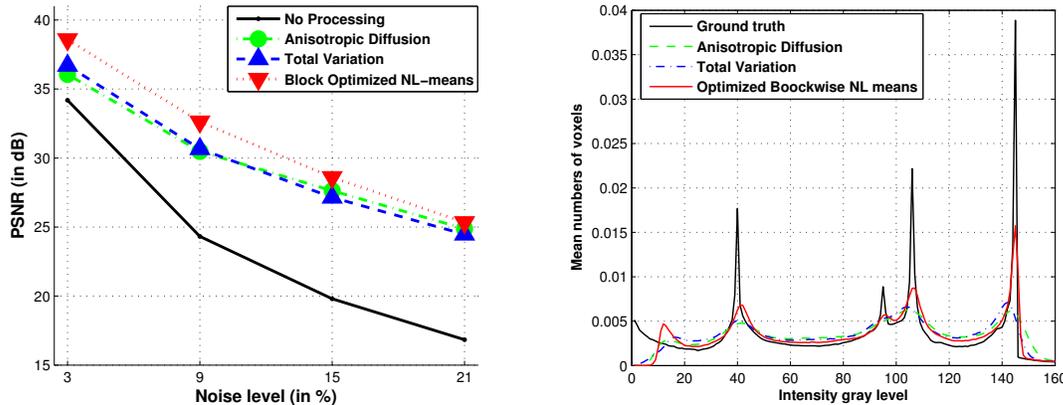


Fig. 15. **Comparison with Anisotropic Diffusion, Total Variation and Optimized Blockwise NL-means denoising.** PSNR values and histograms for different denoising methods on BrainWeb at 9% of Rician noise. Left: the PSNR study shows that the **Optimized Blockwise NL-means** filter outperforms the well-established Total Variation minimization process and the Anisotropic Diffusion approach. Right: When the histograms are compared low values of intensity ($< 20$) are incorrectly restored for all the filters; the Gaussian approximation is not appropriate in that case. Nevertheless, it seems the underlying assumption is well suited to high values ($> 60$). Contrary to others methods, the NL-means based restoration clearly emphasizes the three main peaks representing the white matter, the gray matter and the cerebrospinal fluid. The sharpness of the peaks shows how the **Optimized Blockwise NL-means** filter increases the contrast between denoised biological structures.

[7] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. PAMI.*, vol. 6, pp. 721–741, 1984.

[8] S. J. Mumford, D., "Optimal approximations by piecewise smooth functions and variational problems," *Comm. Pure and Appl. Math.*, vol. 42, pp. 577–685, 1989.

[9] D. Tschumperlé, "Curvature-preserving regularization of multi-valued images using PDE's," in *ECCV*, Graz, 2006, pp. 428–433.

[10] M. Black and G. Sapiro, "Edges as outliers: Anisotropic smoothing using local image statistics." in *Scale-Space Theories in Computer Vision, Second International Conference, Scale-Space'99, Corfu, Greece, September 26-27, 1999, Proceedings*, 1999, pp. 259–270.

[11] P. Saint-Marc, J.-S. Chen, and G. Medioni, "Adaptive smoothing: a general tool for early vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 514–529, June 1991.

[12] D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[13] D. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[14] J. Portilla and E. Simoncelli, "Image restoration using Gaussian scale mixtures in the wavelet domain." in *ICIP '03: International Conference on Image Processing*, 2003, pp. 965–968.

[15] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 1998, p. 839.

[16] S. Smith and J. Brady, "SUSAN–A New Approach to Low Level Image Processing," *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45–78, May 1997.

[17] M. Elad, "On the origin of the bilateral filter and ways to improve it," *IEEE Transactions on Image Processing*, vol. 11, no. 10, pp. 1141–1151, Oct. 2002.

[18] J. van de Weijer and R. van den Boomgaard, "Local mode filtering." in *CVPR '01: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 8-14 December, Kauai, HI, USA*, 2001, pp. 428–433.

[19] T. Chan and H. Zhou, "Total variation improved wavelet thresholding in image compression." in *ICIP*, 2000.

[20] S. Durand and J. Froment, "Reconstruction of wavelet coefficients using total variation minimization," *SIAM J. Sci. Comput.*, vol. 24, no. 5, pp. 1754–1767, 2002.

[21] S. Lintner and F. Malgouyres, "Solving a variational image restoration model which involves $L^\infty$ constraints," *Inverse Problems*, vol. 20, no. 3, pp. 815–831, 2004.

[22] P. Mrazek, J. Weickert, and A. Bruhn, "On robust estimation and smoothing with spatial and tonal kernels," *Preprint 51*, 2004.

[23] D. Barash, "A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 844–847, June 2002.

[24] G. Sapiro, "From active contours to anisotropic diffusion: connections between basic PDE's in image processing," in *ICIP'96: International Conference on Image Processing*, vol. 1, 1996, pp. 477–480 vol.1.

[25] S. V. Polzehl, J., "Adaptive weights smoothing with application to image restoration," *J. Roy. Stat. Soc. B*, vol. 62, pp. 335–354, 2000.

[26] V. Katkovnik, K. Egiazarian, and J. Astola, "Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule," *J. Math. Imaging Vis.*, vol. 16, no. 3, pp. 223–235, May 2002.

[27] C. Kervrann, "An adaptive window approach for image smoothing and structures preserving." in *Computer Vision - ECCV 2004, 8th European*
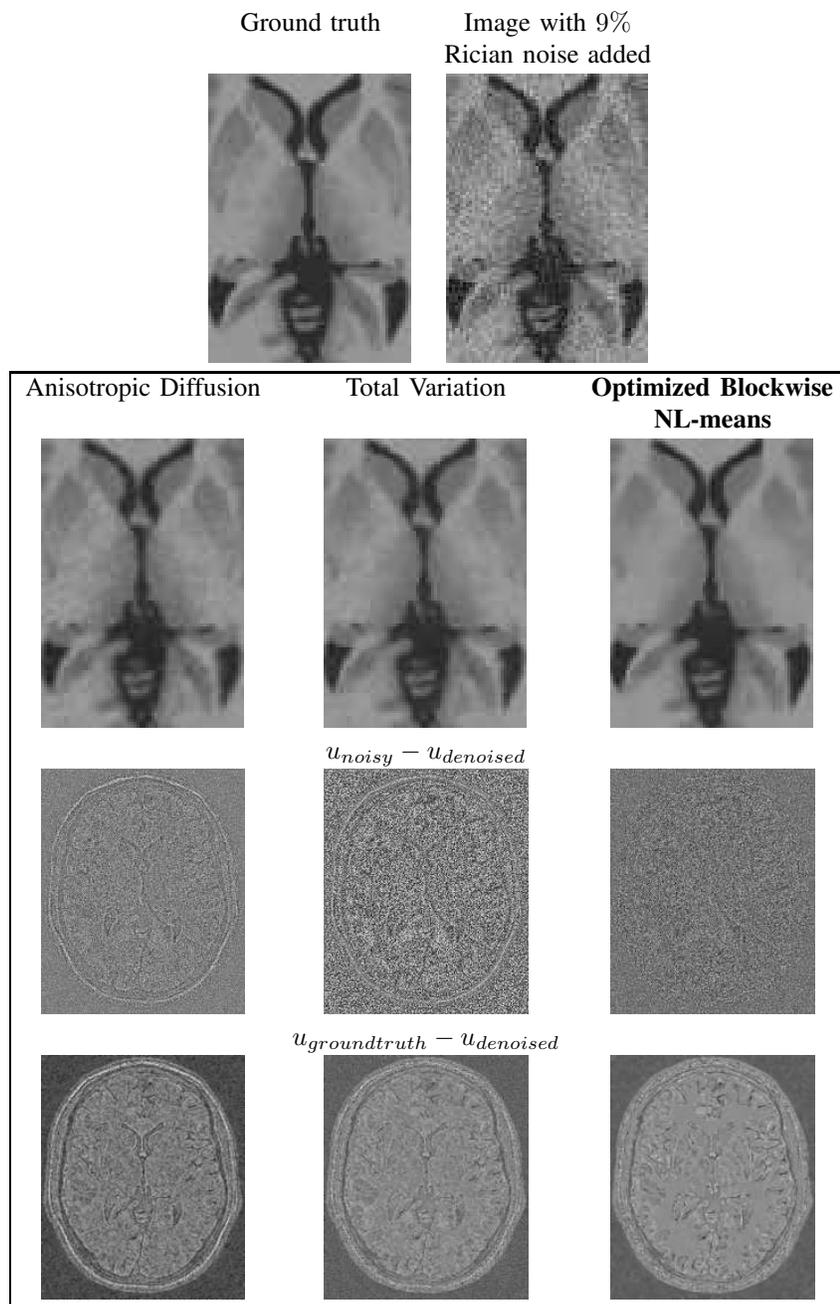
15

Fig. 16. **Comparison with Anisotropic Diffusion, Total Variation and NL-means denoising on synthetic T1-w images.** Top: zooms on T1-w BrainWeb images. Left: the "ground truth". Right: the noisy images with 9% of Rician noise. Middle: the results of restoration obtained with the different methods and the images of the removed noise (i.e. the difference (centered on 128) between the noisy image and the denoised image). Bottom: the "Method Noise" which is the difference (centered on 128) between the denoised image and the ground truth. Left: Anisotropic Diffusion denoising. Middle: Total Variation minimization process. Right: **Optimized Blockwise NL-means** filter. The NL-means based restoration better preserves the anatomical structure in the image while efficiently removing the noise, it can be seen in the image of removed noise.

*Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part III*, 2004, pp. 132–144.

[28] C. Kervrann and J. Boulanger, "Unsupervised patch-based image regularization and representation," in *Proc. European Conf. Comp. Vision (ECCV'06)*, Graz, Austria, May 2006.

[29] ——, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. on Image Processing*, vol. 15, no. 10, 2006.

[30] S. Zhu, Y. Wu, and D. Mumford, "Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling," *Int. J. Comput. Vision*, vol. 27, no. 2, pp. 107–126, 1998.

[31] W. Freeman, E. Pasztor, and O. Carmichael, "Learning low-level vision," *Int. J. Comput. Vision*, vol. 40, no. 1, pp. 25–47, October 2000.

[32] S. Roth and M. Black, "Fields of experts: A framework for learning image priors." in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*, 2005, pp. 860–867.

[33] S. Awate and R. Whitaker, "Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005).* San Diego, CA, USA,: IEEE Computer Society, June 2005, pp. 44–51.

[34] G. Gerig, R. Kikinis, O. Kübler, and F. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Transactions on Medical Imaging*, vol. 11, no. 2, pp. 221–232, June 1992.

[35] J. Weickert, B. M. ter Haar Romeny, and M. A. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering." *IEEE Transactions on*
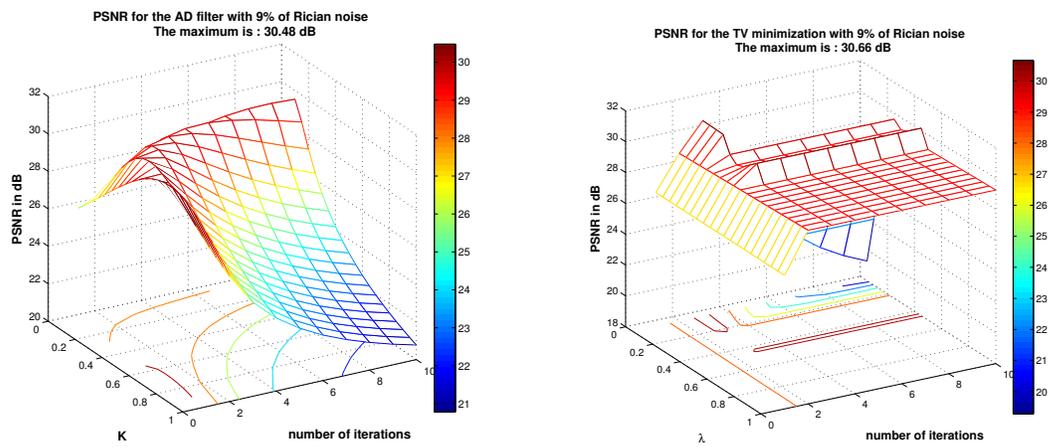
Fig. 17. Result for AD filter and TV minimization on phantom images with 9% of Rician noise. For AD filter K varies from 0.05 to 1 with step of 0.05 and the number of iterations varies from 1 to 10. For TV minimization $\lambda$ varies from 0.01 to 1 with step of 0.01 and the number of iterations varies from 1 to 10.


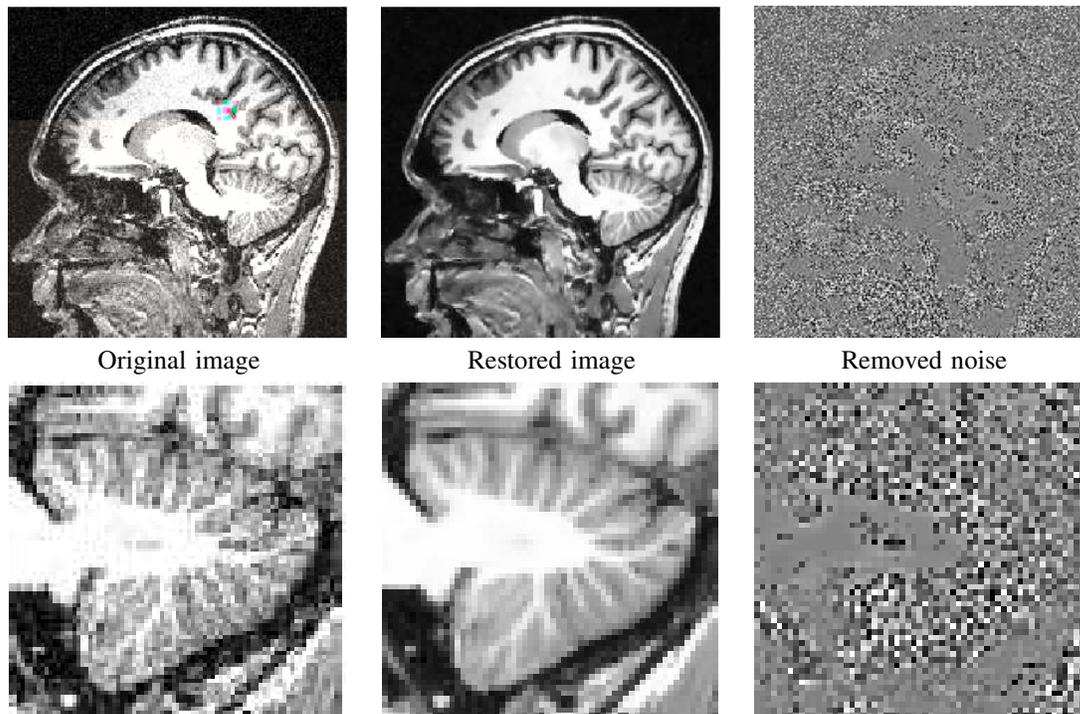
| Original image | Restored image | Removed noise |

Fig. 18. **NL-means filter on a real T1-w MRI.** Fully-automatic restoration obtained with the **Optimized Blockwise NL-means** filter on a 3 Tesla T1-w MRI data of $256^3$ voxels in less than 3 minutes on a Intel(R) Pentium(R) D CPU 3.40GHz with 2Go of RAM . From left to right: Original image, denoised image, and difference image with gray values centered on 128. The whole image is shown on top, and a detail is displayed on bottom.

*Image Processing*, vol. 7, no. 3, pp. 398–410, 1998.

[36] S. Keeling, "Total variation based convex filters for medical imaging," *Appl. Math. Comput.*, vol. 139, no. 1, pp. 101–119, 2003.

[37] W. Wong, A. Chung, and S. Yu, "Trilateral filtering for biomedical images," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Arlington, VA, USA, 15-18 April 2004*, 2004, pp. 820–823.

[38] R. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging ," *IEEE Transactions on Image Processing*, vol. 8, no. 10, pp. 1408–1419, October 1999.

[39] J. Wood and K. Johnson, "Wavelet packet denoising of magnetic resonance images: importance of Rician noise at low SNR." *Magn Reson Med*, vol. 41, no. 3, pp. 631–635, March 1999.

[40] S. Zaroubi and G. Goelman, "Complex denoising of MR data via wavelet analysis: application for functional MRI." *Magn Reson Imaging*, vol. 18, no. 1, pp. 59–68, January 2000.

[41] M. E. Alexander, R. Baumgartner, A. R. Summers, C. Windischberger, M. Klarhoefer, E. Moser, and R. L. Somorjai, "A wavelet-based method for improving signal-to-noise ratio and contrast in MR images." *Magn Reson Imaging*, vol. 18, no. 2, pp. 169–180, February 2000.

[42] P. Bao and L. Zhang, "Noise reduction for magnetic resonance images via adaptive multiscale products thresholding." *IEEE Trans Med Imaging*, vol. 22, no. 9, pp. 1089–1099, September 2003.

[43] A. Ogier, P. Hellier, and C. Barillot, "Restoration of 3D medical images with total variation scheme on wavelet domains (TVW)," in *Proceedings of SPIE Medical Imaging 2006: Image Processing*, San Diego, USA, February 2006.

[44] Y. Wang and H. Zhou, "Total variation wavelet-based medical image denoising," *International Journal of Biomedical Imaging*, vol. 2006, pp. Article ID 89 095, 6 pages, 2006.

[45] M. McDonnell, "Box-filtering techniques," *Computer Vision, Graphics, and Image Processing*, vol. 17, no. 1, pp. 65–70, Sept. 1981.
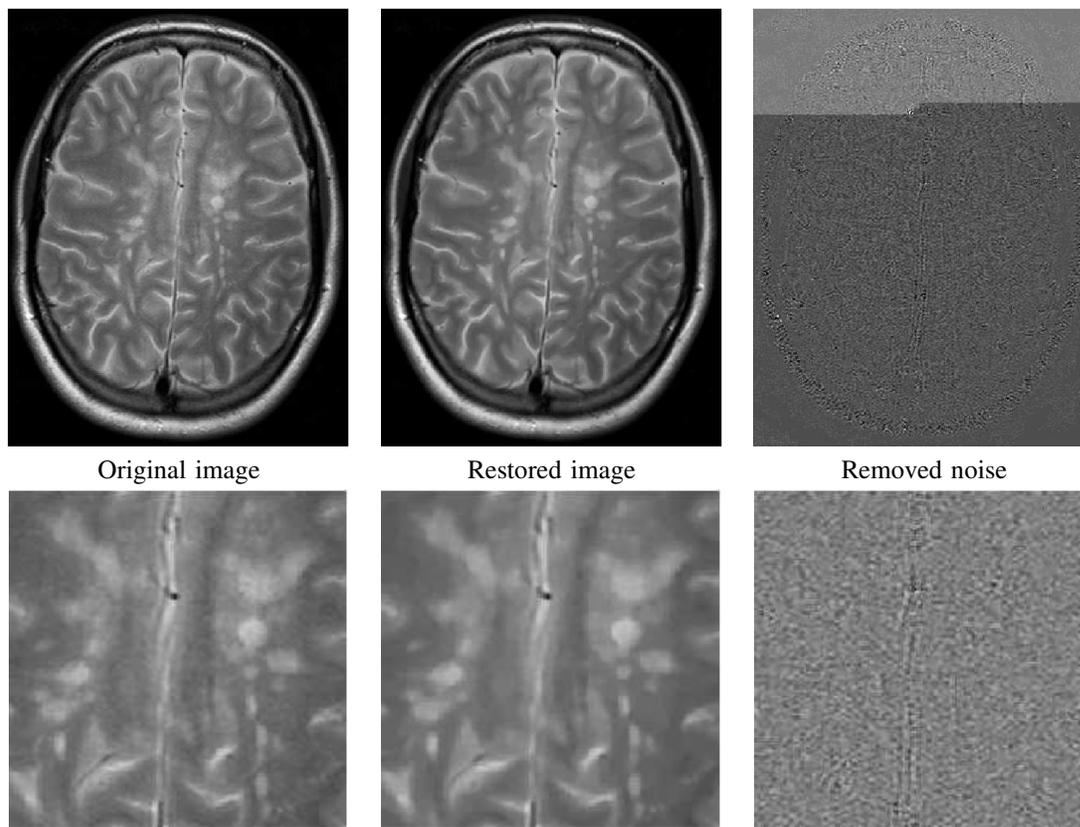
Fig. 19. **NL-means filter on a real T2-w MRI with MS.** Fully-automatic restoration obtained with the **Optimized Blockwise NL-means** filter on a 1.5T T2-w MRI data with MS lesions of $512 \times 512 \times 28$ voxels in less than 2 minute on a Intel(R) Pentium(R) D CPU 3.40GHz with 2Go of RAM. From left to right: Original image, denoised image, and difference image with gray values centered on 128. The whole image is shown on top, and a detail is exposed on bottom.

[46] L. Yaroslavsky, *Digital Picture Processing - An Introduction*, S. Verlag, Ed., 1985.

[47] J. Lee, "Digital image smoothing and the sigma filter," *Computer Vision, Graphics and Image Processing*, vol. 24, pp. 255–269, 1983.

[48] R. Coifman and D. Donoho, "Translation invariant de-noising," in *in Lecture Notes in Statistics: Wavelets and Statistics*, New York, February 1995, pp. 125–150.

[49] T. Gasser, L. Sroka, and C. Steinmetz, "Residual variance and residual pattern in nonlinear regression," *Biometrika*, vol. 73(3), pp. 625–633, 1986.

[50] J. Boulanger, C. Kervrann, and P. Bouthemy, "Adaptive spatio-temporal restoration for 4D fluoresence microscopic imaging," in *Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MIC-CAI'05)*, Palm Springs, USA, October 2005.

[51] A. Buades, B. Coll, and J.-M. Morel, "Image and movie denoising by nonlocal means," CMLA, Tech. Rep. 25, 2006.

[52] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via nonlocal means of similar neighborhoods," *Signal Processing Letters, IEEE*, vol. 12, no. 12, pp. 839–842, 2005.

[53] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magnetic Resonance in Medicine*, vol. 34, pp. 910–914, 1995.

[54] A. Macovski, "Noise in MRI," *Magnetic Resonance in Medicine*, vol. 36, pp. 494–497, 1996.

[55] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 460–489, 2005.

[56] E. Tadmor, S. Nezzar, and L. Vese, "A multiscale image representation using hierarchical $(\mathrm{BV}, L^2)$ decompositions," *Multiscale Model. Simul.*, vol. 2, no. 4, pp. 554–579, 2004.

[57] A. Buades, B. Coll, and J.-M. Morel, "A non local algorithm for image denoising." San Diego, USA: IEEE Computer Society, June 2005, pp. 60–65.

[58] A. Buades, B. Coll, and J. M. Morel, "The staircasing effect in neighborhood filters and its solution," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1499–1505, 2006.