DENOISING DIFFUSION-WEIGHTED MR MAGNITUDE IMAGE SEQUENCES USING LOW RANK AND EDGE CONSTRAINTS

Fan Lam[1]*, S. Derin Babacan*[1]*, Justin P. Haldar*[2]*, Norbert Schuff*[3]*, and Zhi-Pei Liang*[1]

[1]Department of Electrical and Computer Engineering, and
Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign
[2]Signal and Image Processing Institute, the Dornsife Cognitive Neuroscience Imaging Center, and
the Brain and Creativity Institute, University of Southern California
[3]Department of Radiology and Biomedical Imaging, University of California and VA Medical Center, San Francisco

ABSTRACT

This paper addresses the denoising problem associated with diffusion MR imaging. Building on previous approaches to this problem, this paper presents a new method for joint denoising of a sequence of diffusion-weighted (DW) magnitude images. The proposed method uses a *maximum a posteriori* (MAP) estimation formulation to incorporate a Rician likelihood (for modeling the noisy magnitude data), a low rank model (for the DW image sequences) and a spatial prior (for imposing joint edge constraints). An efficient algorithm to solve the associated optimization problem is also described. The proposed method has been evaluated using both simulated and experimental diffusion tensor imaging (DTI) data, which yields very encouraging results both qualitatively and quantitatively.

***Index Terms*—** Diffusion-weighted imaging, diffusion-tensor imaging, Rician noise, low rank, edge constraints.

1. INTRODUCTION

Limited signal-to-noise ratio (SNR) has been a long-standing obstacle for high-resolution diffusion MRI, and various denoising methods have been proposed to address this problem. It has been established that denoising methods that process complex-valued data can significantly outperform methods that use only magnitude images, and are also more easily characterized [1]. However, it is often the case that clinical and research MRI scanners do not save the phase information in diffusion weighted imaging experiments. As a result, denoising diffusion weighted (DW) magnitude images remains an important practical problem, which is the main focus of this paper.

While signal-independent Gaussian noise modeling is appropriate and straightforward for complex-valued data [1], the noise statistics for magnitude images are more complicated. In particular, the noise in magnitude images are signal dependent, and follows a Rician distribution [2, 3]. It is important that these Rician image statistics are modeled explicitly [3], and various Rician denoising approaches have appeared in the literature [4-11]. Existing denoising approaches have made use of various image constraints, including sparse image modeling [4], local smoothness constraints [5–7], and non-local image self-similarities [8–10]. These methods take advantage of both knowledge of the noise distribution and additional prior information about the characteristics of magnitude image sequences.

This work was supported in part by the following research grants: NIH-P41-RR023953, NIH-P41-EB001977 and NIH-1RO1-EB013695-01.

This paper proposes a new method for denoising DW magnitude images, combining three modeling assumptions. First, the intensity in the noisy magnitude images follows Rician distributions. Second, different DW images are smooth but have similar edge structures. This type of edge constraints was first used for denoising diffusion MRI data in [1]. We use the same joint edge-preserving regularization framework to enforce the spatial edge constraints in our new algorithm. Third, the diffusion weighted image sequence has a low rank structure. Low-rank models have previously been used for undersampled image reconstruction [11–13] and denoising [14] in various MR contexts. The combination of these three different but complementary components is utilized within the *maximum a posteriori* (MAP) estimation formulation, which we optimize using an efficient alternating minimization algorithm.

The rest of this paper is organized as follows. Section II describes the proposed joint denoising formulation and explains its components in detail. Section III shows representative results obtained with the proposed method from simulated and experimental data sets. Section IV concludes the paper.

2. PROPOSED METHOD

Suppose we are given a sequence of Q noisy DW magnitude images

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_Q],$$

where the vector $\mathbf{y}_i \in R^M$ is an individual DW image with M pixels and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_Q]$ is the underlying noiseless image sequence. We address the denoising problem using a MAP estimation formulation as follows:

$$\begin{aligned} \hat{\mathbf{X}} &= \arg\max_{\mathbf{X}} \log \mathrm{p}\left(\mathbf{X} \mid \mathbf{Y}\right) \\ &= \arg\min_{\mathbf{X}} -\log \mathrm{p}\left(\mathbf{Y} \mid \mathbf{X}\right) - \log \mathrm{p}\left(\mathbf{X}\right), \end{aligned} \tag{1}$$

where $\mathrm{p}\left(\mathbf{Y} \mid \mathbf{X}\right)$ is the likelihood function that enforces data consistency, and $\mathrm{p}(\mathbf{X})$ is the image prior reflecting the intrinsic properties and side information about the images. In the following, we provide details on how these components are modeled in the proposed method.

2.1. Noise Model

It is well accepted that standard MR magnitude images follow a Rician distribution [3, 15]. Assuming independent univariate Rician distributed noise at each voxel, we can express the likelihood $\mathrm{p}(\mathbf{Y} \mid \mathbf{X})$ as

$$\mathrm{p}(\mathbf{Y} \mid \mathbf{X}) = \prod_{q=1}^{Q} \prod_{m=1}^{M} \frac{y_{mq}}{\sigma^2} \exp\left(-\frac{y_{mq}^2 + x_{mq}^2}{2\sigma^2}\right) I_0\left(\frac{y_{mq} x_{mq}}{\sigma^2}\right), \tag{2}$$

where y_{mq} is the noisy intensity of the m^{th} voxel in the q^{th} image and x_{mq} is the corresponding noise-free intensity. $I_0(\cdot)$ is the modified Bessel function of the first kind with order zero and σ^2 is the variance of the real and imaginary parts of the original complex image noise (prior to discarding the phase image). Estimation of the noise variance σ^2 is discussed in Section 2.4.

2.2. Image Model

The proposed image model contains two components: 1) low rank modeling exploiting the spatial-diffusion correlation, 2) a prior term incorporating the correlated edge information from the entire DW image sequence.

Low rank modeling is used to incorporate the prior information that the noiseless diffusion signal at a given voxel is generally linearly dependent on the diffusion signals in other spatial regions. This dependency (referred to as spatial-diffusion correlation) results in a low dimensional structure in the spatial-diffusion matrix. For simple parametric models like diffusion tensor imaging (DTI) [16], low dimensional structure emerges naturally as a consequence of the limited number of degrees of freedom in the tensor model. In addition, even without simplistic diffusion modeling, it is still reasonable to expect correlated diffusion behavior between different image voxels. In this work, we model spatial-diffusion correlation using the Partial Separability (PS) model, which results in a low rank representation of the image sequence [11]. Based on this, we propose to model $\mathbf{X}$ as [12]

$$\mathbf{X} = \mathbf{U}\mathbf{V}, \tag{3}$$

where $\mathbf{U} \in R^{M \times L}$ and $\mathbf{V} \in R^{L \times Q}$ are low rank matrices with $L \ll Q < M$.

The second part of the image model is the edge prior. It has been shown that a sequence of diffusion-weighted images share highly correlated edge structure, which can effectively be used as a constraint for denoising diffusion MRI data [1]. The proposed formulation employs this modeling and uses the following prior, originally proposed in [1], to enforce the edge constraints

$$\mathrm{p}(\mathbf{X}) \propto \exp\left(-\lambda \sum_{m=1}^{M} \sum_{n \in \Omega_m} L\left(\sqrt{\sum_{q=1}^{Q} w_q^2 \left|x_{mq} - x_{nq}\right|^2}\right)\right), \tag{4}$$

where $L(\cdot)$ is the Huber function, Ω_m is a neighborhood of the m^{th} voxel, w_q are weights for the edge information from the q^{th} image, and λ is a parameter which determines the smoothing strength. The expression in (4) spatial-adaptively smoothes the DW images while preserving edge features by including edge information from all the images with weights determined by w_q. The weights can be assigned in a specific way to emphasize different diffusion features. In this work, we choose w_q and Ω_m as described in [1].

2.3. Joint Denoising Formulation

Substituting (2), (3) and (4) into the MAP formulation in (1) yields the following optimization problem:

$$\begin{aligned} \hat{\mathbf{U}}, \hat{\mathbf{V}} \;=\; & \arg\min_{\mathbf{U},\mathbf{V}} \sum_{q=1}^{Q} \sum_{m=1}^{M} \frac{y_{mq}^2 + (\mathbf{U}\mathbf{V})_{mq}^2}{2\sigma^2} \\ & - \log I_0\left(\frac{y_{mq}\,(\mathbf{U}\mathbf{V})_{mq}}{\sigma^2}\right) + \lambda R(\mathbf{U}, \mathbf{V}), \end{aligned} \tag{5}$$

where the regularization function can be expressed as

$$R(\mathbf{U}, \mathbf{V}) = \sum_{m=1}^{M} \sum_{n \in \Omega_m} L\left(\sqrt{\sum_{q=1}^{Q} w_q^2 \left|(\mathbf{U}\mathbf{V})_{mq} - (\mathbf{U}\mathbf{V})_{nq}\right|^2}\right). \tag{6}$$

It has been shown that the regularization term has a line process interpretation which leads to efficient algorithms [1], which we also use here to obtain a multiplicative half-quadratic approximation algorithm. $\mathbf{U}$ and $\mathbf{V}$ are estimated in an alternating fashion [12, 17], and the final estimated image sequence is computed as $\hat{\mathbf{X}} = \hat{\mathbf{U}}\hat{\mathbf{V}}$. For fixed $\mathbf{V}^{(k)}$ at the k^{th} iteration, we update $\mathbf{U}^{(k)}$ by solving

$$\begin{aligned} \mathbf{U}^{(k)} \;=\; & \arg\min_{\mathbf{U} \in R^{M \times L}} \sum_{q=1}^{Q} \sum_{m=1}^{M} \frac{y_{mq}^2 + \left(\mathbf{U}\mathbf{V}^{(k)}\right)_{mq}^2}{2\sigma^2} \\ & - \log I_0\left(\frac{y_{mq}\left(\mathbf{U}\mathbf{V}^{(k)}\right)_{mq}}{\sigma^2}\right) \\ & + \lambda \sum_{q=1}^{Q} \left\| \mathbf{W}\mathbf{F}\left(\mathbf{U}\mathbf{V}^{(k)}\right)_q \right\|_2^2, \end{aligned} \tag{7}$$

where $\mathbf{W}$ is the weighting matrix from the half-quadratic update and $\mathbf{F}$ is a finite difference operator. For fixed $\mathbf{U}^{(k)}$, the optimization problem for solving $\mathbf{V}^{(k)}$ can be factorized into Q decoupled problems [13] as

$$\begin{aligned} \mathbf{v}_q^{(k)} \;=\; & \arg\min_{\mathbf{v}_q \in R^L} \sum_{m=1}^{M} \frac{y_{mq}^2 + \left(\mathbf{U}^{(k)}\mathbf{v}_q\right)_m^2}{2\sigma^2} \\ & - \log I_0\left(\frac{y_{mq}\left(\mathbf{U}^{(k)}\mathbf{v}_q\right)_m}{\sigma^2}\right) \\ & + \lambda \left\| \mathbf{W}\mathbf{F}\mathbf{U}^{(k)}\mathbf{v}_q \right\|_2^2. \end{aligned} \tag{8}$$

Due to the Rician likelihood, nonlinear programming needs to be used to solve the alternating minimization problems. Here we use the limited-memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [18], which was designed for large scale problems where the approximated Hessian can not be saved explicitly, to solve for $\mathbf{U}^{(k)}$. The standard BFGS method is used for solving each subproblem in Eq. (9). We have also derived the explicit expressions for the gradients of the cost functions, which can be computed efficiently using matrix multiplications. The algorithm is initialized by applying a singular value decomposition (SVD) on $\mathbf{Y}$ and extracting the first L eigenvectors as initial guesses for $\mathbf{U}^{(0)}$ and $\mathbf{V}^{(0)}$.

2.4. Practical Implementation

The noise variance σ^2 can be estimated from the background regions with negligible signal [15]. Notice that when $x_{mq} = 0$, the Rician distribution becomes the Rayleigh distribution with variance $E\left(y_{mq}^2\right) = 2\sigma^2$ [15]. Therefore, σ^2 can be estimated as half of the mean square data intensity in the background region. Moreover, assuming that all DW images have approximately the same noise level, the accuracy of this estimation can be improved by averaging σ^2 estimates from all the frames.

The order L of the low rank model was selected empirically by examining the singular value plot of $\mathbf{Y}$ and choosing a proper cut-off threshold. For the edge constraint prior, we assigned slightly larger w_q for images with higher b-values since these images usually contain more important diffusion information than images with lower b-values (such as non-diffusion weighted images with $b = 0$ usually acquired in a diffusion imaging experiment) [1]. The regularization parameter λ was chosen based on the visual quality of both the denoised images and the estimated diffusion parameter maps. Empirical results suggest that the SNR improvement (by visual inspection) and the quantification accuracy remain stable within a large range of λ.

3. RESULTS

The proposed method has been evaluated using both simulated and experimental diffusion imaging data. Some representative results are presented in this section.

3.1. Simulation

DW images were simulated using an adult mouse diffusion tensor atlas from the Biomedical Informatics Research Network (BIRN) Data Repository (mimicking the simulation setups in [1]). The image sequence was generated using the DTI model [16] for each image voxel. Forty-eight DW images with 256×256 spatial resolution were generated, including 4 images with zero b-value and 44 images with a b-value 1000s/mm^2 and different diffusion encoding directions. Independent Rician noise was simulated by adding white Gaussian noise to the real and imaginary components of the sequence and then taking the magnitude. The noise standard deviation was chosen as $\frac{1}{10}$ of the mean intensity in the center region of the first image (the image with $b = 0$).

We applied the proposed method to the simulated data set, and compared the results with the multi-component non-local means (MNLM) denoising algorithm [10]. The low rank model order of the proposed method was $L = 8$ and the regularization parameter was $\lambda = 200$. The parameters of the MNLM algorithm were chosen such that the best quantification results were obtained. Figure 1 shows one representative image frame, calculated fractional anisotropy (FA) maps and color-coded FA maps for comparing denoising performance. The color represents direction information from the primary eigenvector of the estimated diffusion tensors. It can be observed that the proposed method significantly reduces the noise level both in the DW images and the parameter maps. Figure 2 shows a comparison of quantification results from different methods. Both the comparison of image quality and the comparison of FA estimation accuracy (mean errors and standard deviations) further demonstrate the effectiveness of the proposed method in reducing noise contamination and maintaining the anisotropy information of the diffusion tensors.

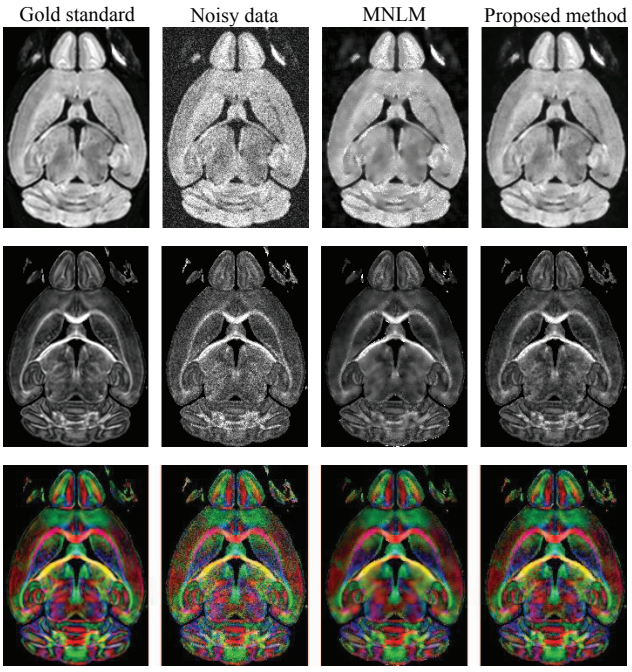


Fig. 1. Denoising results for simulated data. The three rows correspond to the DW images, FA maps and color-coded FA maps from the noiseless images (column 1), noise corrupted images (column 2), images denoised by MNLM algorithm (column 3) and images denoised by the proposed method (column 4). The color scale is based on the orientation of the primary eigenvector of the diffusion tensor: red stands for left-right, green stands for anterior-posterior and blue stands for superior-interior.

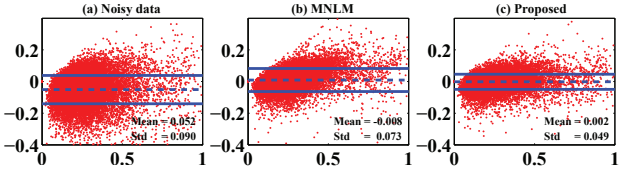


Fig. 2. Plots of the FA estimation errors (true FA value - estimated FA value) versus the true FA values. The mean and standard deviation of the errors corresponding to noisy data, data denoised by MNLM and data denoised by the proposed method are shown.

3.2. Experimental data

In vivo brain DTI data was acquired with a 3T Siemens scanner equipped with a 32 channel receive headcoil and using a double spin-echo EPI sequence (TR/TE=6300/98ms, 128x128 matrix size, 256x256mm^2 FOV, and 35 slices). The DW images were acquired at multiple b-values (b=1000, 2000, 3000s/mm^2), with 30 diffusion encoding directions acquired for each b-value. Partial Fourier reconstruction was applied to each coil as described in [1]. The proposed method was then applied to denoise the reconstructed images with $L = 12$ and $\lambda = 10^5$. Sum-of-squares was used to combine images from all the coils for diffusion tensor estimation.

Figure 3 illustrates the performance of the proposed method with

one slice from the data set. Although quantitative improvement can not be shown due to lack of a gold standard, qualitative improvement is evident in the denoised images. As can be seen, the proposed method not only significantly reduces the noise in the DW image, but also largely recovers the diffusion tensor information originally corrupted by noise.

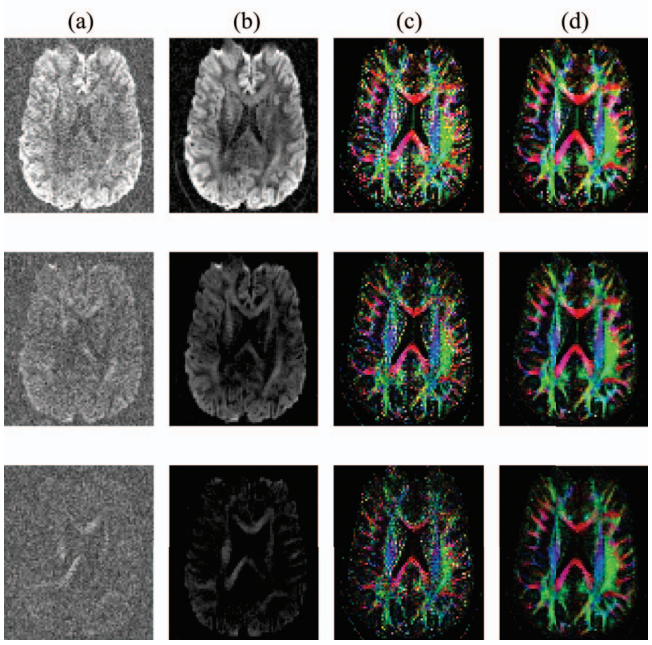


Fig. 3. Denoising results for experimental data (one slice). One representative frame of the noisy DW images (a), the corresponding denoised image (b), and the color-coded FA maps calculated from the noisy data (c) and the denoised data (d) are shown. The corresponding results for different b-values ($b = 1000$ for the first row, $b = 2000$ for the second row, $b = 3000$ for the third row) are shown.

4. CONCLUSIONS

A new method has been proposed for jointly denoising a sequence of diffusion-weighted magnitude images. The proposed method formulates the denoising problem as a MAP estimation problem using Rician signal modeling, low rank modeling and joint edge constraints. Simulated DTI data and experimental multiple b-value diffusion weighted imaging data have been used to evaluate the effectiveness of the proposed method. Qualitative and quantitative improvements have been demonstrated. We expect the proposed method to be useful for achieving higher measurement accuracy and/or reducing data acquisition time for quantitative diffusion MRI with magnitude data.

5. REFERENCES

[1] J. P. Haldar, V. J. Wedeen, M. Nezamzadeh, G. Dai, M. W. Weiner, N. Schuff, and Z.-P. Liang, "Improved diffusion imaging through SNR-enhancing joint reconstruction," *Magn Reson Med*, 2012, In Press.

[2] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Trans Image Process*, vol. 8, pp. 1408 – 1419, 1999.

[3] J. L. R. Andersson, "Maximum a posteriori estimation of diffusion tensor parameters using a Rician noise model: Why, how and but," *NeuroImage*, vol. 42, pp. 1340 – 1356, 2008.

[4] V. Patel, Y. Shi, P. M. Thompson, and A. W. Toga, "K-SVD for HARDI denoising," pp. 1805 – 1808, 2011.

[5] S. Basu, T. Fletcher, and R. Whitaker, "Rician noise removal in diffusion tensor MRI," *MICCAI*, vol. 9, pp. 117 – 125, 2006.

[6] T. McGraw, B. Vemuri, E. Ozarslan, Y. Chen, and T. Mareci, "Variational denoising of diffusion weighted MRI," *Inverse Problems and Imaging*, vol. 3, pp. 625 – 648, 2009.

[7] S. Aja-Fernandez, M. Niethammer, M. Kubicki, M. E. Shenton, and C.-F. Westin, "Restoration of DWI data using a Rician LMMSE estimator," *IEEE Trans Med Imag*, vol. 27, pp. 1389 – 1403, 2008.

[8] N. Wiest-Daessle, S. Prima, P. Coupe, S.P. Morrissey, and C Barillot, "Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: applications to DT-MRI," *MICCAI*, vol. 11, pp. 171 – 179, 2008.

[9] J. V. Manjon, P. Coupe, L. Marti-Bonmati, D. L. Collins, and M Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels," *J Magn Reson Imag*, vol. 31, pp. 192 – 203, 2010.

[10] J. V. Manjon, N. A. Thacker, J. J. Lull, G. Garcia-Marti, L. Marti-Bonmati, and M. Robles, "Multicomponent MR image denoising," *Int J Biomed Imaging*, 2009.

[11] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," pp. 988 – 991, 2007.

[12] J. P. Haldar and Z.-P. Liang, "Spatiotemporal imaging with partially separable functions: A matrix recovery approach," pp. 716 – 719, 2010.

[13] B. Zhao, J. P. Haldar, C. Brinegar, and Z.-P. Liang, "Low rank matrix recovery for real-time cardiac MRI," pp. 996 – 999, 2010.

[14] H. M. Nguyen, X. Peng, M. N. Do, and Z.-P. Liang, "Spatiotemporal denoising of MR spectroscopic imaging data by low-rank approximations," pp. 857 – 860, 2011.

[15] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn Reson Med*, vol. 34, pp. 910–914, 1995.

[16] D. Le Bihan, J. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat, "Diffusion tensor imaging: Concepts and applications," *J Magn Reson Imag*, vol. 13, pp. 534 – 546, 2001.

[17] J. P. Haldar and D. Hernando, "Rank-constrained solutions to linear matrix equations using PowerFactorization," *IEEE Signal Process Lett*, vol. 16, pp. 584 – 587, 2009.

[18] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 2006.