DENOISING OF MEDICAL IMAGES CORRUPTED BY POISSON NOISE

[1,2]*Isabel Rodrigues,* [2,4]*João Sanches and José Bioucas-Dias*[3,4]

Instituto Superior de Engenharia de Lisboa[1]
Instituto de Sistemas e Robótica[2]
Instituto de Telecomunicações[3]
Instituto Superior Técnico[4]

ABSTRACT

Medical images are often noisy owing to the physical mechanisms of the acquisition process. The great majority of the denoising algorithms assume *additive white Gaussian noise*. However, some of the most popular medical image modalities are degraded by some type of non-Gaussian noise. Among these types, we refer the Poisson noise, which is particularly suitable for modeling the counting processes associated to many imaging modalities such as PET, SPECT, and fluorescent confocal microscopy imaging. The aim of this work is to compare the effectiveness of several denoising algorithms in the presence of Poisson noise. We consider algorithms specifically designed for Poisson noise (wavelets, Platelets, and minimum descritpion length) and algorithms designed for Gaussian noise (edge preserving bilateral filtering, total variation, and non-local means). These algorithms are applied to piecewise smooth simulated and real data. Somehow unexpectedly, we conclude that total variation, designed for Gaussian noise, outperforms more elaborated state-of-the-art methods specifically designed for Poisson noise.

Index Terms— Denoising, Poisson, Multiplicative, Bayesian, Regularization, Wavelets, Total Variation, Non-local Means.

1. INTRODUCTION

Noise removal is essential in medical imaging applications in order to enhance and recover anatomical details that may be hidden in the data. The literature is rich in denoising methods assuming the *additive white Gaussian noise (AWGN)* model. However, some important imaging modalities are corrupted by Poisson noise. In fact, imaging acquisition systems using photon-counting devices such as positron emission tomography, single photon emission computed tomography, and confocal microscopy imaging (see Fig. 1) are dominated by Poisson noise. Denoising such images is an ill-posed problem, usually leading to hard optimization problems involving non-quadratic (due to the Poisson observations and non-Gaussian priors), non-negatively constrained, and, possibly, non-convex objective functions.

Let $\mathbf{y} = \{y_{i,j} : i,j = 1,\ldots,N\}$ and $\mathbf{x} = \{x_{i,j} : i,j = 1,\ldots,N\}$ denote the noisy and the original images, respectively. Samples $y_{i,j}$ are contaminated by Poisson noise. Thus, the likelihood of observing $\mathbf{y}$ given the true image $\mathbf{x}$ is

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i,j=1}^{N} \frac{e^{-x_{i,j}} x_{i,j}^{y_{i,j}}}{y_{i,j}!}. \quad (1)$$

Correspondent author: J. Sanches (jmrs@ist.utl.pt). This work was supported by Fundação para a Ciência e a Tecnologia (ISR/IST plurianual funding) through the POS Conhecimento Program which includes FEDER funds and by the Instituto de Telecomunicações under the project IT/LA/325/2005.

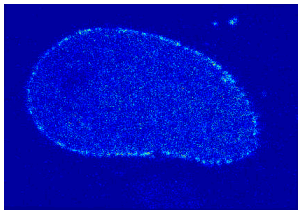

Fig. 1. *Confocal Laser Scanning Microscopy* (CLSM) imaging (Cell)

Let y be a Poisson random variable with mean $E[y] = x$. Defining the signal-to-noise ratio as SNR = $\|\mathbf{x}\|^2 / E\|\mathbf{y} - \mathbf{x}\|^2$, and noting that Var$[y] = x$, we have SNR = $\|\mathbf{x}\|^2 / \sum_{i,j} x_{i,j}$. To control the SNR of simulated images, we generated images according to $\mathbf{y} \sim Po(\alpha\mathbf{x})$, where Po stands for the Poisson density function and $\alpha > 0$. We have then SNR = $\alpha\|\mathbf{x}\|^2 / \sum_{i,j} x_{i,j}$, *i.e.*, the SNR is linearly dependent on α.

Given an estimate $\widehat{\mathbf{x}}$ of the original intensity $\mathbf{x}$, we compute, for comparison purposes, the *Mean Square Error*, MSE $= 1/M^2 \sum_{i,j=1}^{M} (\hat{x}_{i,j} - x_{i,j})^2$, the *Peak Signal-to-Noise Ratio*, PSNR $= 10\log_{10}\left(x_{max}^2/\text{MSE}\right)$, the *Signal-to-Noise Ratio Improvement*, ISNR = SNR$_f$ − SNR$_i$, where SNR$_i$ and SNR$_f$ are the signal-to-noise ratios before and after applying the algorithm, respectively, and T, the CPU time the algorithm takes to run in a Centrino Duo, 2GB RAM.

2. OVERVIEW OF POISSON DENOISING

In the seventies, W.H. Richardson and L. Lucy presented a denoising technique for Poisson noise, named after them as the R-L algorithm. It consists on the iterative minimization of a non-quadratic log-likelihood function with multiplicative corrections. The main shortcoming of this method is that after a few iterations, the algorithm yields highly noisy estimates, in particular when the SNR is low [1].

In 1992 Rudin, Osher and Fatemi [2] proposed the total variation (TV) regularization scheme applied to Gaussian denoising. This methodology constitutes an important achievement in the field of the *edge preserving* denoising algorithms, suitable to deal with the discontinuities associated with anatomical details. A combination of R-L with TV was applied with success by Dey *et al.* [3] to *confocal laser scanning microscopy* images degraded by out-of-focus blur and/or Poisson noise due to photon-limited detection. X. Zhang *et al.* [4] adopted the TV regularizer in tomographic imaging. TV regularization was also used by Bardsley and Luttman [5] jointly with a Poissonian likelihood functional.

Wavelet-based methods are now of widespread use in medical imaging and disease diagnosis. Most of these algorithms operate by applying some type of shrinkage/thresholding to the wavelet image coefficients and then synthesize the denoised image from these coefficients. The most difficult task is the estimation or adoption of an appropriate threshold [6], [7]. Unser *et al.*, in 2003, presented a description of the state-of-the-art on the subject of wavelets in medical imaging in [8]. Besbeas *et al.* presented in [9] a comparative study of wavelet shrinkage methods for estimating the underlying intensity, based on observations from a Poisson regression model. The unnormalized Haar wavelet is an appropriate transform to the case of the Poisson distribution, since it is self-reproducing across scales. Taking advantage of this property, Timmermann and Nowak [10] derived a simple Bayesian intensity estimate procedure, the Multiscale Multiplicative Innovations (MMI) model. MMI was applied to photon-limited imaging. Multiscale analysis is a powerful tool in what concerns denoising procedures. Nowak and Kolaczyk [11] generalized the theory of multiresolution analysis to likelihoods that can be represented as the product of conditional densities, possessing information on the data and on the parameter vector localized in position and scale. Poisson is one of the members of the family of distributions that allow such a factorization. This theory has its analog in wavelets. Willett and Nowak [12] introduced a new multiscale method for nonparametric piecewise polynomial intensity and density estimation of Poisson point processes. The authors developed fast computing piecewise polynomial maximum penalized likelihood methods that become recursive partioning schemes based on multiscale likelihood factorization. These schemes yield nearly-optimal performance dispensing with any *a priori* knowledge of the underlying signal's smoothness. Also using a recursive partioning scheme, Nowak and Figueiredo [13] introduced methods for estimating the underlying intensity functions of spatial Poisson point processes, assuming those functions to be approximately piecewise constant; the splitting of the region is operated according to the Minimum Description Length (MDL) criterium.

The Non-local Means algorithm (NLM), introduced by Buades *et al.* [14], is a non-local averaging technique, operating on all pixels in the image with the same characteristic. Unfortunately the method is very slow. To speed it up, Mahmoudi and Sapiro [15] proposed a scheme of pre-selection of neighborhoods. With the same idea, Coupé *et al.* [16] proposed a similar algorithm where they have used parallel processing. With a similar philosophy, Dabov *et al.* [17] presented an approach to image denoising, based on effective filtering in 3-D transform domain, by combining sliding windows transform processing with block matching. The blocks within the image are processed in a sliding way, which means that given a block, the algorithm searches among the other blocks, which ones match according to a certain criterium. The matching blocks are stacked together forming a 3-D array with high level of correlation. A 3-D unitary transform is applied and noise is attenuated due to the shrinkage of the coefficients of the transform. This 3-D transform produce estimates of all the matched blocks. Repeating this procedure for all blocks in a sliding way, the final estimate is computed as a weighted average of all overlapping block estimates. The authors proposed a fast and efficient algorithm to solve this problem.

3. ALGORITHMS

In this section, the results of a set of six representative algorithms, three of them for Poisson denoising and the other three for Gaussian noise removal are compared using synthetic and real data. In the case of synthetic data, the comparison is based on the *figures of merit* PSNR, ISNR, MSE and CPU time, as described in the previous section.

The **NLM** [18] is based on the non-local averaging of all pixels in the image with the same characteristic. The NLM can be regarded as an evolution of the Yarolavski filter (1985), where the average is performed among similar pixels in the image and the measure of similarity is based on the local intensity. The main difference between this filter and NLM is the way the similarity is measured; the latter is more robust, since not only it compares the gray intensity level in a single point, but also the geometric configuration in a whole neighborhood. In this context, if $\mathbf{Y}$ is a noisy image on a bounded domain $\Omega \in \Re^2$ and $x \in \Omega$, the NLM algorithm estimates the value of x as an average of the values of all the pixels whose neighborhood is similar, in a certain sense, to the neighborhood of x,

$$\hat{x} = NLM(\mathbf{Y})(x) = \frac{1}{C(x)} \int_{\Omega} e^{K(x,z)} \mathbf{Y}(z) dz \tag{2}$$

with

$$K(x,z) = -\frac{1}{h^2} \int_{\Re^2} G_a(t) \|\mathbf{Y}(x+t) - \mathbf{Y}(z+t)\|^2 dt \tag{3}$$

and the normalizing factor $C(x) = \int_{\Omega} e^{K(x,z)} dz$. Buades *et al.* proved that NLM is asymptotically optimal under a generic statistical image model.

The **W** [12] algorithm is a Bayesian approach to Poisson intensity estimation based on the translation invariant (TI) hereditary unnormalized Haar wavelet transform. Since the sum of independent Poisson variates C_i with parameters λ_i is also Poisson distributed with parameter $\sum \lambda_i$, this type of wavelets allows a simple formulation in the case of Poisson data, since every scaling coefficient is the sum of two finer-scale scaling coefficients. This means that the Poisson distribution is self-reproducing across scales. In this algorithm, traditional hard or soft threshold schemes are not applicable; instead, wavelet coefficients are scaled according to their ancestors pruning decisions. The hereditary nature of the pruning process is responsible for the robustness of this algorithm.

The **P** [12] is a nonparametric multiscale platelet algorithm that, unlike traditional wavelet-based methods, is well suited to both photon-limited medical imaging applications involving Poisson data and to piecewise smooth images, *i.e.*, images consisting of smooth regions separated by smooth boundaries. Platelets are localized functions at various scales, locations, and orientations that produce piecewise linear image approximations. A multiscale image decomposition based on these functions is performed. This is a relatively fast, platelet-based, penalized likelihood method. The idea of the method is to find the partition of the *region of interest* which minimizes the following penalized likelihood function:

$$\hat{\mathcal{P}} = \arg\min_{\mathcal{P}} \left[-\log p\left(x|\mathbf{f}\left(\hat{\mathcal{P}}\right)\right) + E_{\textbf{Plat}}\left(\hat{\mathcal{P}}\right) \right], \tag{4}$$

where $p(x|\mathbf{f}(\hat{\mathcal{P}}))$ is the likelihood of observing the counts x given the estimate $\hat{\mathbf{f}} = \mathbf{f}(\hat{\mathcal{P}})$ and $E_{\textbf{Plat}}(\hat{\mathcal{P}})$ is the penalization for using the model of the platelets. $\hat{\mathbf{f}}$ is the model maximum likelihood estimate in each region of the partition $\hat{\mathcal{P}}$ or the penalized likelihood estimator (PLE). The dependence of the estimator on the dyadic partition is attenuated through a process of cycle-spinning that consists on circularly shift the raw data, denoising and shift the estimate back to its initial position.

The Minimum Description Length **MDL** based method [13] gives an estimate of the underlying intensity function of a spatial Poisson point process, assuming it to be approximately piecewise constant. The algorithm builds, in a recursive fashion, a partition of the observation space into regions where the intensity can be considered as constant. The complexity of the partition is measured by information theoretical tools and the MDL criterium is used to find the partition of minimum length. Although the overall scheme is not optimal, it is effective and computationally light.

The bilateral filtering **BIL** algorithm [19] smooths images but preserves edges by means of a nonlinear combination of nearby image values. The method is noniterative, local, simple, and fast. It combines gray levels based on their geometric closeness and their photometric similarity; it gives preference to near values in both domain and range. The combined filtering is given by:

$$\mathbf{h}(\mathbf{x}) = \frac{1}{K(\mathbf{x})} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbf{f}(\mathbf{z}) c(\mathbf{z}, \mathbf{x}) s(\mathbf{f}(\mathbf{z}), \mathbf{f}(\mathbf{x})) d\mathbf{z}, \qquad (5)$$

where $\mathbf{f}(z)$ is the noisy image, $c(\mathbf{z}, \mathbf{x})$ is a measure of the geometric closeness between the neighborhood center $\mathbf{x}$ and the nearby point $\mathbf{z}$ and $s(\mathbf{f}(\mathbf{z}), \mathbf{f}(\mathbf{x}))$ measures the photometric similarity between the pixel at the neighborhood center $\mathbf{x}$ and that of a nearby point $\mathbf{z}$. $K(\mathbf{x})$ is a normalization factor and is given by

$$K(\mathbf{x}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} c(\mathbf{z}, \mathbf{x}) s(\mathbf{f}(\mathbf{z}), \mathbf{f}(\mathbf{x})) d\mathbf{z}. \qquad (6)$$

In fact, the bilateral filtering replaces the pixel value at $\mathbf{x}$ with an average of similar and nearby pixel values.

The denoising optimization problem under the TV regularization and AWGN noise is solved in a very fast way by algorithms by Chambolle [20] and by Figueiredo *et al.* [21]. The former exploits projections onto convex sets, whereas the latter adopts the majorization minimization (MM) framework. Although these algorithms were derived for Gaussian noise, we have included the last one (**TVMM**) due its competitive performance even operating under noise statistic for which it has not been conceived.

4. EXPERIMENTAL RESULTS

In this section experimental results using synthetic and real data are presented, comparing the performance of the algorithms described above.

4.1. Synthetic Data

The results of this section are based on the Shepp-Logan phantom of size 128×128 pixels, in a gray scale ranging from 0 to 255. Fig.2 displays the original Logan phantom, the phantom corrupted with Poisson noise and the denoised images using the described methods.

The Logan phantom was scaled and corrupted by Poisson noise in order to obtain a $15dB$ SNR. This corrupted image was denoised with each denoising method and the *figures of merit* referred in the Introduction section were computed and listed in table 1.

The performance of these algorithms depends a great deal on several specific parameters that are chosen by the user while running the software, such as the number of shifts in the **P** algorithm or the size of the search window in **NL**M. A different tuning of these parameters for a specific image may lead to different results. In our case the parameters were chosen in order to accomplish a compromise between PSNR and CPU time T. Among the methods specifically designed for Poisson denoising, (**W**, **P**, **MDL**), the Wavelets based method presents the most competitive results: the highest PSNR and ISNR and simultaneously the shorter CPU time T. Among the methods for Gaussian denoising, the **TVMM** achieves the better scores. Comparing the estimation errors produced by the six algorithms, **W**, **P**, and **TVMM** displays similar results, which are better than the remaining algorithms. Concerning CPU time, **W** algorithm is the fastest, followed closely by **TVMM** and far way by **P**.

In the next experiment with synthetic data we varied the SNR between 5 and 25 dB. Each noisy image was denoised using the tested algorithms and the final SNR (SNR_f) of the denoised images were computed and displayed in Figure 3 for the six methods. The **TVMM** algorithm shows in general superior performance in the range 5 $dB-$ 20 dB. For 25 dB the **BIL** algorithm presents a slightly higher SNR_f. An interesting fact about **TVMM** is the curve for this algorithm being almost a straight line with a slope of approximately $4/5$, which means that given a value of SNR_i in the range shown in the plot, it is possible to have an idea of the value of the final SNR one is going to get. The Wavelets based method shows an increasing performance for SNR_i in the range $5dB$ - $15dB$. With th exception of the 15 dB SNR_i, the TVMM always outperforms the **W** algorithm.

Alg.	PSNR	ISNR	MSE	T (s)	SNR_f
W	37.406	9.617	11.815	0.14	24.575
P	37.154	9.365	12.523	360.43	24.323
MDL	36.059	8.270	16.112	0.31	23.228
BIL	34.801	7.012	21.527	1.52	21.970
TVMM	37.246	9.457	12.260	0.97	24.415
NLM	35.054	7.257	20.307	9.20	22.215

Table 1. Poisson Denoising Logan 128x128. $x_{max} = 255$ SNR=15dB

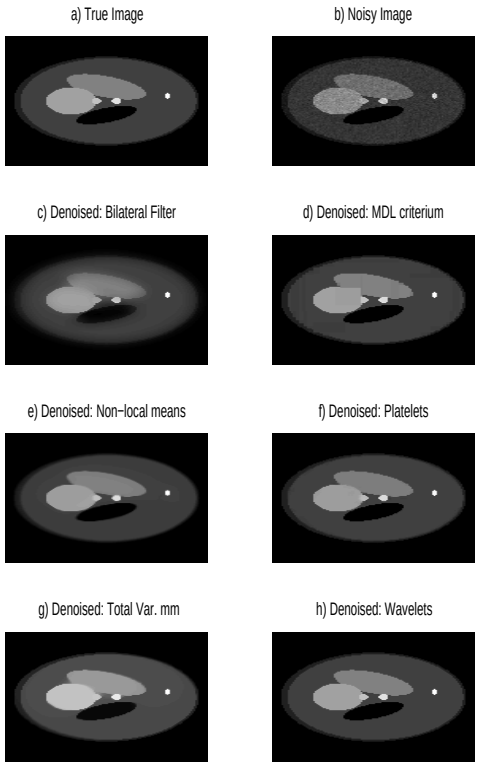

Fig. 2. Denoising with the Shepp-Logan phantom 128×128 pixels.

4.2. Real Data

In this section we apply the algorithms under comparison to real data from confocal microscopy and PET images. Figure 4 shows the original images and the denoising results of the three models we have found more appropriate to these data: W, P and TVMM. Visual inspection indicates that TVMM is the most convenient method for the denoising of the cell image. For the tumor image, the Platelets based method seems to give the best results.

5. CONCLUDING REMARKS

This paper presents a comparison of denoising algorithms applied to images corrupted by Poisson noise. We are particularly interested

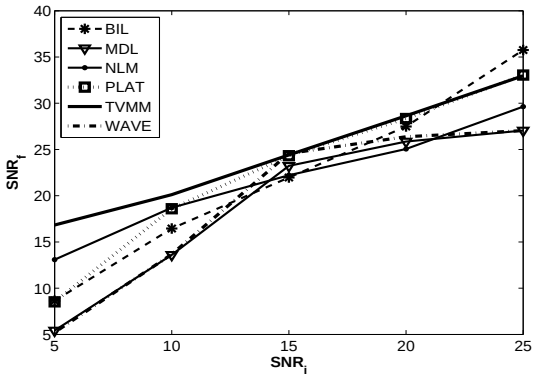

Fig. 3. SNR initial *vs* SNR final for the 6 algorithms.

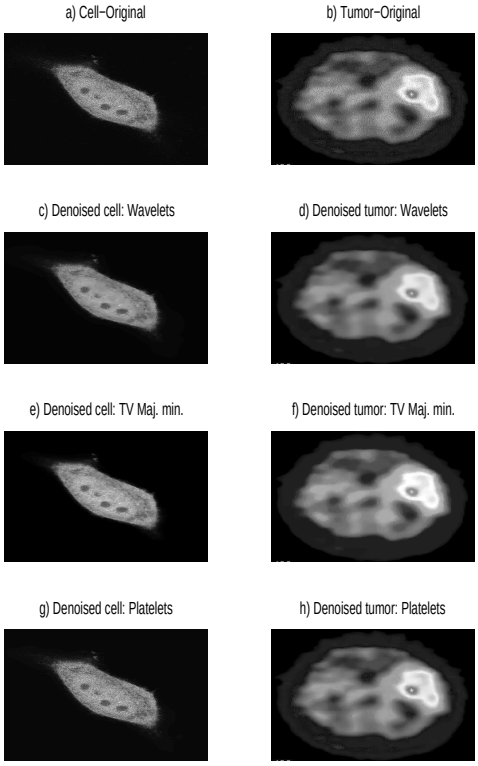

Fig. 4. Denoising Cell and Tumor

in medical images, to which the assumption of piecewise smoothness of the original images is a good model. Six algorithms were tested. Three of them were specifically designed for Poisson noise (wavelet, Platetet , and minimum descritpion length); the remaining three were designed for Gaussian noise (edge preserving bilateral filtering, total variation, and non-local means). The algorithm were applied to simulated and real data. We conclude, perhaps unexpectedly, that total variation denoising designed for white Gaussian noise quasi-uniformly (with respect to SNR) the best algorithm.

6. REFERENCES

[1] R. Vio, J. Bardsley, and W. Wamsteker, "Least-squares methods with poissonian noise: an analysis and a comparison with the richardson-lucy algorithm," *Astronomy and Astrophysics*, vol. 436, pp. 741–755, 2005.

[2] Leonid I. Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," in *Proceedings of the eleventh annual international conference of the Center for Nonlinear Studies on Experimental mathematics : computational issues in nonlinear science*, Amsterdam, The Netherlands, The Netherlands, 1992, pp. 259–268, Elsevier North-Holland, Inc.

[3] N. Dey, L. Blanc-Féraud, C. Zimmer, Z. Kam, J.C. Olivo-Marin, and J. Zerubia, "A deconvolution method for confocal microscopy with total variation regularization," in *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*, 2004.

[4] Xiao-Qun Zhang and J. Froment, "Total variation based fourier reconstruction and regularization for computer tomography," *Nuclear Science Symposium Conference Record, 2005 IEEE*, vol. 4, pp. 2332–2336, 23-29 Oct. 2005.

[5] J.Bardsley and A.Luttman, "Total variation-penalized poisson likelihood estimation for ill-posed problems," Tech. Rep. 8/2006, Department Math Sciences, University of Montana Missoula, 2006.

[6] David L. Donoho and Iain M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

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

[8] M.Unser, A.Aldroubi, and A.Laine, "Guest editorial: Wavelets in medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 285–288, 2003.

[9] Panagiotis Besbeas, Italia De Feis, and Theofanis Sapatinas, "A comparative simulation study of wavelet shrinkage estimators for Poisson counts.," *Int. Stat. Rev.*, vol. 72, no. 2, pp. 209–237, 2004.

[10] K.E. Timmermann and R.D. Nowak, "Multiscale modeling and estimation of poisson processes with application to photon-limited imaging," *Information Theory, IEEE Transactions on*, vol. 45, no. 3, pp. 846–862, Apr 1999.

[11] R.D. Nowak and E.D. Kolaczyk, "A statistical multiscale framework for poisson inverse problems," *Information Theory, IEEE Transactions on*, vol. 46, no. 5, pp. 1811–1825, Aug 2000.

[12] R.M. Willett and R.D. Nowak, "Fast multiresolution photon-limited image reconstruction," *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pp. 1192–1195 Vol. 2, 15-18 April 2004.

[13] R.D. Nowak and M.A.T. Figueiredo, "Unsupervised progressive parsing of poisson fields using minimum description length criteria," *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, vol. 2, pp. 26–30, 1999.

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

[15] 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, Dec. 2005.

[16] P. Coupé, P. Yger, and C. Barillot, "Fast Non Local Means Denoising for 3D MR Images," in *9th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI'2006*, R. Larsen, M. Nielsen, and J. Sporring, Eds., Copenhagen, Denmark, October 2006, vol. 4191 of *Lecture Notes in Computer Science*, pp. 33–40, Springer.

[17] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3d-transform domain collaborative filtering," *IEEE Trans. on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[18] Jérôme Boulanger, Charles Kervrann, and Patrick Bouthemy, "Space-time adaptation for patch-based image sequence restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1096–1102, 2007.

[19] 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, 1998, p. 839, IEEE Computer Society.

[20] Antonin Chambolle, "Total variation minimization and a class of binary mrf models," in *EMMCVPR*, 2005, pp. 136–152.

[21] M.A.T. Figueiredo, J.B. Dias, J.P. Oliveira, and R.D. Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparisonwith wavalet denoising," *Image Processing, 2006 IEEE International Conference on*, pp. 2633–2636, 8-11 Oct. 2006.