FAST DENOISING TECHNIQUES FOR TRANSVERSE RELAXATION TIME ESTIMATION IN MRI

Marcus Björk and Petre Stoica*

Department of Information Technology, Uppsala University, Uppsala, Sweden

ABSTRACT

Estimating the transverse relaxation time (T_2) from magnitude spin echo images is a common problem in Magnetic Resonance Imaging. The standard approach is to use pixelwise estimates; however these estimates can be quite noisy when only two images are available. By imposing inter-pixel information the estimates can be improved. An optimal formulation of the problem is nonlinear and typically time consuming to solve. Here we propose two fast methods to reduce the variance of the T_2 estimates: 1) a simple local least squares (LS) method, and 2) a total variation based approach that can be cast as a linear program. The two approaches are evaluated using both simulated and in vivo data. We show that the variance of the proposed T_2 estimates is smaller than the pixelwise estimates without compromising tissue contrast.

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a useful tool for non-invasively studying internal structures in the body, as for example the brain. Signal processing has many applications in MRI, for example parameter estimation and image denoising [1, 2]. Quantitative MRI has brought new ways of imaging based on tissue specific physical quantities, such as mapping the longitudinal and transverse relaxation times, denoted T_1 and T_2 respectively. The relaxation times describe the decay of the magnetization vector after excitation. One standard approach to estimate T_2 is to acquire several images at different echo times using a sequence called the spin echo sequence; and then fit a decaying exponential to the magnitude data in each pixel individually [3]. The magnitude images are Rice distributed meaning that the least squares (LS) estimate will be suboptimal. A few suggestions of how to solve this problem are available in the literature [4, 5], where the authors apply maximum likelihood (ML) methods taking the Rice distribution into account. To further denoise the images or the resulting estimates, techniques like total variation (TV) can be used [6]. The idea is to regularize the problem by penalizing high total variation in the image, usually defined by the L1 norm of some first order difference measure. The resulting estimates tend to be piecewise constant, which can be a problem in MRI where more gradual changes also can occur. The authors in [6] also present a total generalized variation regularization that can be used for image denoising while suppressing these so called staircase artifacts. Solving the optimization problem resulting from a TV based approach can however be nonlinear and time consuming, and the authors of [6] propose an implementation on a graphics processing unit to reduce computation time.

In this paper we treat the case when only two magnitude images are available, in which case the T_2 estimates can be rather noisy. Applying advanced statistical methods to account for the Rice distributed noise is not likely to be fruitful when only two samples are available in each pixel. Moreover, at common signal-to-noise ratios (SNR) the Rice distribution will be accurately approximated by a Gaussian. We focus on the problem of variance reduction of the T_2 estimates. The idea is to improve the estimates by using inter-pixel information in the images while preserving the contrast between tissues. We propose two methods: 1) a fast local least squares (LS) method which is easy to implement, and 2) a more general TV based method that can be cast as a linear program (LP). Due to the efficiency of solving LPs, this approach is computationally more efficient than the standard TV. Both methods provide an intuitive way to choose the user parameters. We then compare the performance and computation time of the proposed methods to the pixelwise counterpart.

2. THEORY

2.1. Signal model

The received signal in MRI is commonly complex-valued and Gaussian distributed. In this paper we solve the estimation problem using magnitude data, which is typical for spin echo T_2 estimation. The resulting data samples are Rice distributed with the probability distribution function (PDF)

$$p_{\mathrm{R}}(x|\nu,\sigma) = \frac{x}{\sigma^2} e^{\frac{-(x^2+\nu^2)}{2\sigma^2}} I_0\left(\frac{x\nu}{\sigma^2}\right), \tag{1}$$

where ν is the signal magnitude; σ is the Gaussian noises standard deviation in the real and imaginary parts of the corresponding complex-valued data; I_0 is the modified Bessel

*Correspondence to marcus.bjork@it.uu.se.

This work was partially funded by the European Research Council (ERC) Advanced Grant, number 247035.

Thanks to Dr. Joel Kullberg in the section of Radiology at the Uppsala University Hospital for providing the in vivo data.

function of the first kind and order zero. The data at an arbitrary pixel can then be modeled as an observation $s_{\mathrm{R}}(t_i)$ of a Rice distributed stochastic variable S_{R} parameterized by ρ, T_2 and σ [7, 8], that is

$$\begin{aligned} s_{\mathrm{R}}(t_i) &= S_{\mathrm{R}}(\rho, T_2, \sigma, t_i), \\ S_{\mathrm{R}}(\rho, T_2, \sigma, t_i) &\sim \mathrm{Rice}(\rho e^{-t_i/T_2}, \sigma). \end{aligned} \tag{2}$$

where $t_i < t_{i+1}, \forall i$ are the echo times of the images. However, the Rician mean and variance converge towards the Gaussian counterparts as SNR increases, as is shown in Fig. 1. The SNR is defined as $\mathrm{SNR} = (s_1 + s_2)/2\sigma$, which is the definition commonly used in MRI. The reason for avoiding the root mean square SNR definition here is because it will give less weight to the second sample of the damped signal. At an SNR larger than 20 dB, the Rice distribution is accurately approximated by a Gaussian, meaning that least squares (LS) will be close to maximum likelihood. We can then approximately model the data as

$$s(t_i) = \rho e^{-t_i/T_2} + v(t_i), \tag{3}$$

where $v(t_i)$ is independent Gaussian noise. Due to the high complexity associated with using a Rician distribution compared to the accuracy gained, we propose to use the Gaussian approximation model of (3) in regions containing signal.

By assuming that there is no noise, the transverse relaxation time T_2 can be estimated from (3) by simple algebra. Given that ρ and T_2 are positive real-valued, we have

$$\lambda \triangleq \frac{s(t_2)}{s(t_1)} = e^{(t_1 - t_2)/T_2}, \tag{4}$$

$$T_2 = (t_1 - t_2)/\ln(\lambda). \tag{5}$$

By applying (5) for each pixel p, we can estimate T_2 in an entire image. We will refer to this method as the pixelwise approach. The variable ρ in (3) represents the initial magnetization or the proton density. An estimate of ρ can be obtained by LS assuming that T_2 is known, but it will not be discussed further here. The estimate in (5) is actually the nonlinear least squares (NLS) estimate, obtained by solving the problem

$$\underset{\rho, T_2}{\text{minimize}} \sum_{i=1}^{2} \left(s_i - \rho e^{-t_i/T_2} \right)^2, \tag{6}$$

where we have denoted $s_i = s(t_i)$ for ease of notation. This follows from the fact that there is always an exact solution to (6) for $s_i \geq 0$, and we can therefore reformulate this problem to implicitly estimate only T_2 as

$$\underset{\lambda}{\text{minimize}} \; (s_2 - \lambda s_1)^2, \tag{7}$$

which clearly has the solution in (5).

2.2. Noise variance estimation

To estimate the standard deviation of the Gaussian noise σ from the signal free background of the images, the Rician

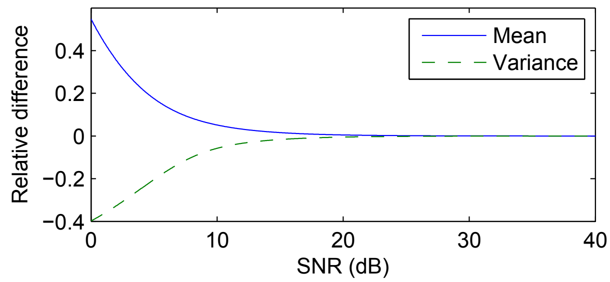

Fig. 1. The relative difference between in mean and variance for the Rice distribution and those of the Gauss distribution, versus the SNR.

distribution is used. More exactly, the Rayleigh distribution is used which is a special case of the Rician distribution when $\nu = 0$ in (1). The maximum likelihood estimate of σ is [4]

$$\hat{\sigma} = \sqrt{\frac{1}{4N_b} \sum_{p \in R_b} (s_{1p}^2 + s_{2p}^2)}, \tag{8}$$

where R_b is the set of all background pixels and N_b is its cardinality. Similarly, we will denote the cardinality of any set R_x as N_x. Note that the corresponding pixels in both images are used, since σ is assumed to be constant. In practice the pixels containing signal are separated from the noise only background by thresholding the first image and keeping pixels above a fixed percentage of the maximum intensity. This procedure defines the set R_b in (8), as well as the set of signal pixels R_s used in the proposed algorithms.

3. METHOD

By using the definition in (4), we can express the problem in (7) in matrix form by stacking the data from each image in column vectors $\mathbf{s}_i$

$$\underset{\boldsymbol{\lambda}}{\text{minimize}} \; \|\mathbf{s}_2 - \mathrm{diag}(\mathbf{s}_1)\boldsymbol{\lambda}\|_2^2, \tag{9}$$

where the bold lower case variables are column vectors and $\mathrm{diag}(\mathbf{v})$ is a square matrix with the elements of $\mathbf{v}$ along its diagonal. The problem in (9) is redundant in its current form since there is no coupling between the pixels. However, if we impose the constraint that the estimates are smooth over the image, in a certain sense (see below), we get a coupling between pixels and an optimization formulation becomes vital.

3.1. Local Least Squares approach

A simple method to reduce the variance is to assume that locally over a small region around the pixel of interest, the value of T_2 is constant. By using this assumption the estimate of λ in (7) can be obtained by linear LS over this region. Here we will avoid smoothing across tissue borders by detecting sudden large changes in the pixelwise T_2 estimates obtained from (5). Based on this information we adjust the neighborhood region R_p used in LS (see below). The algorithm then estimates λ, and hence T_2, assigns this estimate to the center pixel only, and then moves to the next pixel. We call this method the localLS.

We can formulate the problem in each region R_p, identified by the center pixel p, as

$$\underset{\lambda_p}{\text{minimize}} \sum_{j \in R_p} (s_{2j} - \lambda_p s_{1j})^2, \qquad \forall p \in R_s, \tag{10}$$

where R_s is the set of all pixels containing signal. Denoting the vector of neighborhood data points by $\mathbf{s}_{1p}$ and $\mathbf{s}_{2p}$ we get

$$\hat{\lambda}_p = \frac{\mathbf{s}_{1p}^T \mathbf{s}_{2p}}{\mathbf{s}_{1p}^T \mathbf{s}_{1p}}, \qquad \hat{T}_{2p} = -\frac{t_2 - t_1}{\ln(\hat{\lambda}_p)}. \tag{11}$$

The region R_p is initially defined as the intersection of R_s and the 8 pixels surrounding pixel p. The pixelwise estimates are then computed and checked for feasibility, that is $0 < T_2 < \infty$, and any infeasible pixels are removed from R_p. If the center pixel is infeasible, it is initially set to the median value of the feasible surrounding estimates. The outlier removal procedure then further limits the set R_p. This is done by computing the Cramér-Rao Bound (CRB) [9] of T_2 using the center pixel as the true value and the estimated noise standard deviation from (8). The CRB gives a lower bound on the parameter standard deviation, which is used as an estimate of the variation in the current area. Any pixel T_2 estimate further away than $k_{\text{LS}}\sigma$ from the center T_2 estimate is considered an outlier. The threshold k_{LS} is a user parameter that will be discussed further in section 4. It is straightforward to choose a different size of the surrounding neighborhood than the 3x3 pixels used here, but a preliminary study showed no clear gain in doing so, and it is therefore kept constant and not discussed further in this paper.

The problem stated in (10) is not equivalent to the NLS as in the pixelwise case (5). This is due to the reparameterization using λ instead of T_2. By expressing the NLS for the region R_p in vector form, where $\mathbf{r}(T_{2p}) = [e^{-t_1/T_{2p}} \; e^{-t_2/T_{2p}}]^T$ and $\tilde{\mathbf{s}}_j = [s_{1j} \; s_{2j}]^T$, we get

$$\underset{\rho_j, T_{2p}}{\text{minimize}} \sum_j \|\tilde{\mathbf{s}}_j - \rho_j \mathbf{r}(T_{2p})\|_2^2. \tag{12}$$

By LS we have that for each j

$$\rho_j(T_{2p}) = \frac{\mathbf{r}^T(T_{2p})\tilde{\mathbf{s}}_j}{\mathbf{r}^T(T_{2p})\mathbf{r}(T_{2p})}. \tag{13}$$

Substituting this into (12) we can rewrite the NLS criterion as

$$\underset{T_{2p}}{\text{minimize}} \sum_{j \in R_p} \tilde{\mathbf{s}}_j^T \left(I - \frac{\mathbf{r}(T_{2p})\mathbf{r}^T(T_{2p})}{\mathbf{r}^T(T_{2p})\mathbf{r}(T_{2p})} \right) \tilde{\mathbf{s}}_j, \tag{14}$$

or by introducing

$$A(T_{2p}) = I - \frac{\mathbf{r}(T_{2p})\mathbf{r}^T(T_{2p})}{\mathbf{r}^T(T_{2p})\mathbf{r}(T_{2p})}, \tag{15}$$

$$S = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1N_p} \\ s_{21} & s_{22} & \dots & s_{2N_p} \end{bmatrix}, \tag{16}$$

we get

$$\underset{T_{2p}}{\text{minimize}} \; \text{tr}(S^T A(T_{2p}) S). \tag{17}$$

By using the estimates obtained form (11) as an initial guess, a nonlinear minimization method can be applied to solve the problem in (12) and obtain the NLS estimate of T_2 in the region R_p. The method obtained by adding this NLS tuning step is abbreviated as the localNLS.

3.2. L1 Total Variation approach

In this approach we suggest imposing a global smoothing criterion such that each estimate of T_2 is *close* to the neighboring estimates. The problem is typically formulated as a quadratic fitting with a total variation constraint based on the L1 norm. However, this leads to a quadratic program, which can be quite demanding to solve for the problem size at hand. Therefore we propose to use a L1 fitting leading to an LP that can be solved efficiently. Furthermore, we formulate the problem as a minimization of the total variation, subject to a bound on the L1 norm of the error. We have

$$\begin{aligned} \underset{\boldsymbol{\lambda}}{\text{minimize}} \quad & \sum_{p \in R_s} \sum_{j \in R_p} |\lambda_p - \lambda_j| \\ \text{subject to} \quad & \|\mathbf{s}_2 - \text{diag}(\mathbf{s}_1)\boldsymbol{\lambda}\|_1 \leq k_{\text{TV}} \hat{\sigma} N_s \\ & \mathbf{0} < \boldsymbol{\lambda} < \mathbf{1}, \end{aligned} \tag{18}$$

where $\mathbf{0}$ and $\mathbf{1}$ are the column vectors of appropriate length containing zeros and ones respectively, and $\boldsymbol{\lambda}$ is the vector of all $\lambda_p \in R_s$. Using this formulation, the right hand side of the constraint can be estimated by applying the Cauchy-Schwarz inequality to the error vector $\boldsymbol{\epsilon}$

$$\|\boldsymbol{\epsilon}\|_1 = \mathbf{1}^T |\boldsymbol{\epsilon}| \leq \|\epsilon\|_2 \|\mathbf{1}\|_2 = \hat{\sigma} N_s, \tag{19}$$

where the last equality holds if the errors have zero mean. The proportionality constant k_{TV} is a user parameter that will be discussed further in section 4. The use of the L1 norm allows for piecewise smoothness, and does not penalize sudden jumps in the estimate as severely as the L2 norm. We also require that T_2 is positive and finite, corresponding to $\mathbf{0} < \boldsymbol{\lambda} < \mathbf{1}$. It is possible to set stricter bounds on T_2 using prior information, but this possibility will not be pursued in this paper. By introducing two new variables $\mathbf{w}$ and $\mathbf{z}$, we can write the problem as

$$\begin{aligned} \underset{\boldsymbol{\lambda}, \mathbf{z}, \mathbf{w}}{\text{minimize}} \quad & \mathbf{1}^T z \\ \text{subject to} \quad & -\mathbf{z} \leq \big(\text{bdiag}(\mathbf{1}_{N_p}) - B\big)\, \boldsymbol{\lambda} \leq \mathbf{z} \\ & -\mathbf{w} \leq \mathbf{s}_2 - \text{diag}(\mathbf{s}_1)\boldsymbol{\lambda} \leq \mathbf{w} \\ & \mathbf{1}^T \mathbf{w} \leq k_{\text{TV}} \sigma N_s \\ & \mathbf{0} < \boldsymbol{\lambda} < \mathbf{1}, \end{aligned} \tag{20}$$

where the B matrix specifies the unique neighbors of each pixel and $\mathrm{bdiag}(\mathbf{1}_{N_p})$ is the block diagonal matrix with the vectors $\mathbf{1}$ of length N_p in each diagonal block. The problem in (20) is an LP [10], and solving it for $\boldsymbol{\lambda}$ gives the variance reduced estimates of T_2. We call this method L1TV.

4. RESULTS

4.1. Simulations

It can be shown empirically that the optimal echo times, in terms of the CRB, for estimating T_2 can be accurately described by the relation $t_2 = 1.1T_2 + t_1$. This relies on the Gaussian assumption and hence only holds approximately for high SNR. In practice the true T_2 is not known and will vary over the image, but the tissue of interest can still provide an approximate expected T_2. In terms of actual CRB values, choosing t_1 as low as possible is favorable, which can be expected since the signal decays in time.

A simulated dataset in Rician noise was created to be able to quantitatively compare the performance of the proposed methods to the pixelwise approach. The echo times where chosen as $t_1 = 21$ ms and $t_2 = 100$ ms to mimic the in vivo data in section 4.2. The simulated data was set to contain a range of T_2 values from 50 ms to 300 ms, and several types of T_2 variations: gradients, large and small abrupt changes, fine detail and larger constant areas; while ρ was kept constant, since it only affects the SNR. The total mean square error (MSE) was computed for different values of the user parameters k_{LS} and k_{TV}, and is shown in Fig. 2. As can be seen the MSE is minimized for $k_{\mathrm{LS}} > 2.5$ in the localLS method, and L1TV has a clear minimum around $k_{\mathrm{TV}} = 0.8$. The results were similar for different SNRs, although a small variation could be seen. As a compromise between visual appearance, bias, and MSE, default values of $k_{\mathrm{LS}} = 2$ and $k_{\mathrm{TV}} = 0.5$ were set for the localLS and L1TV respectively. It can also be seen from Fig. 2 that the localNLS has similar performance to localLS. Visually the results where indistinguishable and the localNLS will therefore not be studied separately in the following.

The resulting T_2 estimates are shown in Fig. 3. The localLS and L1TV have similar performance in terms of MSE, and show a clear reduction in the variance. However, the L1TV has a slight problem resolving the thin bright lines on the dark background compared to localLS.

4.2. In vivo data

Two spin echo images of a brain at echo times $t_1 = 21$ ms and $t_2 = 100$ ms where acquired using a 1.5 T scanner. The average image SNR was 31 dB, meaning that the Gaussian assumption should be valid. By employing the user parameters suggested from the simulations, the T_2 estimates were calculated using the proposed methods, localLS and L1TV, and compared to the pixelwise approach. The results are shown in Fig. 4. The images are truncated at 300 ms to make the results in the target (lower) range of T_2 values more clear.

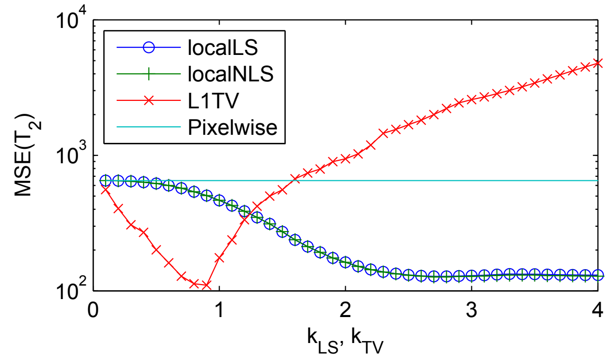

Fig. 2. The average MSE of the T_2 estimates for the simulated data versus the parameters k_{LS} and k_{TV} that controls the smoothness enforced.

Again, both the localLS and L1TV show a clear reduction in the variance without compromising contrast. Visually there is no clear gain in using the L1TV compared to the simpler localLS approach. By examining the difference between the estimates obtained by the proposed methods and pixelwise approach, it can be seen that the localLS has a more uniform noise reduction, indicated by the low amount of structure in the difference image (not shown, due to space limitations). Only minor staircase artifacts can be seen in the L1TV estimates for the chosen degree of smoothing. The relative difference of the LS estimates compared to the NLS (not shown) were less than 1% in over 99 % of the pixels. This indicates that the LS gives a good approximation of the NLS, and the extra computational burden involved in acquiring the NLS estimates is not motivated.

The MOSEK linear program solver was run through MATLAB on a Intel Core i7 860 system at 2.93 GHz and with 16 GB of memory. The average run times for the different methods when applied to the brain data (25800 pixels) were 2.3 s for the localLS and 27.5 s for L1TV, while the pixelwise approach is instant.

5. CONCLUSIONS

We have proposed two methods that can be used to reduce the variance of the T_2 estimates obtained from two spin echo images, without compromising the resolution at tissue boundaries. Both localLS and L1TV include a way of choosing the user parameters, k_{LS} and k_{TV} respectively, and automatically adapt to the local image conditions. The L1TV decides what is considered to be an outlier based on the whole image, not just the center pixel versus the surrounding as in LS. Futhermore, L1TV is set in a more general optimization framework. The localLS algorithm, on the other hand, is easy to use, computationally efficient, and generally gives similar or superior performance compared to the L1TV approach. It is also easy to implement and memory efficient enough to be generally applicable for T_2 estimation on any standard computer.

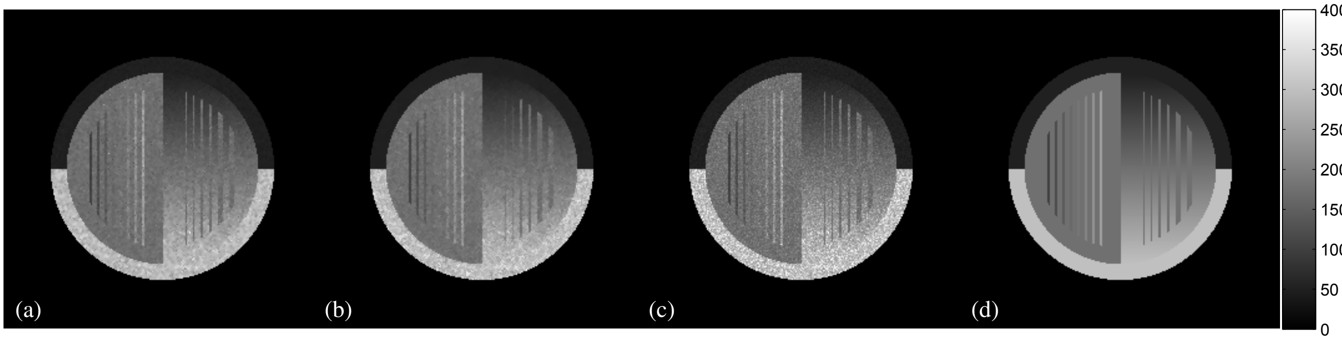

Fig. 3. Estimates of T_2 obtained by: (a) localLS ($\mathrm{MSE} = 172$), (b) L1TV ($\mathrm{MSE} = 209$), and (c) pixelwise ($\mathrm{MSE} = 652$), together with (d) the true values.

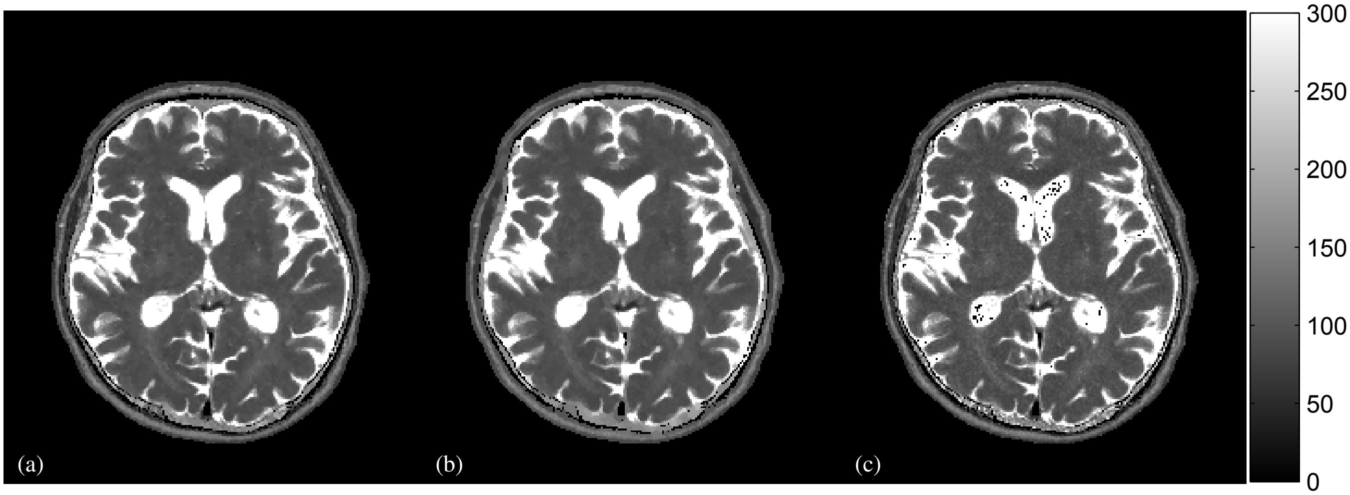

Fig. 4. Estimates of T_2 obtained by: (a) localLS, (b) L1TV, and (c) pixelwise approach.

6. REFERENCES

[1] Y.-Y. Chang, K. A. Kraft, and A. Docef, "On improving the accuracy of T1 mapping of the human brain," in *Proc. 19th EUSIPCO*, Barcelona, Spain, 2011.

[2] J. Aelterman, B. Goossens, A. Pizurica, and W. Philips, "Removal of correlated rician noise in magnetic resonance imaging," in *Proc. 16th EUSIPCO*, Lausanne, Switzerland, 2008.

[3] H. Y. Carr and E. M. Purcell, "Effects of diffusion on free precession in nuclear magnetic resonance experiments," *Physical Review*, vol. 94, pp. 630–638, 1954.

[4] J. Sijbers, A. J. den Dekker, E. Raman, and D. Van Dyck, "Parameter estimation from magnitude MR images," *International Journal of Imaging Systems and Technology*, vol. 10, no. 2, pp. 109–114, 1999.

[5] S. Walker-Samuel, M. Orton, L. D. McPhail, J. K. R. Boult, G. Box, S. A. Eccles, and S. P. Robinson, "Bayesian estimation of changes in transverse relaxation rates," *Magnetic Resonance in Medicine*, vol. 64, no. 3, pp. 914–921, 2010.

[6] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, "Second order total generalized variation (TGV) for MRI," *Magnetic Resonance in Medicine*, vol. 65, no. 2, pp. 480–491, 2011.

[7] J.-M. Bonny, M. Zanca, J.-Y. Boire, and A. Veyre, "T2 maximum likelihood estimation from multiple spin-echo magnitude images," *Magnetic Resonance in Medicine*, vol. 36, no. 2, pp. 287–293, 1996.

[8] J. Sijbers, A. J. den Dekker, M. Verhoye, E. Raman, and D. Van Dyck, "Optimal estimation of T2 maps from magnitude MR images," *Proc. of SPIE Medical Imaging*, vol. 3338, pp. 384–390, 1998.

[9] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice Hall, 2005.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.