Magnetic Resonance Imaging 57 (2019) 176–193

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.elsevier.com/locate/mri

Original contribution

Denoising of MR images using Kolmogorov-Smirnov distance in a Non Local framework

Fabio Baselice[a,*], Giampaolo Ferraioli[b], Vito Pascazio[a], Antonietta Sorriso[a]

[a] Dipartimento di Ingegneria, Università degli Studi di Napoli Parthenope, Napoli, Italy
[b] Dipartimento di Scienze e Tecnologie, Università degli Studi di Napoli Parthenope, Napoli, Italy

ARTICLE INFO

Keywords:
Image denoising
Magnetic Resonance Imaging
Statistical signal processing
Non Local Mean

ABSTRACT

Data coming from any acquisition system, such as Magnetic Resonance Imaging ones, are affected by noise. Although modern high field scanners can reach high Signal to Noise Ratios, in some circumstances, for example in case of very weak signals due to a specific acquisition sequence, noise becomes a critical issue that has to be properly handled. In the last years methods based on the so called Non Local Mean have proven to be very effective in denoising tasks. The idea of these filters is to find similar patches across the image and to jointly exploit them to obtain the restored image. A critical point is the distance metric adopted for measuring similarity. Within this manuscript, we propose a filtering technique based on the Kolmogorov-Smirnov distance. The main innovative aspect of the proposed method consists of the criteria adopted for finding similar pixels across the image: it is based on the statistics of the points rather than the widely adopted weighted Euclidean distance. More in details, the Cumulative Distribution Functions of different pixels are evaluated and compared in order to measure their similarities, exploiting a stack of images of the same slice acquired with different acquisition parameters. To quantitatively and qualitatively assess the performances of the approach, a comparison with other widely adopted denoising filters in case of both simulated and real datasets has been carried out. The obtained results confirm the validity of the proposed solution.

1. Introduction

The topic of noise reduction in case of Magnetic Resonance Imaging (MRI) data is considered within this manuscript. Noise is common to any kind of acquired data, nevertheless its statistical behavior depends on the characteristics of the acquisition system. The current trend of MRI scanner manufacturers is to produce systems with higher magnetic field in order to limit this problem. Nevertheless, noise is still a critical issue in case of low field scanners and in case of some specific kind of acquisition: methodologies for noise reduction (denoising) are worthy to be considered.

In literature several denoising filters have been proposed in the past years [1,2]. Among all, we cite linear (e.g. Wiener [3]), nonlinear(e.g. Anisotropic Diffusion [4]) and statistical filters (e.g. MAP [5,6]), the latter taking into account the noise statistical description. Recently, methodologies that find pixels with similar characteristics across the images have been proposed, i.e. the so-called Non Local Mean (NLM) approaches. The idea is that several similar regions (patches) found across the whole image can be exploited for a more effective denoising [7-9] or for image up-sampling [10]. NLM approaches divide the image in patches, and evaluate their similarity by comparing the textures, using a specific defined distance metric. In this manuscript, the focus is on the NLM filters group. Commonly, the Euclidean norm is adopted, defined as the sum of the squared differences between textures. The setting of a proper distance metric is a critical step to achieve good denoising performances and to limit the amount of artifacts that affect the filtered image.

Several operators have been defined for measuring the distance between patches with the aim of better identifying similar ones, reducing the amount of mistakes [11-13]. An analysis of different available distance measures can be found in Foi and Boracchi [14]. Recently, similarity metrics based on statistical properties of the patches have been proposed. In particular, the ones based on the distance between the Cumulative Distribution Functions (CDFs) of the image patches have proved to be effective [15,16]. These methods evaluate the CDF difference by means of Kolmogorov-Smirnov (KS) test or distance. In this framework, Rajan et al. in [17] proposed to adopt KS test for filtering MRI. The authors test the similarity of the statistical distribution

* Corresponding author.
E-mail addresses: fabio.baselice@uniparthenope.it (F. Baselice), giampaolo.ferraioli@uniparthenope.it (G. Ferraioli), vito.pascazio@uniparthenope.it (V. Pascazio), antonietta.sorriso@uniparthenope.it (A. Sorriso).

https://doi.org/10.1016/j.mri.2018.11.022
Received 2 July 2018; Received in revised form 5 November 2018; Accepted 24 November 2018

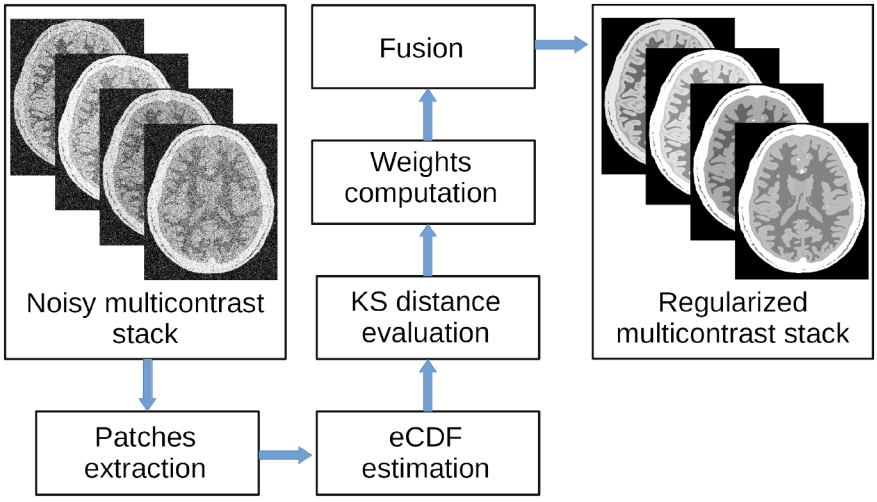

Fig. 1. KSNLM processing chain.

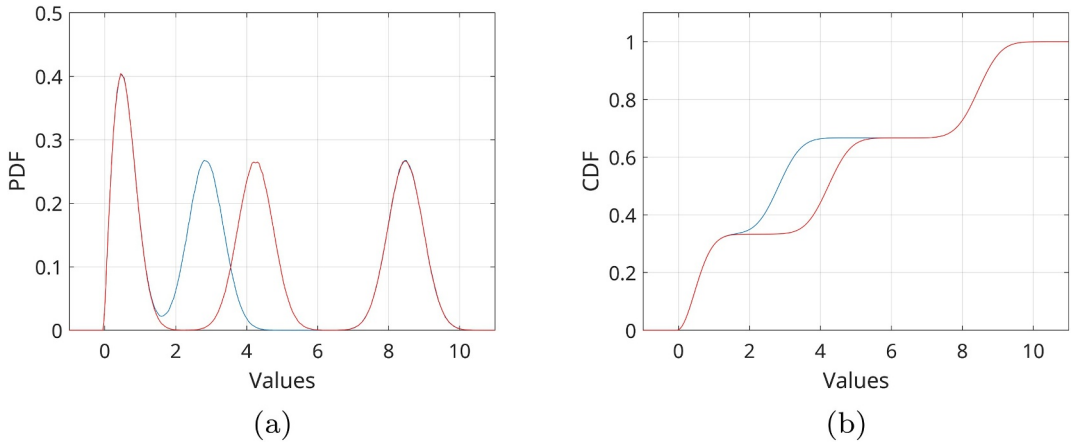

Fig. 2. Normalized histograms (a) and estimated CDFs (b) computed at two locations of a 3 MR images stack. The values do not perfectly match, and a deviation between the two CDFs appears.

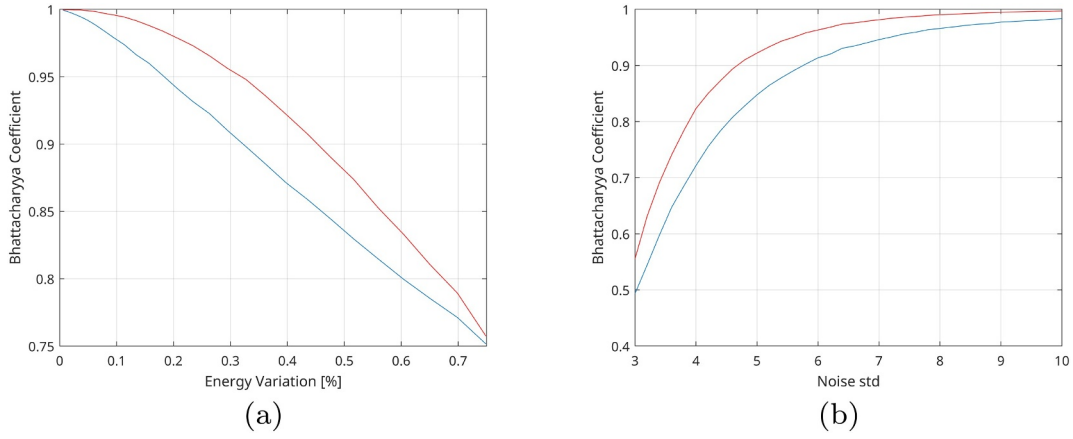

Fig. 3. BC between the two D^{KS} distributions (blue line) and the two D^{E} distributions (red line) computed for different energy variations (a) and for different noise standard deviations (b).

of the difference between two patches with a Normal one with known variance.

The algorithm proposed in this manuscript is an evolution of the approaches proposed in [15] and [16]. The idea is to compute the CDF of a given pixel exploiting the values it assumes in different MR acquisitions, instead of considering its spatial neighbors (i.e. the patch). As a matter of fact, in clinical analysis several MR images of the same slice are available, acquired with different scanner parameters in order to acquire the so-called multi-contrast images (e.g. Proton Density weighted, T2 Weighted, T1 Weighted, Diffusion Weighted, fluid attenuated inversion recovery) [1,9]. As it will be shown, working with different images of the same slice allows to consider small patches,

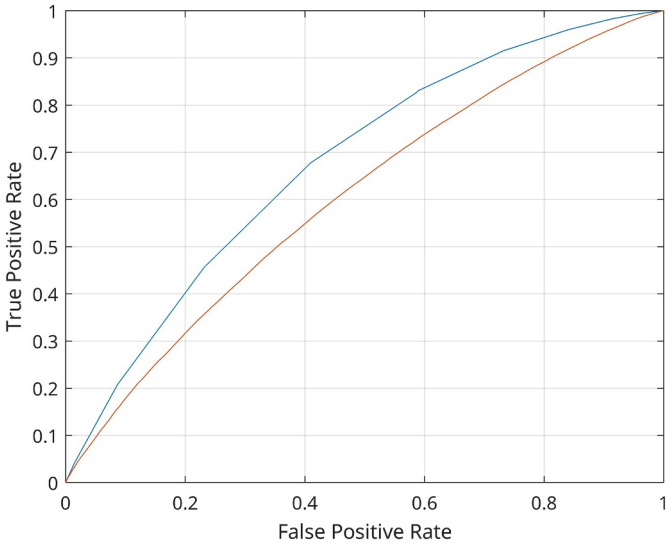

Fig. 4. ROC curves computed by applying a threshold to the KS distance (blue line) and to the Euclidean distance (red line). The false alarm probability and the detection probability are on the horizontal and on the vertical axes, respectively.

Table 1
Filter parameters adopted for comparison.

Method	Parameters
AD	$n_{Iter} = 4$, $\Delta_t = 1/7$, $k = 70$, $c = 1/[1 + (\nabla_I k)^2]$
BM3D	$\sigma = \sigma_{noise}$
LMMSE	$Ws = 9 \times 9$
NLM	$W_{search} = 21 \times 21$, $W_{similarity} = 7 \times 7$, $h = 0.05$
MAP	-
MNLM	$W_{search} = 21 \times 21$, $W_{similarity} = 3 \times 3$, $h = \sqrt{2}\sigma_{noise}$

ideally 1×1 in case a sufficient number of acquisitions, that can greatly reduce the arise of artifacts in the filtered image. In particular, the ghost effect that often characterizes NLM filtered image, i.e. signal-like patches that appear in flat areas, originated by random noise and reinforced through the patch selection process [18], can be greatly limited.

The methodology requires large data availability in order to perform a robust CDF estimation. In the case of ultrasound data considered in Baselice [15] and in case of radar images considered in Ferretti et al. [16], several images of the same scene (the frames of an ultrasound video and the radar acquisitions, respectively) are jointly processed under the hypothesis of stationary, i.e. constant signal and different noise realization. Subsequently, the KS distance is implemented for evaluating the similarity across the image. In case of MRI data, the hypothesis of several images available with the same signal and with different noise realizations is generally not met. However, several MR images of the same slice are available, although acquired with different scanner parameters. Therefore a peculiar approach has been developed to handle heterogeneous MRI images.

The proposed approach, namely Kolmogorov-Smirnov Non Local Mean (KSNLM), has several innovative aspects. First of all, it is different from the so called 3D MRI denoising filters proposed in literature [7,19]. These methodologies have been developed in order to work with different slices (i.e. 3D data), acquired with the same image sequence, and to jointly restore the whole imaged volume. The proposed approach makes use of several images of the same slice, obtained using different acquisition sequences or parameters (e.g. Proton Density weighted, T2 Weighted). Moreover, the adopted KS distance shows interesting capabilities in discriminating slightly different patches compared to the Euclidean one, which is widely adopted in other NLM methodologies proposed in literature. Compared to the denoising approach proposed in Rajan et al. [17], based on the KS test, the proposed KSLNM, based on the KS distance, is able to jointly exploit and to simultaneously filter a stack of images, instead of a single one. Finally, the algorithm is not limited to specific noise distribution in the acquisition model and thus it does not require the knowledge of its statistical description.

The paper is organized as the following: In Section 2 NLM theory is briefly reported; in Section 2 the proposed methodology is presented. The comparison framework is described in Section 3, while the Simulated and Real data results are reported in Sections 4 and 5, respectively.

2. Material and methods

2.1. Non Local approaches

Let us consider a single MR image. The pixel at the location p can be modeled as:

$$y_p = y_p^R + iy_p^I = x_p + w_p, \qquad y_p, x_p, w_p \in \mathbb{C} \tag{1}$$

where $x = x^R + ix^I$ is the complex noise free signal, $w = w^R + iw^I$ is the circular complex additive noise with zero mean and σ^2 variance and $p = 1,2,\ldots,P$, with P being the number of pixels of the image. The modulus of the measured signal $|y_p|$ is Rician distributed [5], and it will be characterized by the following statistical distribution:

$$f(|y_p|; |x_p|) = \frac{|y_p|}{\sigma^2} exp\left(\frac{-|y_p|^2 - |x_p|^2}{2\sigma^2}\right) I_0\left(\frac{|y_p||x_p|}{\sigma^2}\right) \tag{2}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind with order zero. In this case, the additive model is no more correct. However, in case of high Signal to Noise Ratio (SNR), it can be considered valid [9,20]:

$$|y_p| = |x_p| + z_p, \quad y, x \in \mathbb{C}, \quad z \in \mathbb{R} \tag{3}$$

where the noise term z can be assumed to be Gaussian distributed, i.e. $\mathcal{N}(\mu_z, \sigma^2)$. In this case, the probability density function of $|y_p|$ can be written as:

$$f(|y_p|; |x_p|) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-(|y_p| - |x_p| - \mu_z)^2}{2\sigma^2}\right] \tag{4}$$

Therefore, depending on the SNR of the acquired data, the Rician model of Eq. (2) or the additive Gaussian model of Eq. (4) can be assumed.

Considering a pixel at position p and its surrounding domain $\mathbf{\Delta}_p$ (the neighborhood of p), it is possible to define $\mathbf{y}_p = \{|y_n|\}$, with $n \in \mathbf{\Delta}_p$, and $\mathbf{x}_p = \{|x_n|\}$, with $n \in \mathbf{\Delta}_p$ as the noisy and the noise free patches, respectively.

The NLM methodology requires the evaluation of the distances between patches in order to find the most similar ones. For this task, the Euclidean distance is widely adopted. Given two pixels p and q and their surrounding patches $\mathbf{y}_p$ and $\mathbf{y}_q$, the Euclidean distance is given by:

$$D_{p,q}^E = \|\mathbf{y}_p - \mathbf{y}_q\|_2^2 \tag{5}$$

Once the distance $D_{p,q}^E$ is computed, a fusion step is implemented by properly merging pixels with lower distances in order to obtain the regularized pixel values at the location p. More in detail, the estimated value $|\hat{x}_p|$ is computed according to:

$$|\hat{x}_p| = k \sum_{q=1}^{P} |y_q| \exp\left(-\frac{D_{pq}^E}{h^2}\right) \tag{6}$$

where k is a normalization factor and h is the filtering parameter controlling the decay of the exponential function [14]. The latter

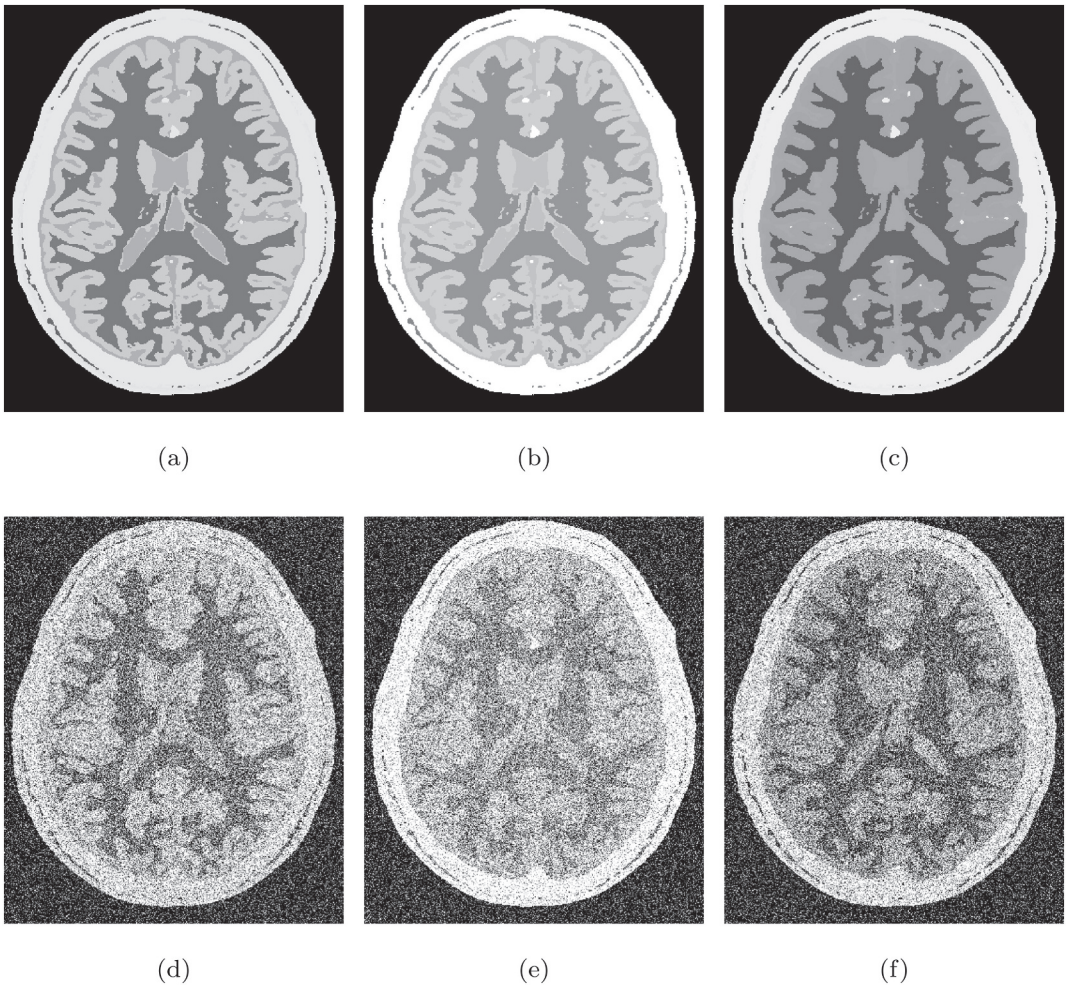

Fig. 5. First row: three noise free images acquired with different imaging parameters. Second row: the same images in the noisy case with SNR = 15 dB case. The images simulate an axial acquisition of a head slice.

parameter is adopted for handling excessively large D^E, and it is often set equal to the noise standard deviation. In other words, it is a normalization factor for the distance. From Eq. (6) it is evident that the denoising effectiveness greatly depends on how many similar patches are found across the image. If most of the locations q are characterized by a very large distance, the denosing effect will be limited.

2.2. Proposed approach

Let us consider L co-registered MR images of the same slice acquired with different sequences and imaging parameters. It is possible to define $\mathbf{Y}_p$ the vector collecting the stack of the L observed patches $\mathbf{y}_p$. The idea is to exploit the statistical behavior of the vector $\mathbf{Y}_p$ for finding similar regions: it can be considered as a vector collecting the realizations of a random variable, and thus it can be described by the Cumulative Distribution Function (CDF) $F_{\mathbf{Y}_p}$. Let us consider two pixels at locations p and q. The aim is to determine whether they are similar or not. The statistical distribution of the noise term is assumed to be the same. If the noise free patches centered in p and q are equal in all the L images (i.e. $\mathbf{X}_p = [\mathbf{x}_p(1),\mathbf{x}_p(2),\cdots,\mathbf{x}_p(L)] = \mathbf{X}_q = [\mathbf{x}_q(1),\mathbf{x}_q(2),\cdots,\mathbf{x}_q(L)]$), it is expected that the two CDFs $F_{\mathbf{Y}_p}$ and $F_{\mathbf{Y}_q}$ will be similar. On the contrary, the two CDFs will not show the same behavior. The measure of the distance between the two CDFs can thus be considered as a similarity metric between patches. The first step consists of estimating the empirical CDFs (eCDFs) from the data by means of the Kaplan-Meier nonparametric estimator [21]. Subsequently, the Kolmogorov-Smirnov distance can be adopted for evaluating such similarity [22]:

$$D_{p,q}^{KS} = \max\left[|F_{\mathbf{Y}_p} - F_{\mathbf{Y}_q}|\right] \tag{7}$$

Note that the $D_{p,q}^{KS}$ value does not depend on the image index l. Once the distance is computed, the estimation of the regularized pixel at location p for each of the L images is carried out by averaging all the location with a distance lower than the threshold T:

$$|\hat{x}_p(l)| = k \sum_{q=1}^{P} |y_q(l)| u(T - D_{pq}^{KS}), \quad l = 1, \cdots, L \tag{8}$$

where $u(\cdot)$ is the unitary step function and k is a normalizing factor. The threshold T has a similar usefulness to parameter h of Eq. (6): it helps in controlling the filter strength. Differently from Eq. (6), as the D^{KS} is a distance between probabilities, it is defined within the [0,1] range, the normalization is not needed. The processing steps of the methodology, schematized in Fig. 1, can be summarized as follows:

- Acquire L heterogeneous MR images;
- Estimate the CDFs for each location p by jointly exploiting the different images of the stack;
- Compute the distances D^{KS} between each location and all the others based on the estimated CDFs;
- Fuse similar pixels according to Eq. (8) for each of the L images.

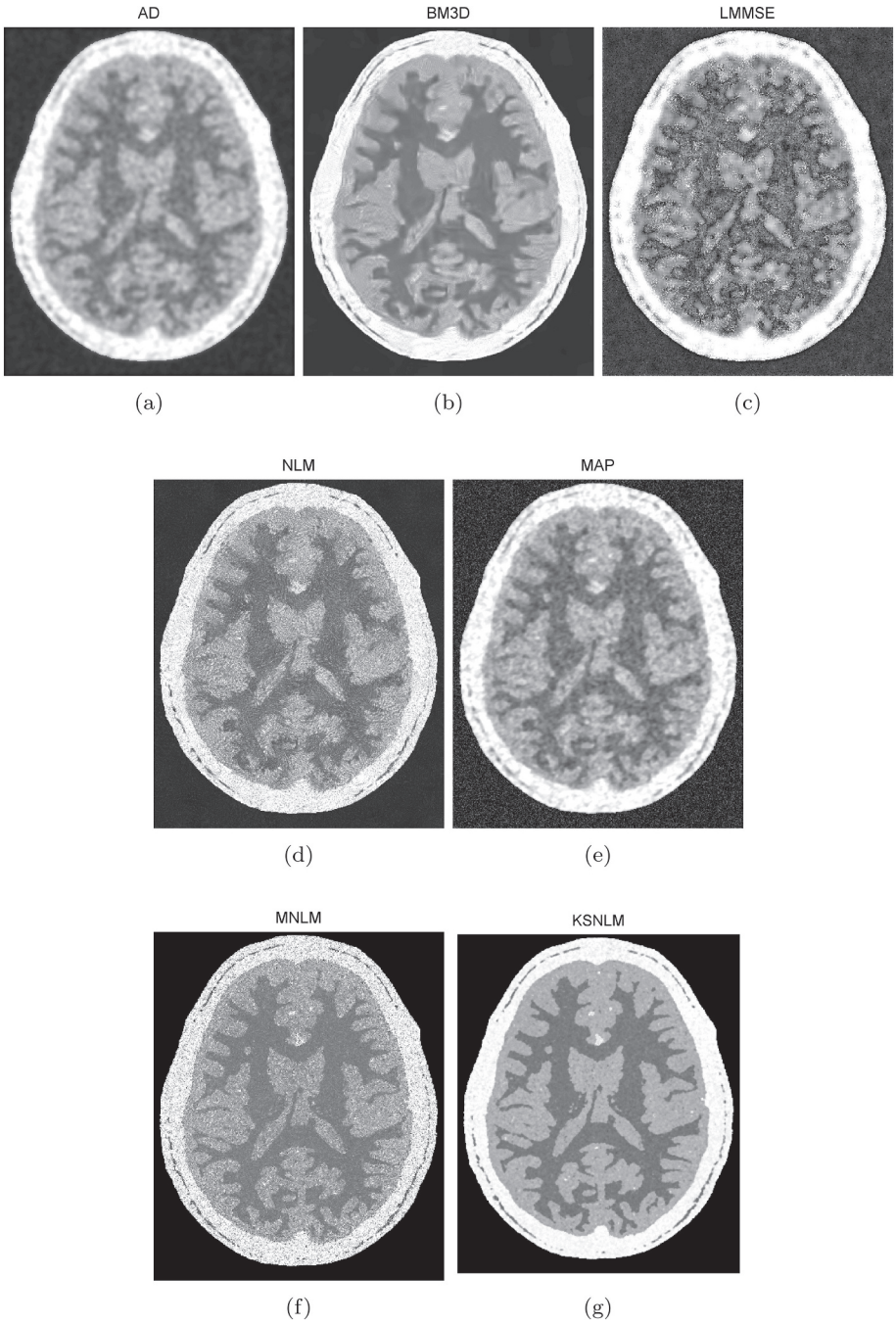

Fig. 6. Simulated dataset filtering results: Anisotropic Diffusion (a), Block Matching 3D (b), Linear Minimum Mean Square Error (c), Non-Local Means (d), Maximum A Posteriori (e), Multicontrast NLM (f), Kolmogorov-Smirnov NLM (g) filters. Results refer to SNR = 15 dB.

The output of the methodology are L filtered images. It has to be underlined that the methodology will fail in case the vector $\mathbf{X}_q$ contains a permutation of $\mathbf{X}_p$ values. However, such condition is very unlikely to appear.

In case of low SNR, i.e. when the Gaussian model cannot be assumed, the weighted mean of Eq. (8) will produce a bias in the regularized data [23]. Such bias is proportional to σ and can be compensated in a post-processing step [24].

According to the proposed approach, the pixels at location q that have the same noise free component as location p on the whole stack will strongly contribute to the regularization of the latter. On the contrary, if the noise free parts of the pixels at location q differ from the ones at position p in one or more images of the stack, the contribution of the former in regularizing the latter will be limited. To better clarify this point, we can make use of a simple example. Let us consider a stack of $L = 3$ images and two 1×1 patches whose noise free values are $\mathbf{X}_p = [1,4,9]$ and $\mathbf{X}_q = [1,3,9]$. In a Monte Carlo simulation, noise is added in order to generate the observed patches $\mathbf{Y}_p$ and $\mathbf{Y}_q$. In Fig. 2, the normalized histograms and the estimated CDFs related to the two locations p and q are reported. The normalized histograms show that one peak of the blue line (related to the location q) is not aligned with the corresponding peak of the red line (related to the location p). By considering the CDFs (Fig. 2b), a deviation appears between the two curves, which is measured by the KS distance $D_{p,q}^{KS}$. For each of the 3 images composing the stack, the pixels at the location q will contribute with a sensibly small weight (due to the large distance) in the

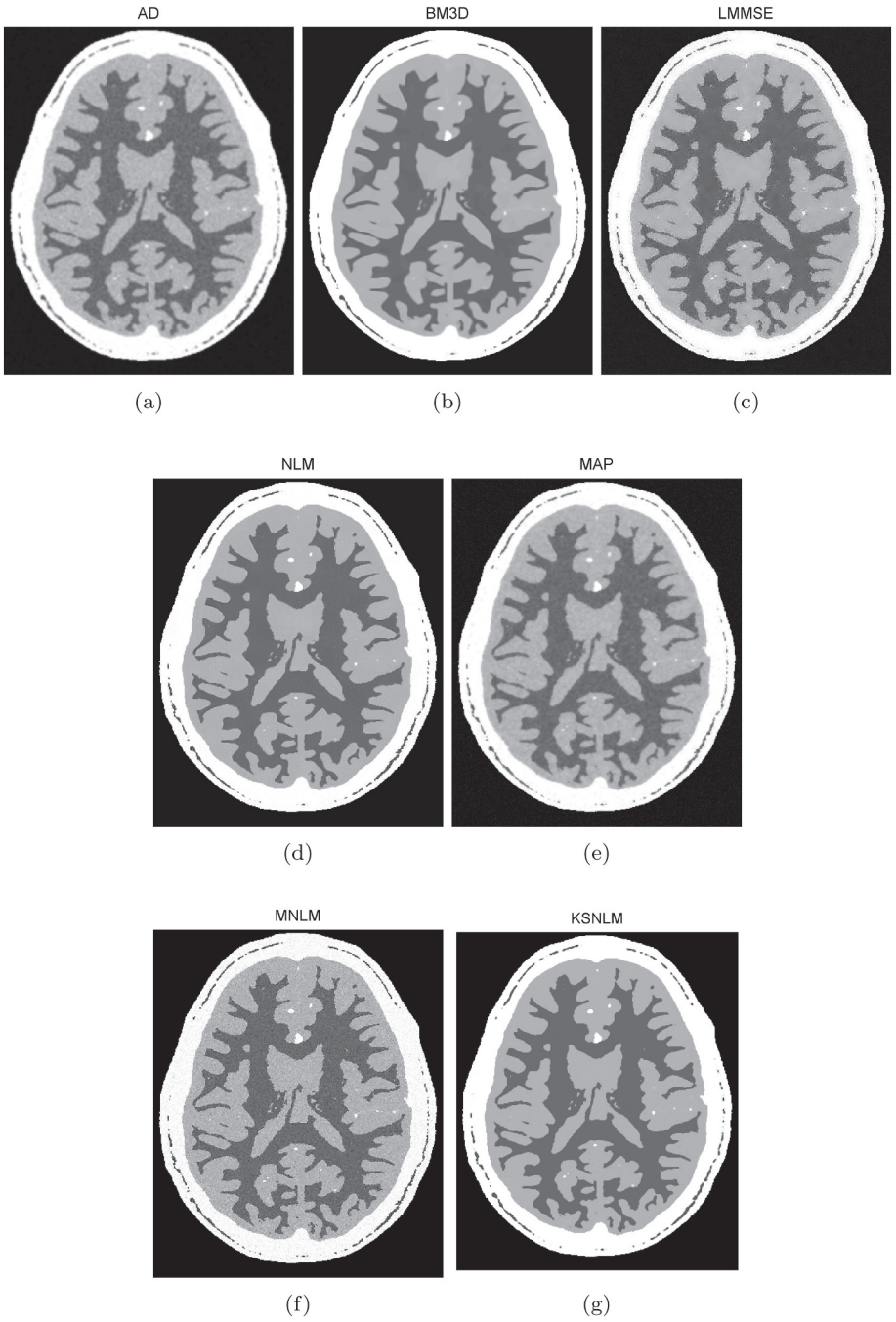

Fig. 7. Simulated dataset filtering results: Anisotropic Diffusion (a), Block Matching 3D (b), Linear Minimum Mean Square Error (c), Non-Local Means (d), Maximum A Posteriori (e), Multicontrast NLM (f), Kolmogorov-Smirnov NLM (g) filters. Results refer to SNR = 30 dB case.

computation of Eq. (8).

Given the previous considerations, the proposed methodology based on the KS distance is particularly sensitive to discriminate small variations between patches. This will allow to discard in the fusion step the patches characterized by slightly different textures.

In order to analyze the sensitivity of the proposed KS distance in measuring similarity with respect to the classical Euclidean one, a Monte Carlo simulation with 10^5 iterations is set up. Three stacks, namely A, B and C, of 8 constant patches made of 3×3 pixels, each patch with a different amplitude level, are considered. Each stack is corrupted by a zero mean additive Gaussian noise, with a standard deviation set with respect to the maximum amplitude value of the stack. The stacks A and B only differ in the noise realizations, while one of the patches of the stack C is different from A and B. The change of the patch is such that the energy of the stack C (i.e. the sum of the squared modulus), compared to the others, is slightly increased.

The KS distances $D^{KS}(A,B)$ and $D^{KS}(A,C)$ between A and B (same energy, different noise) and A and C (different energy and noise) are computed. The idea is that if a slight change appears (one patch is different in the C stack compared to the other stacks), the distributions of the distances $D^{KS}(A,B)$ and $D^{KS}(A,C)$ should not overlap anymore. For a fixed change, the larger is the shift between the distributions, the higher is the sensitivity. Clearly, the same can be stated for the euclidean distances $D^{E}(A,B)$ and $D^{E}(A,C)$. To evaluate the overlapping (or the shifting) the Bhattacharyya Coefficient (BC) between $D^{KS}(A,B)$ and $D^{KS}(A,C)$ and between $D^{E}(A,B)$ and $D^{E}(A,C)$ is computed. The BC is a

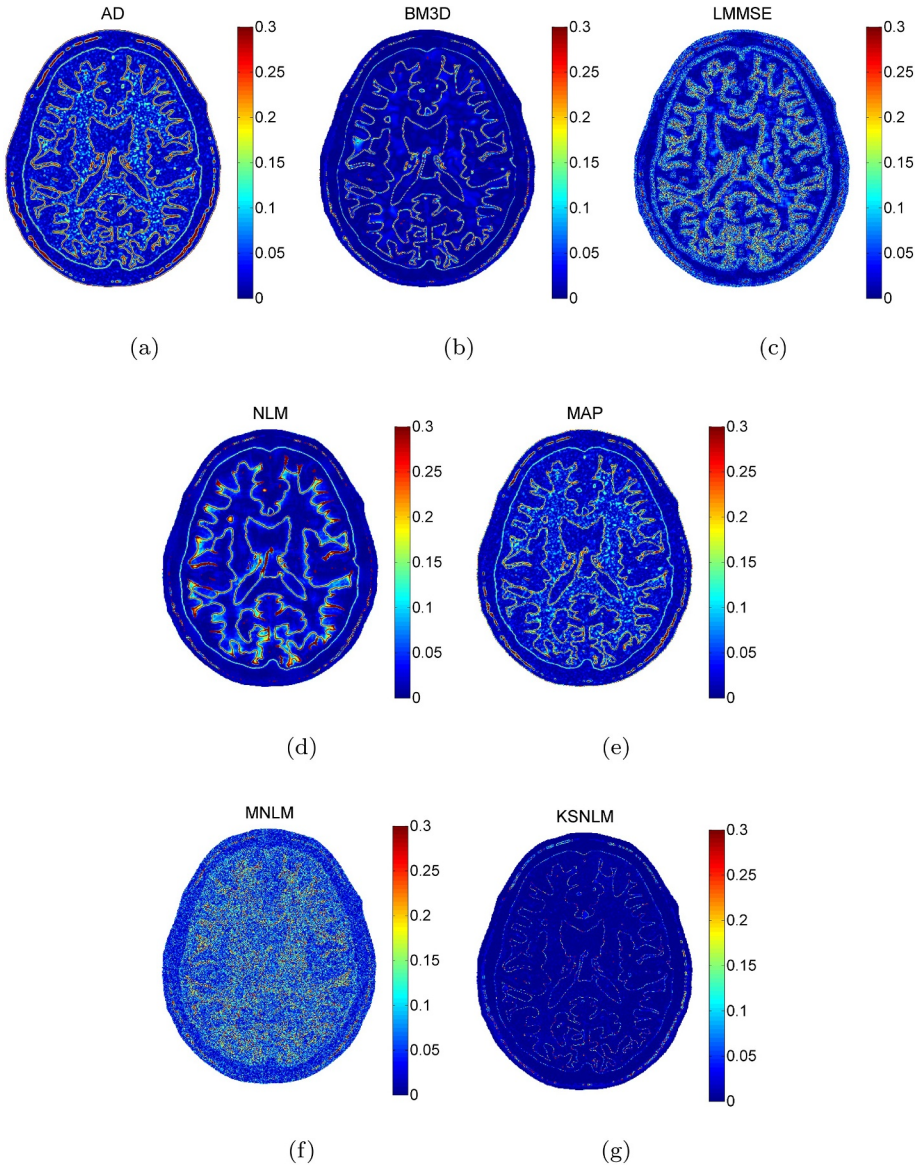

Fig. 8. Simulated dataset, residual error: Anisotropic Diffusion (a), Block Matching 3D (b), Linear Minimum Mean Square Error (c), Non-Local Means (d), Maximum A Posteriori (e), Multicontrast NLM (f), Kolmogorov-Smirnov NLM (g) filters. Results refer to SNR = 30 dB case.

commonly adopted index able to measure the similarity between images probability distribution. In order to have a high sensitivity, small texture perturbations should correspond to large BC index variations. The BC assumes values in the [0,1] range. In case of perfect overlapping BC is equal to 1, while in case of non-overlapping BC is equal to 0. The analysis is conducted for different noise standard deviations and for different energy variations. The results are reported in Fig. 3.

In Fig. 3 (a), the BC values between $D^{KS}(A,B)$ and $D^{KS}(A,C)$ (blue line) and between $D^{E}(A,B)$ and $D^{E}(A,C)$ (red line) are reported, fixing the noise standard deviation to 7% of the maximum value of the stack and modifying the energy variation between 0% and 0.7%. In Fig. 3 (b), the BC values between $D^{KS}(A,B)$ and $D^{KS}(A,C)$ (blue line) and between $D^{E}(A,B)$ and $D^{E}(A,C)$ (red line) are reported, fixing the energy variation to 0.2% and varying the noise standard deviation from the 3% to the 10% of the maximum value of the stack. The higher sensitivity of the KS distance compared to the Euclidean one in discriminating slightly different patches is evident. Considering Fig. 3 (a), in case there is no change in the patches, BC correctly assumes in both cases (red and blue lines) its maximum value (BC = 1). On the contrary, when the energy changes, the BC between $D^{KS}(A,B)$ and $D^{KS}(A,C)$ (blue line) is always smaller than the red line. This is an index of a larger separation and of a better discrimination capability. The same consideration can be inferred for different noise levels, as in Fig. 3 (b). The BC values of the KS distances are again smaller with respect to the Euclidean case.

A further analysis on the sensitivity can be conducted based on the false alarm and detection probabilities of the two distances. Let us consider the same Monte Carlo simulation, but fixing the energy change to 0.2% and the noise standard deviation to 7%. If we set a threshold T, it is possible to define a correct detection if, in the hypothesis of stacks A and B, the computed distance is smaller than the threshold, and a false alarm if, in the hypothesis of stacks A and C, the computed distance is smaller than T. By varying the value of T, the Receiver Operating Characteristic (ROC) curve is computed and reported in Fig. 4 for the KS (blue line) and the Euclidean (red line) cases. It is evident that, being the farthest from the bisector, the KS distance (the blue line) is more effective in discriminating slightly different patches compared to the Euclidean one (the red one).

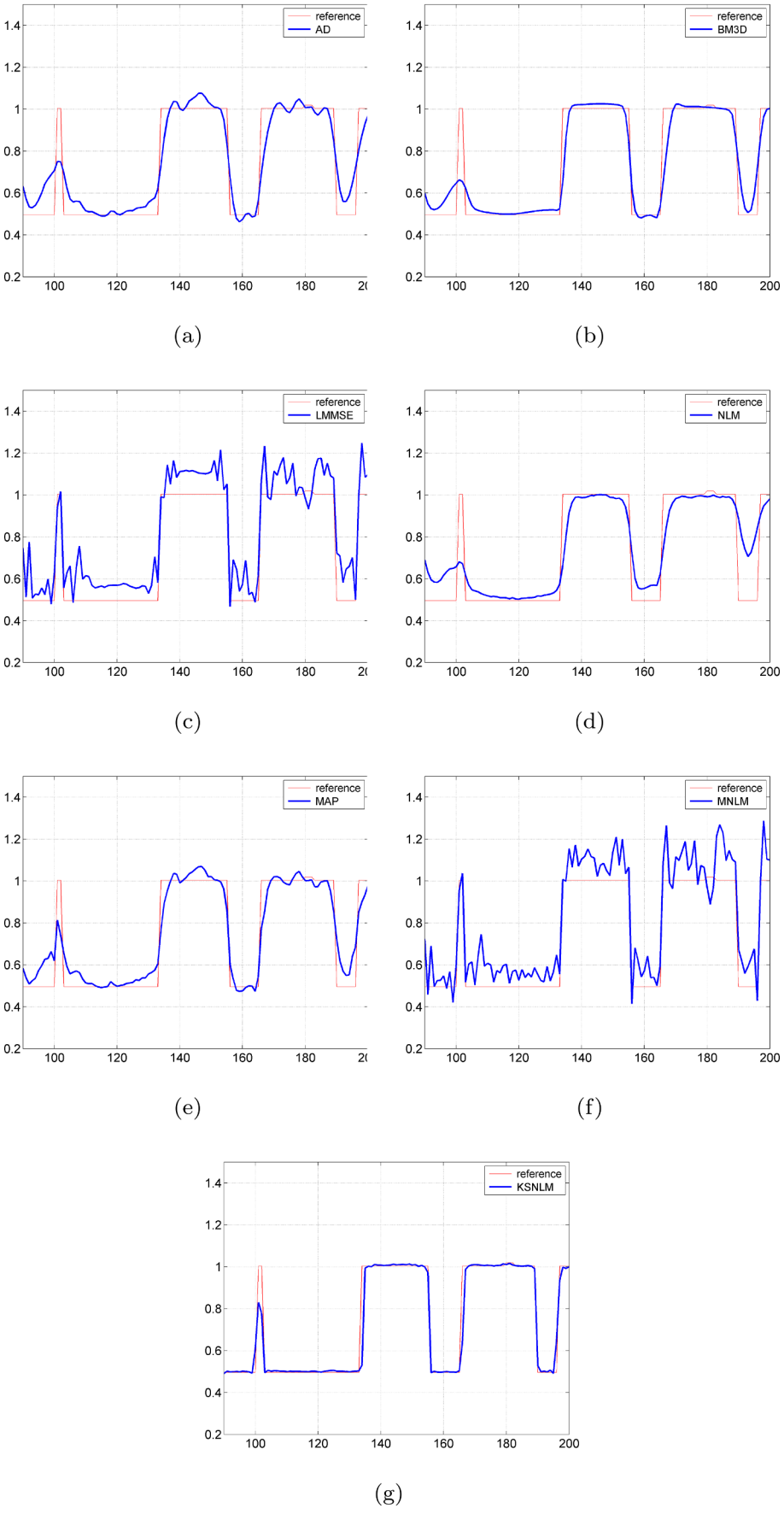

Fig. 9. Simulated dataset, filtered and reference image (one row): Anisotropic Diffusion (a), Block Matching 3D (b), Linear Minimum Mean Square Error (c), Non-Local Means (d), Maximum A Posteriori (e), Multicontrast NLM (f), Kolmogorov-Smirnov NLM (g) filters. Results refer to SNR = 30 dB case.

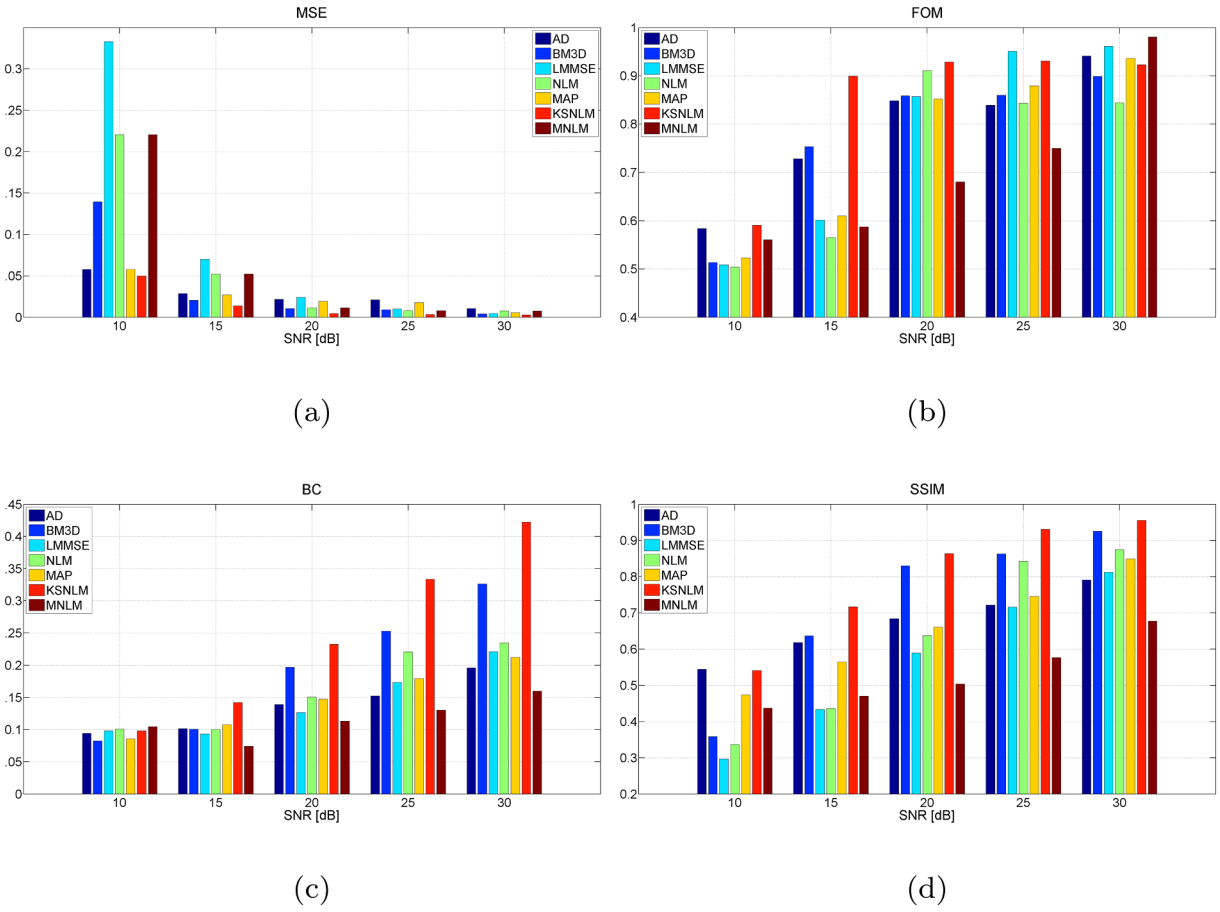

Fig. 10. Simulated dataset: MSE (a), FOM (b), BC (c) and SSIM (d) metrics for all the considered filters in case of different SNR values.

Table 2
Performances comparison in case of stacks composed of a different number of images.

	Number of acquired images			
	1	2	4	8
MSE	0.031	0.012	0.006	0.004
FOM	0.46	0.64	0.92	0.94
BC	0.13	0.16	0.19	0.24
SSIM	0.51	0.63	0.74	0.86

Table 3
Performances comparison for Gaussian and Rician distributions.

	MSE		SSIM	
	Gaussian	Rician	Gaussian	Rician
AD	0.0210	0.0367	0.7223	0.7002
BM3D	0.0092	0.0247	0.8630	0.8368
LMMSE	0.0101	0.0098	0.7168	0.7368
NLM	0.0079	0.0352	0.8424	0.7917
MAP	0.0179	0.0358	0.7460	0.6989
MNLM	0.0325	0.0435	0.5758	0.5285
KSNLM	0.0034	0.0036	0.9309	0.9309

3. Results

3.1. Comparison framework

The proposed KSNLM methodology has been compared to other widely adopted denoising filters, and the achieved performances have been analyzed in simulated and real datasets. In particular, the following approaches have been included in the comparison:

- Anisotropic Diffusion (AD), an approach based on Partial Differential Equation [4];
- Block Matching 3D (BM3D), which belongs to the NLM family [25];
- Linear Minimum Mean Square Error (LMMSE) estimator developed for MRI data [26];
- Non Local Means (NL-means) algorithm, the classical approach exploited by Buades et al. [27];
- Maximum-A-Posteriori (MAP), a statistical filter based on Bayesian approach [28].
- Multicontrast NLM (MNLM), a Non Local Means filter developed for multicontrast MR image stacks [9].

For each methodology, the parameters adopted in the comparison have been found by manually tuning the optimal configuration suggested by the authors. The values are reported in Table 1.

The KSNLM methodology requires the estimation of the CDF of the data, thus a proper number of samples is mandatory. Based on the estimation accuracy, the complexity and the patch dimension, a number of points between 50 and 100 is recommended. Therefore, since available stacks are often smaller than this size, 3×3 patches can be considered. Such dimension allows an effective CDF estimation and avoids block-like effects typical of spatial constraints [29]. As reported in Section 2.2, in order to identify similar pixels across the image a threshold based criterion that minimizes the probability of false alarms is adopted. In all the results, a value of threshold T equal to 0.05 has been set. Such a value has turned to be effective in all the test cases

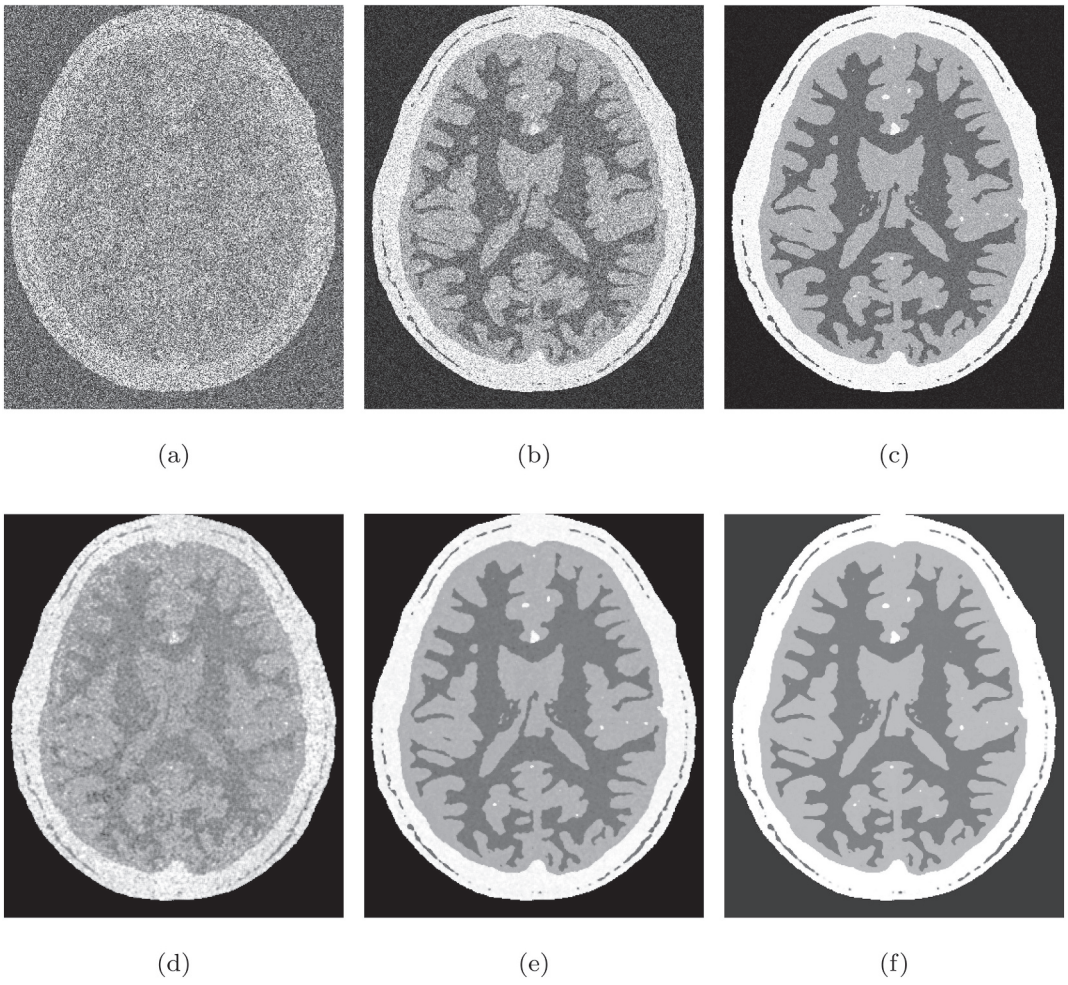

Fig. 11. Simulated dataset filtering results in case of Rician distribution: noisy data (a,b,c) and filtered images (d,e,f) in case of SNR = 10 dB, 20 dB and 30 dB, respectively.

considered in the following. For the following analyses, a searching window of 101 × 101 pixels is adopted.

Concerning the other methods used in the comparison, the parameters have been set by tuning the configuration suggested by the authors in order to find the best trade-off between noise reduction and edges preservation (i.e. between under- and over-regularization).

For the simulated dataset, a quantitative evaluation has been performed based on the following indexes.

- *Mean-Squared Error* (MSE), defined as:

$$MSE = E[(\hat{x} - x)^2] \tag{9}$$

where $E[\cdot]$ computes the expected value. The MSE is an effective parameter for measuring the quality of proposed algorithm, comparing the estimated image $\hat{x}$ and reference image x.

- *Figure Of Merit* (FOM), proposed by Pratt for measuring the deviation of edge points from the ideal edge [30], defined as:

$$FOM = \frac{1}{\max(I_I, I_A)} \sum_{i=1}^{I_A} \frac{1}{1 + \delta e^2(i)} \tag{10}$$

where I_A is the actual number of detected edge points, I_I is the number of edge points on the ideal edge, $e(i)$ is distance between edge point and ideal edge and δ is a scaling constant. FOM is effective in measuring the detail preservation capability of a filter.

- *Bhattacharyya Coefficient* (BC), which is an index used to measure the similarity between images probability distribution [31]. BC is defined as

$$BC = \sum_{x \in \mathcal{X}} \sqrt{p_{\hat{X}}(x) p_X(x)} \tag{11}$$

where $\mathcal{X}$ is the interval of the possible image values, $p_{\hat{X}}(\cdot)$ and $p_X(\cdot)$ are the probability density functions of the filtered and of the reference images, respectively.

- Structural Similarity Index (SSIM) is another metric used in image processing to measure the similarity and for predicting the perceived quality [32]. The SSIM mathematical formulation is the following:

$$SSIM = [l(\hat{x}, x)]^{\alpha} \cdot [c(\hat{x}, x)]^{\beta} \cdot [s(\hat{x}, x)]^{\gamma} \tag{12}$$

This index assesses the visual impact by calculating three characteristics of image: the luminance $l(\cdot)$, the contrast $c(\cdot)$ and the structural $s(\cdot)$ term.

Clearly, for the real data sets, it is not possible to provide a quantitative evaluation, due to the lack of a true reference image. The evaluation will be only qualitatively, based on the visual inspection.

3.2. Simulated data results

3.2.1. Gaussian noise

In order to adopt a realistic human brain phantom, the "BrainWeb: Simulated Brain Database", available at http://brainweb.bic.mni.

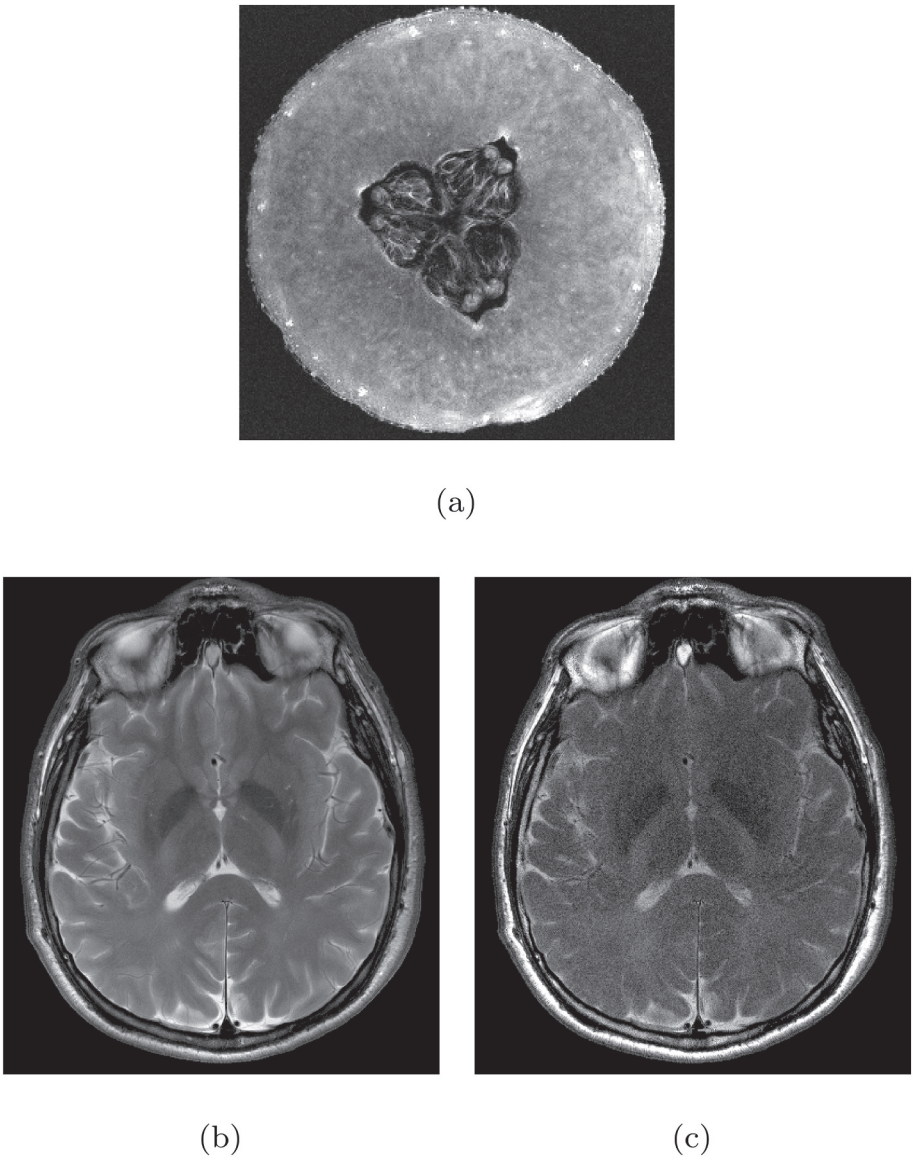

Fig. 12. Real 3T datasets: melon slice (a) and two clinical head acquisitions (b, c). All the scans adopted SE sequences. The sequence timings for the reported images are T_R = 3500ms and T_E = 110ms (a), T_R = 700ms and T_E = 140ms (b) and T_R = 2080ms and T_E = 100ms (c).

mcgill.ca/brainweb/, has been considered [33]. The phantom simulates an axial slice of the head, made of 10 gray levels (i.e. 9 tissues and the background) and 434 × 362 pixels. Based on this phantom, several weighted Spin Echo (SE) MR images have been generated with different combinations of Echo Times (T_E) and Repetition Times (T_R). Gaussian noise has been added in order to achieve Signal to Noise Ratio (SNR) values varying from 10 to 30 dB. In order to compare the considered noise level with those provided by the BrainWeb database, we report that SNR of 10, 20 and 30 dB correspond to noise levels equal to 54%, 17% and 7%.

Globally, 8 images with the following T_E and T_R combinations have been generated: (20, 500), (130, 900), (50, 1500), (70, 2500), (100, 3000), (90, 5000), (80, 3500), (110, 2000) [ms]. Among all, the images related to combinations (20, 500), (70, 2500) and(110, 2000) are reported in Fig. 5, both in the noise-free (top row) and the noisy (SNR equal to 15 dB) (bottom row) cases. The filtering methodologies have been applied to the whole stack, nevertheless in the following only the results related to the image with T_E= 110 ms and T_R= 2000 ms are reported. The filtered images related to this acquisition, are reported in Figs. 6 and 7 in case of SNR = 15 dB and 30 dB, respectively.

In Fig. 8 the normalized residual error maps are reported, which are defined according to:

$$r(n_i, m_j, k) = \frac{|x(n_i, m_j, k) - \hat{x}(n_i, m_j, k)|}{x(n_i, m_j, k)} \tag{13}$$

The lower amount of residual noise that characterizes the KSNLM solution is evident. Compared to another Non Local approach, as BM3D, KSNLM provides less errors on the edges, differently from KSNLM.

The values of MSE, FOM, BC and SSIM computed for all the considered filters in case of SNR spanning between 10 dB and 30 dB are reported in Fig. 10. These indexes provide quantitative information about different aspects of the considered filters.

All the considered filters are characterized by different computational complexity, with NLM approaches requiring higher CPU time. Among the proposed filters, KSNLM is the slowest one. This can be explained by considering that the code has been developed in Mathworks® Matlab© environment, without any code optimization. Moreover no parallelization has been implemented, although the filter is well suited for such kind of processing. Efforts in this direction could

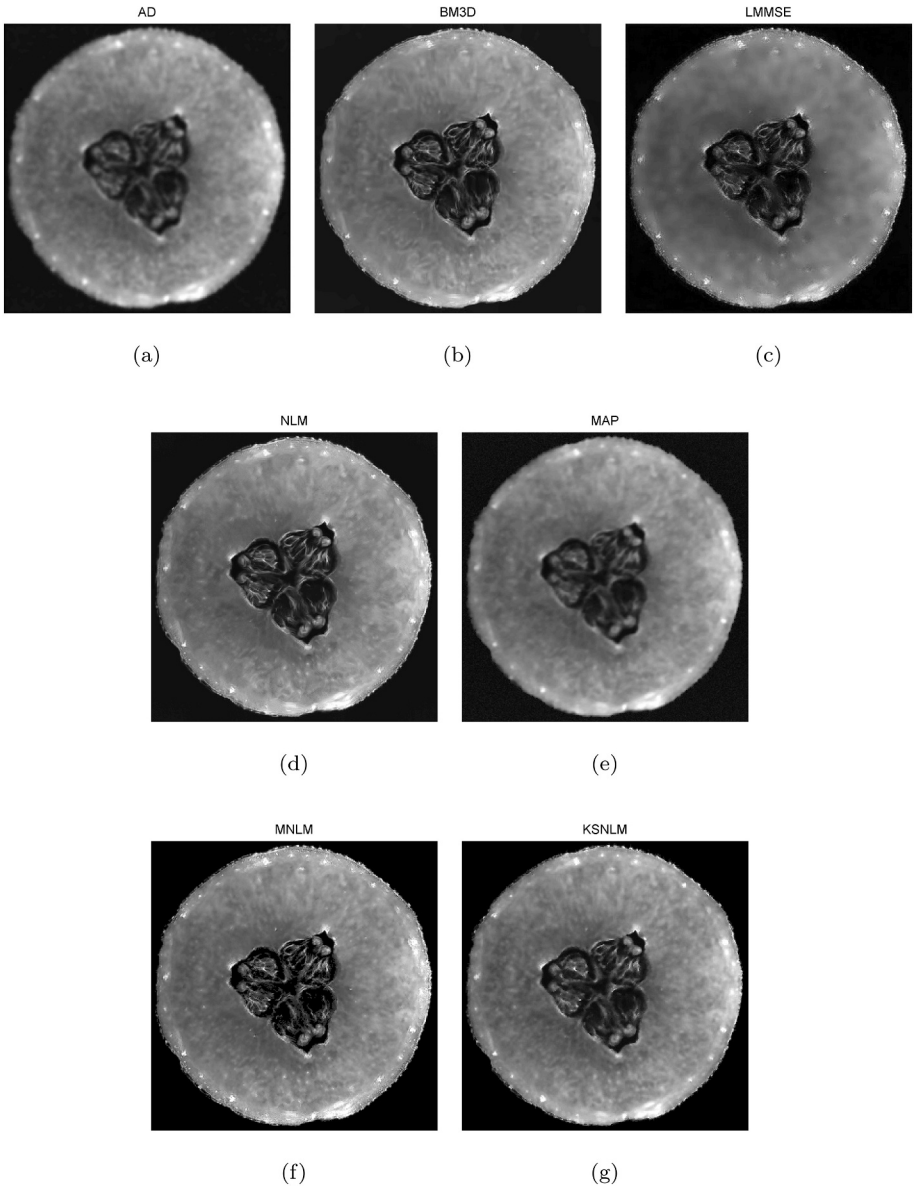

Fig. 13. First test case - Filtering results: Anisotropic Diffusion (a), Block Matching 3D (b), Linear Minimum Mean Square Error (c), Non-Local Means (d), Maximum A Posteriori (e), Multicontrast NLM (f) and Kolmogorov-Smirnov NLM (g) filters.

be done in order to greatly speed up the application of the proposed filter, but theirs are out of the scope of this manuscript.

In order to evaluate the role of the availability of several images, a further analysis has been conducted. The proposed KSNLM filter has been applied to a stack composed of a different number of images, namely 1, 2, 4 and 8 images, in case of SNR equal to 20 dB. The considered qualitative indexes have been computed and reported in Table 2

3.2.2. Rician noise

In the previous Section, Gaussian additive noise was assumed. As reported in the methodology description, this assumption can be made only in case of high SNR values, otherwise the Rice distribution has to be considered. Within this Section the differences in terms of filtering effectiveness in case of Rician data instead of Normal ones, are assessed. For this aim, the Brainweb simulated dataset has been considered and corrupted by noise according to Coupe et al. [7]:

$$y = \sqrt{(x + w_R)^2 + w_I^2} \tag{14}$$

with w_R, $w_I \sim \mathcal{N}(0,\ \sigma_W^2)$. For this analysis, we assumed σ_W equal to the 7% of the value of the brightest tissue in the image.

In Table 3, the MSE and the SSIM values of the considered methodologies are reported both in case of Gaussian and Rician models.

An analysis of the achievable performances in case of different noise levels has also been conducted. The algorithm has been tested on the Brainweb dataset corrupted by a Rician noise with SNR equal to 10 dB, 20 dB and 30 dB. In Fig. 11 the noisy images are reported, together with the regularized solutions of KSNLM.

3.3. Real data results

Three real test cases have been considered to validate the performances of the proposed KSNLM algorithm. All the datasets have been acquired via a Philip Achieva 3.0T scanner [34] and are characterized by a spatial resolution of 0.45 × 0.45 mm. The first test case is related to a 301 × 298 pixels slice of a cantaloupe. The dataset is composed of 15 images obtained using Spin Echo acquisition sequence with different Echo and Repetition times combinations. In particular, the following

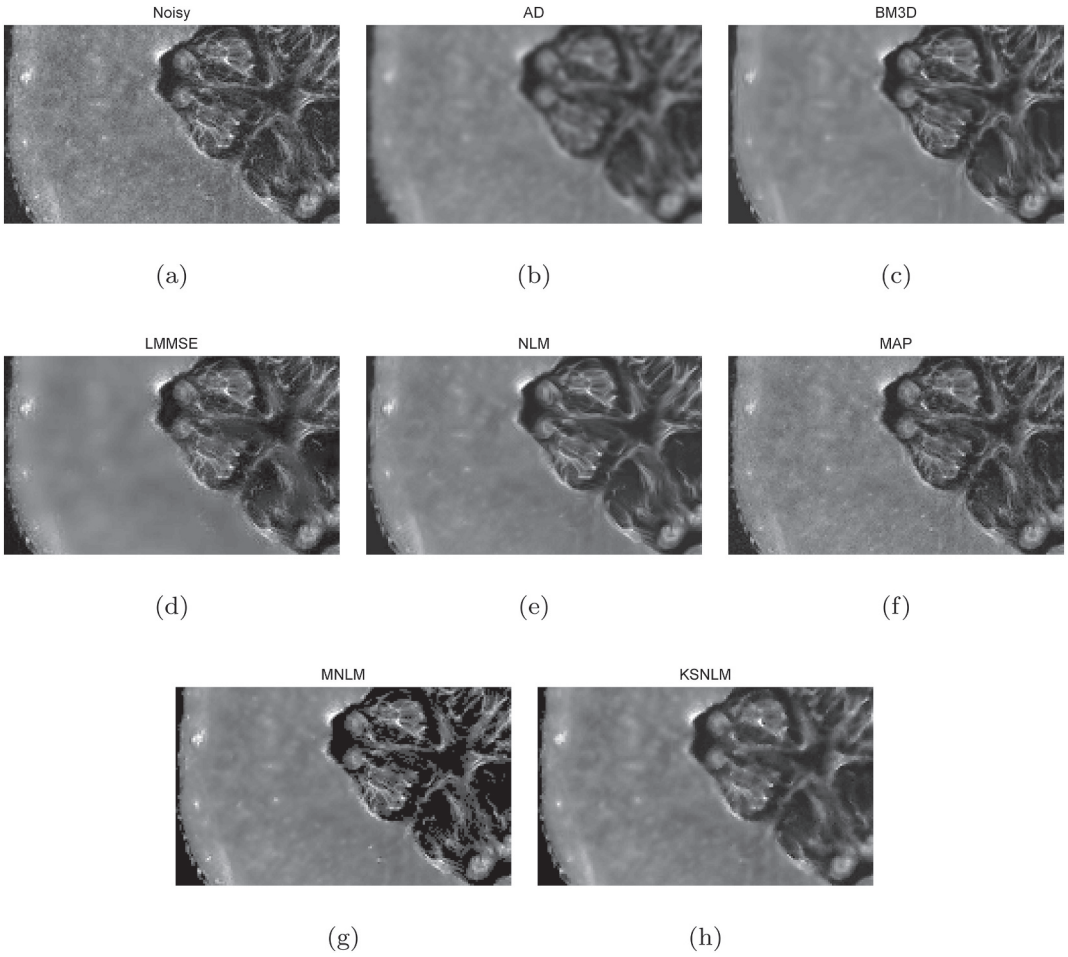

Fig. 14. First test case - Filtering results, enlargements over a ROI: acquired image (a), Anisotropic Diffusion (b), Block Matching 3D (c), Linear Minimum Mean Square Error (d), Non-Local Means (e), Maximum A Posteriori (f), Multicontrast NLM (g) and Kolmogorov-Smirnov NLM (h) filters.

combination of times (in [ms]) has been adopted: (100,3000), (82.12,3000), (125,3000), (138,3000), (160,3000), (82.12,2500), (82.12,2200), (82.12,2080), (82.12,1800), (100,2500), (100,2200), (125,2500), (100,2080), (125,2200), (138,2500). The image with $T_E = 100$ ms and $T_R = 2080$ ms is reported in Fig. 12a.

The second and third test cases are related to an axial acquisition of the head, with a resolution of 405 × 466 points. The dataset stack is composed of 8 SE images with the following T_E and T_R values [ms]: (82.5,700), (82.5,3500), (110,3500), (140,700), (140,2800), (170,2100), (200,700), (200,3500). We recall that, among the considered filters, the proposed KSNLM method in the only one able to jointly regularize a heterogeneous stack. As the SNR values of the images vary considerably, we choose to perform the comparison with the other filters in case of two images. In particular the image with $T_E = 110$ ms and $T_R = 3500$ ms (higher SNR), shown in Fig. 12b, will be considered for the second test, while the image with $T_E = 140$ ms and $T_R = 700$ ms (lower SNR), reported in Fig. 12c, will be considered as the third test case.

Filtering results for the first test case are reported in Fig. 13, while enlargements over selected ROIs are shown in Fig. 14.

Moving to the clinical dataset, results are for the second and the third test case are reported in Figs. 15 and 17, respectively. The images have a small to severe amount of noise and, in order to highlight the performance levels of the considered methodologies, enlargements over a ROI are reported in Figs. 16 and 18.

4. Discussion

4.1. Simulated data results

4.1.1. Gaussian noise

In Figs. 6 and 7, denoising results in case of SNR = 15 dB and 30 dB are reported, respectively. Although the first test case (SNR = 15 dB) is more challenging, the second one (SNR = 30 dB) is closest to the real case, in particular in case of high field MRI scanners. By looking at the filtered images in the severe noise scenario, reported in Fig. 6, very heterogeneous results appear. Anisotropic Diffusion methods tend to over-smooth the image, while LMMSE shows a noisy solution in the edges. Better results are achieved by NLM and especially by BM3D method (both Non Local based), with good noise reduction and edges preservation, although they are both characterized by ghost effects in some areas. MAP approach stays in the middle, with worse denoising performances balanced by nice edges preservation. The MNLM algorithm is able to preserve sharp edges, although the noise reduction is not strong. Compared to the others, the proposed KSNLM filters give very good results: noise is well reduced and image sharpness is remarkable. This peculiarity is confirmed in the SNR = 30 dB case reported in Fig. 7. Although all the methods give globally good results, KSNLM produces also in this case a very neat solution.

By looking at the normalized residual error maps reported in Fig. 8, the lower amount of residual noise that characterizes the KSNLM

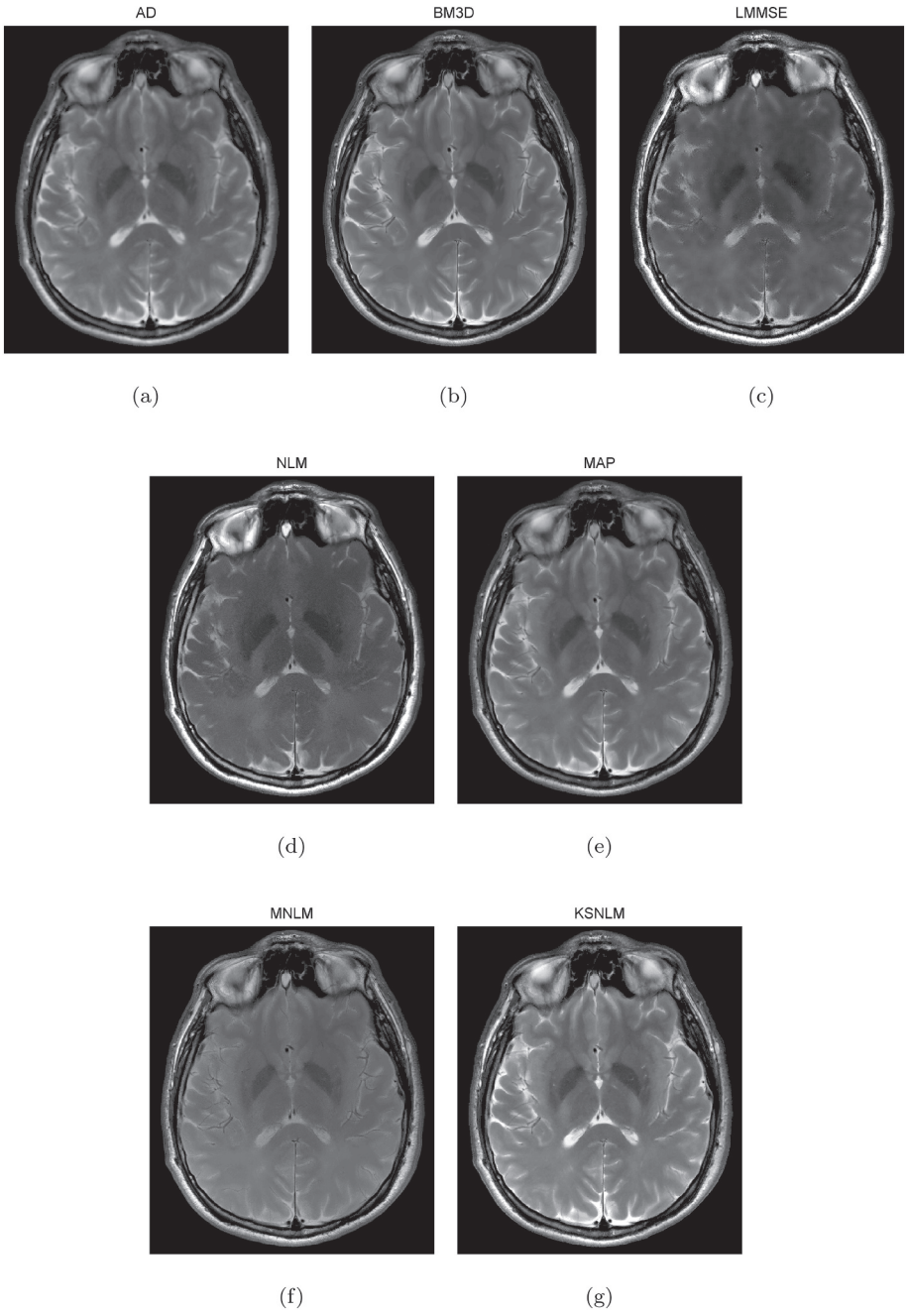

Fig. 15. Second test case - Filtering results: Anisotropic Diffusion (a), Block Matching 3D (b), Linear Minimum Mean Square Error (c), Non-Local Means (d), Maximum A Posteriori (e), Multicontrast NLM (f) and Kolmogorov-Smirnov NLM (g) filters.

solution can be appreciated. By focusing on the Non Local approaches, KSNLM provides less errors on the edges compared to BM3D, and stronger noise reduction compared to MNLM.

In Fig. 9, one line of the image is reported for better evaluating the denoising performances and the edge preservation capabilities of each filter. Such comparison has been carried out in the SNR = 30 dB case. It confirms the previous finding, with KSNLM solution being the most accurate one. In particular, the flat regions are well retrieved, and only a minimal ripple can be seen. Moreover, the sharp discontinuities are well preserved, and only some underestimation of the left peak characterizes the solution. The closest results are the BM3D, followed by NLM, but discontinuities are less sharp, and the left peak is wrongly reconstructed. All the other filters produce worse profiles.

In Fig. 10, the considered quality indexes are reported. As previously stated, MSE (Fig. 10a) is mainly indicative of the noised reduction capability. KSNLM shows interesting performances in almost all noise levels. MAP and AD approaches have similar results, showing nice performances in low noise datasets. BM3D, except in case of severe noise (SNR $\leq$ 10 dB), shows interesting results. In case of low noise (SNR = 30 dB), the MNLM is characterized by interesting performances.

FOM (Fig. 10b) evaluates the effectiveness of each filter in retrieving the edges of the image, and in preserving details. As expected from the previous results, the proposed KSNLM is one of the best of the group. In particular, it is characterized by a big improvement moving from an SNR value of 10 dB to 15 dB. Only in case of very weak noise (SNR = 30 dB), it is slightly outperformed by some other filters.

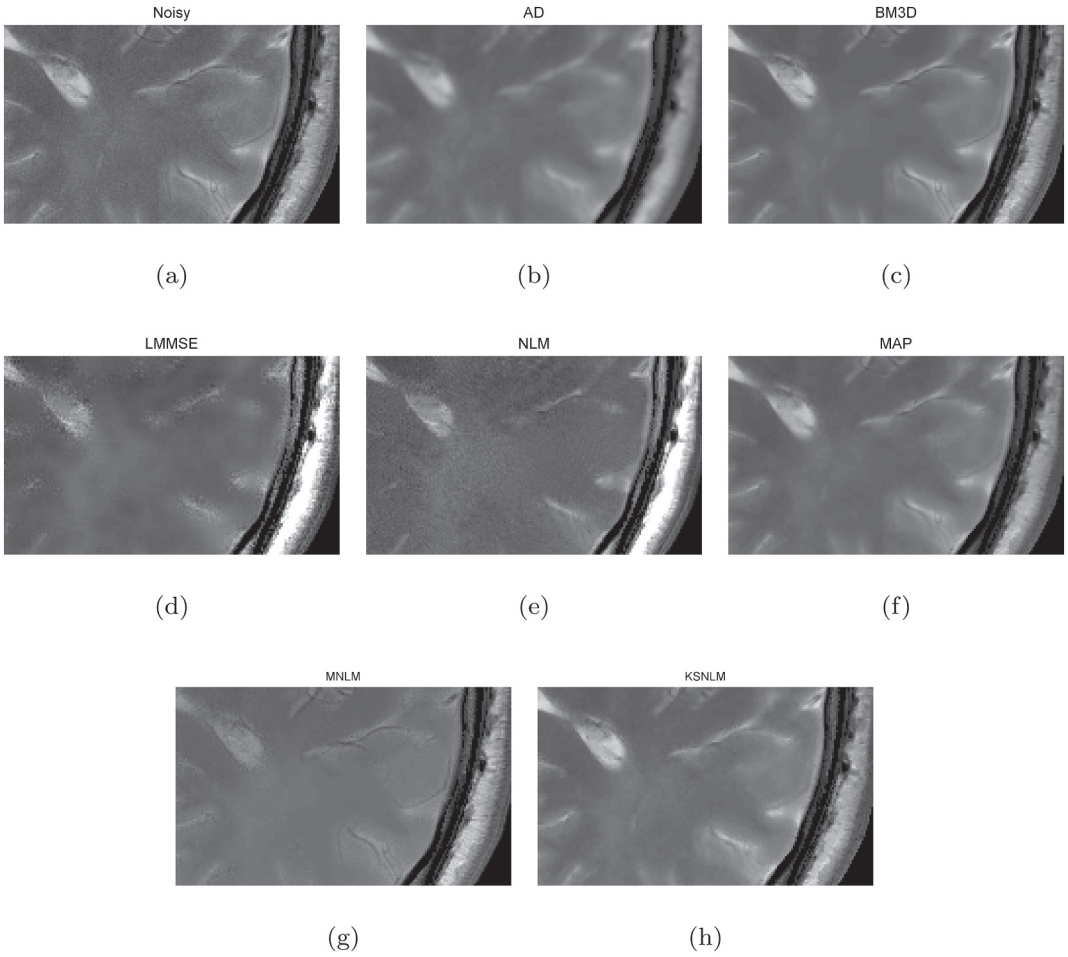

Fig. 16. Second test case - Filtering results, enlargements over a ROI: acquired data (a), Anisotropic Diffusion (b), Block Matching 3D (c), Linear Minimum Mean Square Error (d), Non-Local Means (e), Maximum A Posteriori (f), Multicontrast NLM (g) and Kolmogorov-Smirnov NLM (h) filters.

BC metric (Fig. 10c) measures the similarity between the probability distribution of the original true image and of the restored one. In case of severe noise (SNR $\leq$ 10 dB) the performances of the filters are very similar, while, decreasing the noise, KSNLM overcomes the others.

Finally, the similarity index SSIM shows that KSNLM is able to restore a solution which is more similar to the true one, compared to the others, for the different SNRs. Good performances in terms of SSIM are provided by BM3D.

It has to be underlined that, among all, in the FOM table the proposed methodology is not the leader of the group in case of high SNR (above 25 dB), where its performance level is slightly below the LMMSE approach. However, globally the considered quality indexes of Fig. 10 confirm that the proposed filter is capable of effective results, both in terms of noise reduction and details preservation, compared to the other algorithms of the comparison.

Moving to Table 2, the improvement due to the presence of several images within the stack is evident. The enhancement is significant for all the considered indexes, as expected.

4.1.2. Rician noise

A comparison of the achievable results in terms of MSE and SSIM in case of Gaussian and Rician noise is reported in Table 3. Among all, the KSNLM approach is the less sensitive to noise distribution, as both MSE and SSIM values show almost constant behavior. Moreover, such values confirm to be the best ones of the group. By considering all the values, moving to the Rician distributed signals implies a worsening of the performances, except for the KSNLM and, marginally, for the LMMSE. Similar results can be obtained by considering the other qualitative indexes and in case of different noise levels.

Results reported in Fig. 11 confirm the effectiveness of the proposed approach in reducing noise also in case of Rician distribution and low SNR.

4.2. Real data

Let us consider the results reported in Figs. 13 and 14. By analyzing the reconstructions, it is evident that the pulp area is generally well retrieved, with some filters that tend to a more smooth solution, e.g. MAP and AD, and others characterized by a more sharp reconstruction, such as BM3D, MNLM and KSNLM. The noise reduction effectiveness is globally satisfactory for all the considered approaches. In order to evaluate the preservation of the details, the most interesting area is the central region, which is full of seeds, fibers and a lot of small edges (Fig. 14). AD and LMMSE filters produce a quite blurred solution, while Non Local approaches (BM3D, MNLM and KSNLM) preserve sharpness. MAP filter stays in the middle, with a good combination of details preservation and noise reduction. The KSNLM solution is satisfactory: it is characterized by a good tradeoff between noise reduction and details preservation.

By moving to the clinical dataset of Fig. 15, let us focus on the ROI

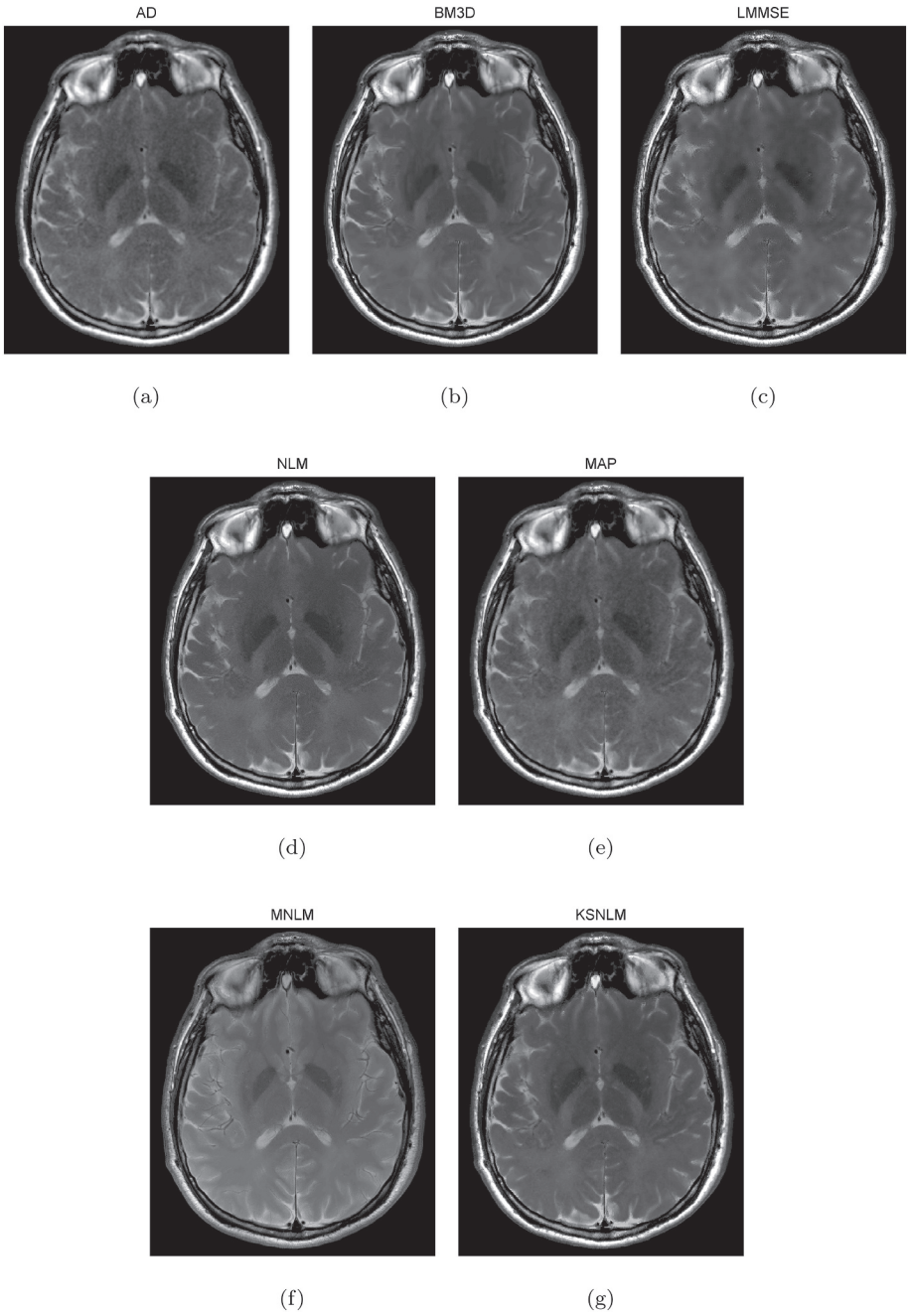

Fig. 17. Third test case - Filtering results: Anisotropic Diffusion (a), Block Matching 3D (b), Linear Minimum Mean Square Error (c), Non-Local Means (d), Maximum A Posteriori (e), Multicontrast NLM (f) and Kolmogorov-Smirnov NLM (g) filters.

of the high SNR case of Fig. 16. As expected, all of the approaches produce well regularized images, with the exception of the AD filter which over-smooths the data. A limitation of the BM3D method starts appearing in the central area, where false contours can be found. Moreover, in case of the LMMSE solution, a noisy reconstruction on the edges is evident.

A more challenging comparison is the low SNR case (third test case) of Figs. 17 and 18. It is evident that the best results are achieved by BM3D, MNLM and KSNLM filters, while the others are characterized by excessive blurring (AD) or by the presence of artifacts (LMMSE). The MAP approach stays in the middle, being effective in removing the noise but losing some details across the image. Globally, the KSNLM results are satisfactory, as fine details of the image are preserved and the image looks much clear than the original one.

5. Conclusions

A novel methodology for denoising magnetic resonance image stacks has been presented. The approach belongs to the non local mean family, and adopts a statistical criterion based on the Kolmogorov-Smirnov distance for detecting similar pixels across the image. The approach exploits stacks of multiple images of the same slice, acquired with different imaging sequence parameters. Once identified, similar pixels are fused in order to produce the filtered output. The proposed methodology has been compared to other widely adopted denoising algorithms, both local and non local approaches. A realistic dataset has been simulated in order to compute quantitative performance indexes in case of different noise intensity. Moreover, two real datasets acquired by a 3T scanner have also been exploited. Within the comparison, the

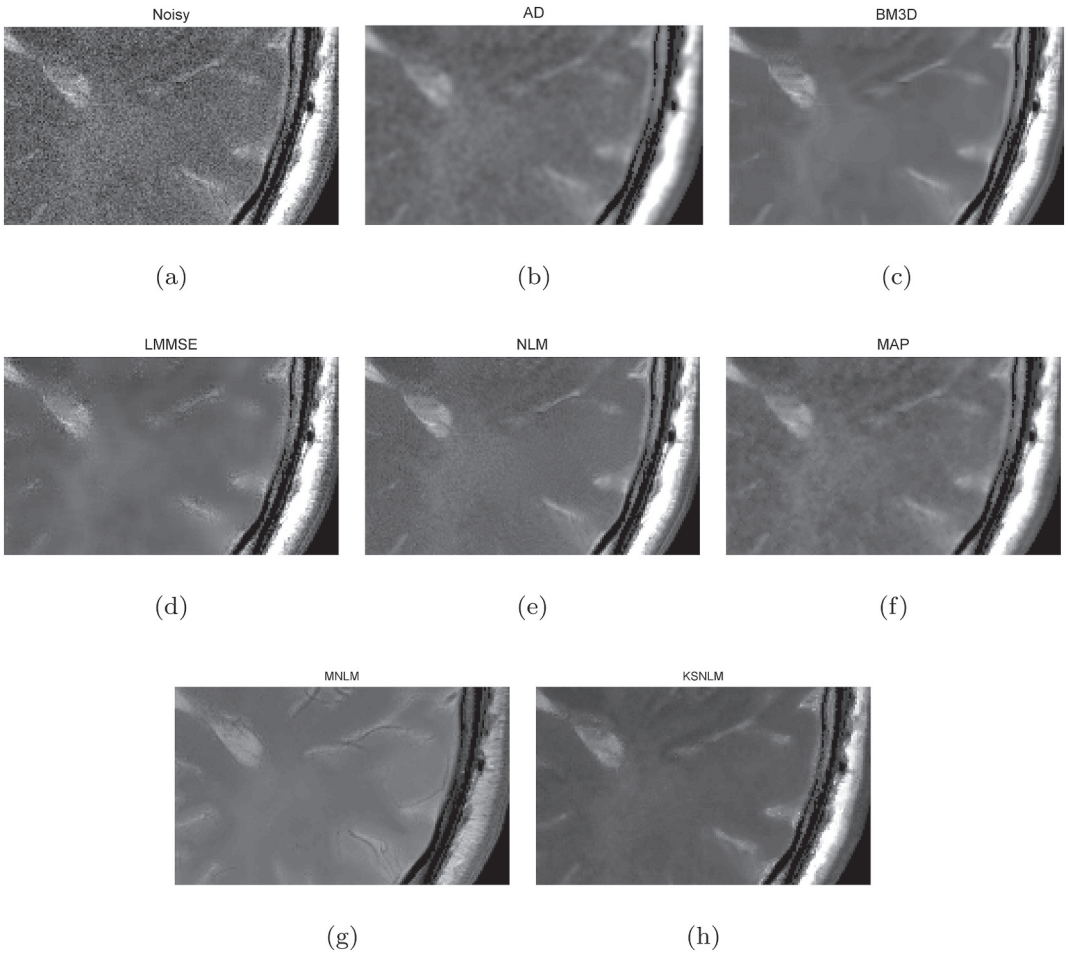

Fig. 18. Third test case - Filtering results, enlargements over a ROI: acquired data (a) and Anisotropic Diffusion (b), Block Matching 3D (c), Linear Minimum Mean Square Error (d), Non-Local Means (e), Maximum A Posteriori (f), Multicontrast NLM (g) and Kolmogorov-Smirnov NLM (h) filters.

proposed filter has shown interesting results, being characterized by a good combination of noise reduction and details preservation. At present, the main drawback of the approach is its processing time, but this issue can be addressed as the filter could greatly benefit of code parallelization.

Acknowledgments

The authors would like to thank Rocchina Caivano and Aldo Cammarota, Department of Radiology, IRCSS CROB, Rionero in Vulture, Italy, for supplying the real datasets.

Conflict of interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Funding

The present work has been partially funded by the University of Naples Parthenope, Italy, within the framework of the “Bando per il sostegno alla ricerca individuale per il triennio 2015–2017”.

References

[1] Kwon K, Kim D, Park H. Multi-contrast MR image denoising for parallel imaging using multilayer perceptron. Int J Imaging Syst Technol 2016;26:65–75.
[2] Mohan J, Krishnaveni V, Guo Y. A survey on the magnetic resonance image denoising methods. Biomed Signal Process Control 2014;9:56–69.
[3] Wiener N. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. New York: Wiley; 1949.
[4] Gerig G, Kubler O, Kikinis R, Jolesz F. Nonlinear anisotropic filtering of MRI data. IEEE Trans Med Imaging 1992;11:221–32.
[5] Baselice F, Ferraioli G, Pascazio V, Sorriso A. Bayesian MRI denoising in complex domain. Magn Reson Imaging 2017;38:112–22.
[6] Lam F, Babacan SD, Haldar JP, Weiner MW, Schuff N, Liang ZP. Denoising diffusion-weighted magnitude MR images using rank and edge constraints. Magn Reson Med 2014;71:1272–84.
[7] Coupe P, Yger P, Prima S, Hellier P, Kervrann C, Barillot C. An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images. IEEE Trans Med Imaging 2008;27:425–41.
[8] Buades A, Coll B, Morel JM. A review of image denoising algorithms, with a new one. IAM J Multiscale Model Simul SIAM Interdiscip J 2005;4:490–530.
[9] Manjon JV, Thacker NA, Lull JJ, Garcia-Martí G, Martí-Bonmatí L, Robles M. Multicomponent MR image denoising. Int J Biomed Imaging 2009:1–10.
[10] Jafari-Khouzani K. MRI upsampling using feature-based nonlocal means approach. IEEE Trans Med Imaging 2014;33:1969–85.
[11] Froment J. Parameter-free fast pixelwise non-local means denoising. Image Process On Line 2014;4:300–26.
[12] Tasdizen T. Principal neighborhood dictionaries for nonlocal means image denoising. IEEE Trans Image Process 2009;18:2649–60.
[13] Torre DL, Vrscay ER, Ebrahimi M, Barnsley MF. Measure-valued images, associated fractal transforms, and the affine self-similarity of images. SIAM J Imag Sci 2009;2:470–507.
[14] Foi A, Boracchi G. Foveated nonlocal self-similarity. Int J Comput Vis 2016;120:78–110.
[15] Baselice F. Ultrasound image despeckling based on statistical similarity. Ultrasound Med Biol 2017;43:2065–78.
[16] Ferretti A, Fumagalli A, Novali F, Prati C, Rocca F, Rucci A. A new algorithm for processing interferometric data-stacks: Squeesar. IEEE Trans Geosci Remote Sens 2011;49:3460–70.
[17] Rajan J, den Dekker AJ, Sijbers J. A new non-local maximum likelihood estimation method for Rician noise reduction in magnetic resonance images using the Kolmogorov-Smirnov test. Signal Process 2014;103:16–23.
[18] Pierazzo N, Lebrun M, Rais ME, Morel JM, Facciolo G. Non-local dual image

denoising. 2014 IEEE International Conference on Image Processing (ICIP). 2014. p. 813–7.
[19] Tristán-Vega A, García-Pérez V, Aja-Fernández S, Westin CF. Efficient and robust nonlocal means denoising of MR data based on salient features matching. Comput Methods Prog Biomed 2012;105:131–44.
[20] Sijbers J, den Dekker AJ, Van Audekerke J, Verhoye M, Van Dyck D. Estimation of the noise in magnitude MR images. Magn Reson Imaging 1998;16:87–90.
[21] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457–81.
[22] Massey FJ. The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc 1951;46:68–78.
[23] Nowak R. Wavelet-based Rician noise removal for magnetic resonance imaging. IEEE Trans Image Process 1999;10:1408–19.
[24] Manjon JV, Carbonell-Caballero J, Lull JJ, García-Martí G, Martí-Bonmatí L, Robles M. MRI denoising using non-local means. Med Image Anal 2008;12:514–23.
[25] Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Trans Image Process 2007;16:2080–95.
[26] Aja-Fernández S, Alberola-López C, Westin CF. Noise and signal estimation in magnitude MRI and Rician distributed images: a LMMSE approach. IEEE Trans Image Process 2008;17:1383–98.
[27] Buades A, Coll B, Morel JM. A non-local algorithm for image denoising. Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE; 2005. p. 60–5.
[28] Fabio Baselice, Giampaolo Ferraioli, Vito Pascazio. A 3D MRI denoising algorithm based on Bayesian theory. Biomed Eng OnLine 2017;16:25.
[29] He L, Greenshields IR. A nonlocal maximum likelihood estimation method for Rician noise reduction in MR images. IEEE Trans Med Imaging 2009;28:165–72.
[30] Pratt WK. Digital Image Processing: PIKS Inside. 3rd ed. Wiley-Interscience; 2001.
[31] Kailath T. The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans Commun Technol 1967;15:52–60.
[32] Wang Z, Bovik A, Sheikh H, Simoncelli E. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13:600–12.
[33] Collins DL, Zijdenbos AP, Kollokian V, Sled JG, Kabani NJ, Holmes CJ, et al. Design and construction of a realistic digital brain phantom. IEEE Trans Med Imaging 1998;17:463–8.
[34] Baselice F, Caivano R, Cammarota A, Ferraioli G, Pascazio V. T1 and T2 estimation in complex domain: first results on clinical data. Concepts Magn Reson Part A 2014;43:166–76.