PARAMETER SELECTION FOR OPTIMIZED NON-LOCAL MEANS FILTERING OF TASK FMRI

Jian Li, Richard M. Leahy

Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089, USA

ABSTRACT

Non-local means (NLM) filtering of fMRI can reduce noise while preserving spatial structure. We have developed a variant called temporal-NLM (tNLM) which uses similarity in time-series between voxels as the basis for computing the weights in the filter. Using tNLM, dynamic fMRI data can be denoised while spatial boundaries between functionally distinct areas in the brain tend to be preserved. The degree of smoothing in tNLM is determined by a parameter h. Here we describe a procedure for selection of h to optimize our ability to differentiate functionally discrete brain regions. We demonstrate the method in application to optimized filtering of task fMRI data.

***Index Terms*—** Optimization, non-local means, fMRI, networks

1. INTRODUCTION

Dynamic functional MRI indirectly reflects neuronal activity by measuring blood-oxygen level dependent (BOLD) changes in image contract. Strong temporal correlations are evident in these data between physiologically related cortical regions, even during periods of rest [1]. Quantifying correlations between multiple brain areas forms the basis for brain network identification from resting fMRI [2]. While the role and analysis of task fMRI are quite different, the BOLD time-series in these data will similarly show strong correlations within and between regions that respond to a particularly cognitive challenge [3].

BOLD-related contrast changes are small and noise in the data limit our ability to reliably identify regions of spontaneous or task-related activity. Both resting and task-related fMRI are typically spatially filtered to reduce noise, in addition to other preprocessing steps. Even after extensive preprocessing, identification of brain networks from correlation patterns in resting fMRI and identification of active regions using a generalized linear model (GLM) in task fMRI can still be challenging.

Spatial smoothing is applied to the fMRI data either volumetrically with 3D isotropic Gaussian kernel [4], or on data mapped onto a 2D representation of the cortical surface using the Laplace-Beltrami operator [5]. Both methods suffer from the common problem that in addition to reducing noise, they also inevitably spatially mix signals between adjacent regions of functional specialization.

Recently, non-local means (NLM) filtering has been applied for structural-preserving denoising of anatomical MRI [6], [7], fMRI [8], [9] and diffusion MRI [10]. All of these applications compute the NLM kernels based on spatial similarity measures similar to that in the paper that originally described this approach [11]. Recently we described a variant on NLM that filters spatio-temporal data based on measures of similarity in the time series rather than spatial similarity [12]. We refer to this as temporal NLM (tNLM) and demonstrated its effectiveness in denoising of resting fMRI data. By filtering based on temporal similarity, tNLM will reduce noise by averaging over regions in the image that have similar functional roles without blurring across functional boundaries between regions with distinct temporal activity.

Performance of tNLM is dependent on a parameter h that determines the form of the mapping from temporal correlation to the filter kernel weights. Too small a choice will cause inadequate smoothing and less SNR improvement. On the other hand, too large a value will result in over-smoothing and blurring between functional regions. Buades *et al.* [11] empirically suggested setting $h = 10 \times \sigma$, where σ is the standard deviation of the noise. In application to MRI, Manjn *et al.* [6] exhaustively searched the parameter space and found $h = 1.2 \times \sigma$ to be the best choice. Since the optimal h is not only a function of the noise level, but also a function of the block size over which filtering is performed, Coupé *et al.* [7] developed a method for selecting h automatically by normalizing the l_2 distance and estimating the noise variance. They found that an optimal choice of $h = \sqrt{2\beta}\hat{\sigma}$, where β is a manually-tuned constant. None of these approach extends directly to tNLM where the temporal nature of the data needs to be accounted for. Further these methods are not optimal for the key application considered here: preservation of discrete functional regions while also reducing noise.

Here we describe a new approach to selecting the tNLM parameter h to optimize differentiation of regions that are functionally connected from those that are not. We describe the approach below and present evaluation based on simulated and experimental task fMRI data.

This work supported by NIH Grants R01 NS089212, R01 EB009048 and R01 NS074980

2. METHODS

2.1. Temporal NLM Filtering

We apply tNLM filtering to fMRI data defined on the vertices of a tessellated representation of the mid-cortical surface. Let $s(i,t)$ be the time series data at vertex i and time t. Let $\mathcal{N}(i)$ denote the set that contains i and all of its k-hop neighbors. Then tNLM filtering is defined as

$$s'(i,t) = \frac{1}{\sum_{j\in\mathcal{N}(i)} w(i,j)} \sum_{j\in\mathcal{N}(i)} s(j,t)w(i,j) \quad (1)$$

where $w(i,j)$ is the weight applied when averaging across vertices. This weight depends on a temporal measure of similarity, which we define as

$$w(i,j) = \exp\left(-\frac{\frac{1}{T}\left\|\frac{\vec{s}(i)}{||\vec{s}(i)||} - \frac{\vec{s}(j)}{||\vec{s}(j)||}\right\|^2}{h^2}\right) \quad (2)$$

where $\vec{s}(i) = [s(i,1),\ldots,s(i,T)]^\top$ is the vector representation of the time series at vertex i with length T and h is the scalar parameter that controls the degree of smoothing. As noted in [12], the distance $\frac{1}{T}\left\|\vec{s}(i)/||\vec{s}(i)|| - \vec{s}(j)/||\vec{s}(j)||\right\|^2$ between any pair of vertices in (2) can be expressed as $2 - 2\times\hat{r}(\vec{s}(i),\vec{s}(j))$ where $\hat{r}$ is the sample correlation coefficient between $\vec{s}(i)$ and $\vec{s}(j)$.

We can rewrite tNLM filtering in matrix form. Let $X \in \mathbb{R}^{N\times T}$ be the data matrix with N vertices and T time samples. Then the weighting matrix W can be expressed as

$$W(i,j) = \begin{cases} \exp\left(-\frac{2(1-A(i,j))}{h^2}\right) & , j\in\mathcal{N}(i) \\ 0 & , j\notin\mathcal{N}(i) \end{cases} \quad (3)$$

where $A = XX^\top \in \mathbb{R}^{N\times N}$ is the data correlation matrix. We can further define the degree matrix D to be a diagonal matrix whose diagonal element $d_{ii} = \sum_j W_{i,j}$. Then the tNLM filtered signal Y can be written as

$$Y = D^{-1}WX \quad (4)$$

2.2. Optimization of tNLM Parameter h

The parameter h in tNLM filtering determines the degree of noise reduction and smoothing. Here we focus on its application to spatio-temporal fMRI data representing brain networks. Each network is made up of a number of discrete areas of functional specialization in the brain. We want to select the parameter to maximize the SNR in the tNLM-filtered data within each network but without mixing the signals between them. To do this we assume that the matrix $W = f(A)$, equation (3), defines a graph with nodes representing the vertices on the cerebral cortex and the edge strength between them given by the elements of W. Our goal is to select h so that the graph W optimally differentiates, in terms of edge strength, between pairs of vertices in the same network, and pairs in different networks. In this way the tNLM filter will lead to improved SNR while minimizing mixing of signals between distinct functional networks.

Let the observed signal be $x_i = s_i + n_i$, a superposition of the true signal s_i and the noise n_i at vertex i. Assume s_i and n_i are independent with $s_i \sim \mathcal{N}(0,\sigma_s^2)$ and $n_i \sim \mathcal{N}(0,\sigma_n^2)$. Further assume perfect correlation within each network (H_1) and zero between networks (H_0) with respect to the true signal s_i. Then the correlation has the following form:

$$\rho = \frac{E[x_i x_j]}{\sigma_{x_i}\sigma_{x_j}} = \begin{cases} 0 & , H_0 : \frac{E[s_i s_j]}{\sigma_s^2} = 0 \\ \frac{\sigma_s^2}{\sigma_s^2+\sigma_n^2} & , H_1 : \frac{E[s_i s_j]}{\sigma_s^2} = 1 \end{cases} \quad (5)$$

where σ_{x_i} represents the standard deviation of x_i. The sample correlation $\hat{r}$ will vary from this expected value according to the distribution [13]

$$P(\hat{r}) = \frac{(M-2)\Gamma(M-1)(1-\rho^2)^{\frac{M-1}{2}}(1-\hat{r}^2)^{\frac{M-4}{2}}}{\sqrt{2\pi}\Gamma(M-\frac{1}{2})(1-\rho\hat{r})^{M-\frac{3}{2}}} \times {}_2F_1\left(\frac{1}{2},\frac{1}{2},\frac{2M-1}{2},\frac{\rho r+1}{2}\right) \quad (6)$$

where M is the number of samples and ${}_2F_1(a,b;c;z)$ is the Gaussian hypergeometric function.

An example of the distribution of the elements of correlation matrix A, for cases with (red curve) and without (blue curve) correlation are shown in Fig. 1. For large M and small ρ, cases typical in fMRI, we can approximate these distributions as normal.

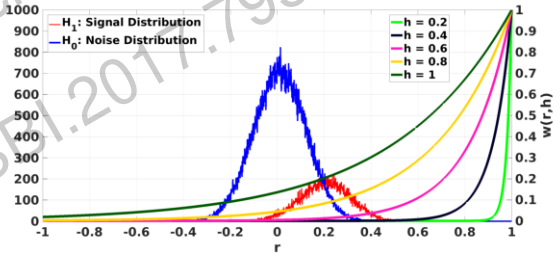

Fig. 1. The distribution of the elements of the correlation matrix A for zero true correlation (blue) and 0.2 correlation (red), with the kernel function in (2) evaluated for different values of the parameter h.

To optimally differentiate connections within and between networks, we select h to maximize the expected value of $W(i,j)$ for H_1 and minimize it for H_0. To account for different frequencies of occurrence of H_0 and H_1, we weight these expected values by their respective probabilities $P(H_0)$ and $P(H_1)$. To achieve this we solve the optimization problem

$$\begin{aligned} \hat{h} &= \underset{h}{\operatorname{argmax}}\, E[w(r,h)|H_1]P(H_1) - E[w(r,h)|H_0]P(H_0) \\ &= \underset{h}{\operatorname{argmax}} \int_{-1}^{1} w(r,h)P_{H_1}(r)dr\,P(H_1) - \int_{-1}^{1} w(r,h)P_{H_0}(r)dr\,P(H_0) \\ &= \underset{h}{\operatorname{argmax}} \int_{-1}^{1} w(r,h)(P(H_1)P_{H_1}(r) - P(H_0)P_{H_0}(r))dr \end{aligned} \quad (7)$$

where $w(r,h) = \exp(-2(1-r)/h^2)$. $P(H_1)$ and $P(H_0)$ are the relative frequencies of H_1 and H_0 with $P_{H_1}(r)$ and $P_{H_0}(r)$. For illustration, we have overlaid the function $w(r,h)$ for several values of h on the sample correlation distributions in Fig. 1. Note that to solve (7) we must first learn the pdfs. We do this by fitting a bimodal Gaussian mixture model (GMM) to the correlation data in A, rather than using the precise form in (6). We then use a directed line search to find the optimal h. This then determines the mapping $W = f(A)$ which is used in turn to generate the filtered data according to equation (4). We summarize our algorithm as follows:

Algorithm 1 Optimal Selection of h

1: Compute the correlation matrix $A = XX^\top$
2: Fit a bimodal GMM to elements of A using an EM algorithm to find $P_{H_1}(r)$, $P(H_1)$, $P_{H_0}(r)$, $P(H_0)$
3: Find the optimal $\hat{h}$ by solving equation (7) using $P_{H_1}(r)$, $P(H_1)$, $P_{H_0}(r)$, $P(H_0)$ obtained from the GMM

2.3. Performance Evaluation

2.3.1. Simulation

We simulated data matrix X with $K = 5$ networks and 100 vertices in each network, $T = 80$ time samples and SNR $= \sigma_s^2/\sigma_n^2 = 0.25$. These parameters were chosen so the distribution of the data in the correlation matrix A from the simulated data visually matched that for our fMRI data. The neighborhood $\mathcal{N}(i)$ was chosen to be the entire set of vertices. The optimal h was obtained by running Algorithm 1 for 1000 Monte Carlo trials. The value of the cost function in (7) was then evaluated as a function of h for each of the trials as shown in Fig. 2, which yielded an average optimal value of $h = 0.49$ ($\pm.03$ s.d.)

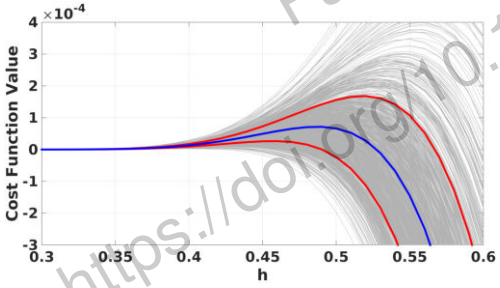

Fig. 2. Cost function in equation (7) evaluated as a function of h under the simulation settings. The light gray curves are 1000 individual Monte Corlo trials, the blue curve is the mean and the two red curves are one standard deviation away from the mean for each h.

To evaluate the impact of different values of h on identifying networks, we generated the filtered data Y according to (4), and used the correlation matrix $C = YY^\top$ to partition the data into 5 networks using a normalized cut algorithm [14]. We then computed a performance measure using the Adjusted Rand Index (ARI) [15] between the resulting network labels and the ground truth. Finally, we plotted the ARIs as a function of h in Fig. 3. The value that actually produced the highest ARI, averaged over 1000 trials, was $h = 0.52$ ($\pm.06$ s.d.), compared to the average optimal value $h = 0.49$ ($\pm.03$ s.d.) selected by our algorithm.

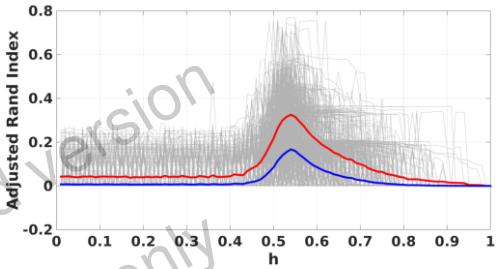

Fig. 3. ARI metric of the clustering results evaluated as a function of h under the simulation settings. The colors of curves have the same meaning as those in Fig. 2.

2.3.2. Task fMRI Data

Our earlier application of tNLM was to resting fMRI [12], here we focus on filtering of task fMRI. To evaluate performance we used a single subject minimally processed task fMRI data set from the Human Connectome Project (HCP) [16]. The minimal preprocessing pipeline for HCP task fMRI data is described in [16], [17]. We briefly summarized here: fMRI data were acquired with TR = 720ms, TE = 33.1ms, $2 \times 2 \times 2$ mm voxels with two independent sessions. Acquisition artifacts including head motion and spatial distortion were corrected and the data were co-registered with the T1 structural image and resampled onto the 32K-vertex cortical surface.

We applied tNLM to two data sets: the Language Processing (LP) task and the Social Cognition (SC) task [16], [18]. We evaluated performance as described below using the contrast of "Story" vs rest for LP and "Random" Interaction vs rest for SC because these two contrasts showed activated adjacent functional areas, allowing us to evaluate whether optimal-tNLM is able to preserve the separation of these areas after filtering.

Fig. 4. Cost function in equation (7) evaluated as a function of h for the Language Processing (LP) task (blue) and the Social Cognition (SC) task (red), respectively.

To evaluate the efficacy of our approach, the LP and SC

fMRI data were processed separately to find the optimal h for tNLM filtering and analyzed as described below. To achieve computational tractability, the cortical surface was downsampled to 11K vertices prior to filtering. We used the data from the right hemisphere only to obtain the optimal h based on Algorithm 1 for LP ($h = 0.43$) and SC ($h = 0.45$) separately. The value of the cost function in (7) was also evaluated as a function of h as shown in Fig. 4 for the two tasks.

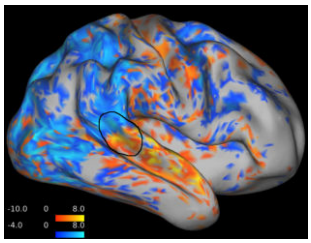

Fig. 5. Combined representative z-score maps ($\alpha = 0.05$) for "Story" contrast in LP (orange) and "Random" contrast in SC (blue). ROI for analysis containing activated regions in both contrasts is indicated in superior temporal gyrus.

Each data set was filtered using tNLM (with the neighborhood chosen as the 11-hop neighbors for each vertex i, as in [12]) for multiple different values of the parameter h. For comparison the data was also isotropically smoothed with FSL (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL) for multiple different values of the full-width-half-maximum (FWHM) parameter s of the Gaussian smoothing kernel. We then applied FSL's level 1 and level 2 analyses, which fits a GLM to the data [16]. As a result of this processing, we obtained a vertex-wise cortical map of z-scores for the "Story" and "Random" contrasts for multiple levels of Gaussian and tNLM smoothing. The z maps were thresholded for two different uncorrected α levels (0.02, 0.05). The uncorrected values were used as we were interested in exploring consistency of and differences in smoothing results across different α levels over all vertices.

Based on a visual comparison of results for the two tasks, we selected an ROI in the superior temporal gyrus of the right hemisphere, which contains adjacent regions of activation from the two tasks as shown in Fig. 5. Let R_s denote the region within the ROI exceeding the α-threshold for "Story" and R_r its counterpart for "Random". We computed the mean z-score over the entire ROI vs the Dice coefficient between R_s and R_r as a function of both h and s and for two different α values. The results are shown in Fig. 6. We can conclude from these figures that

1. For isotropic linear smoothing: with increased smoothing the average z-score (reflecting strength of response) increases along with the Dice coefficient. Thus, filtering will help detection of activation (increased z-score) but at the expense of blurring (increased Dice coefficient) between the two distinct but adjacent functional areas.

2. For tNLM smoothing, with increased levels of smoothing there is a range over which the average z-score increases without a corresponding increase in Dice coefficient. This indicates improved ability to detect activation, but now without blurring between functional regions. At a certain point, there is a knee in the curves at which point the Dice coefficient starts increasing while the mean z-score actually starts decreasing. The part of the curve above the knee indicates poor performance of tNLM because the value of h is too large causing blurring between the adjacent functional areas.

3. Interestingly, the knee in the tNLM curves was found to be 0.45 which matches well the optimal values found using equation (7) for LP ($h = 0.43$) and SC ($h = 0.45$). This represents the value which produces the maximum value of average z-score within the ROI without producing any significant increase in Dice coefficient or equivalently, blurring between functional areas.

4. We found this result consistent for multiple α levels ranging from 0.05 to 0.01. Space limitations only allow us to show two in Fig. 6

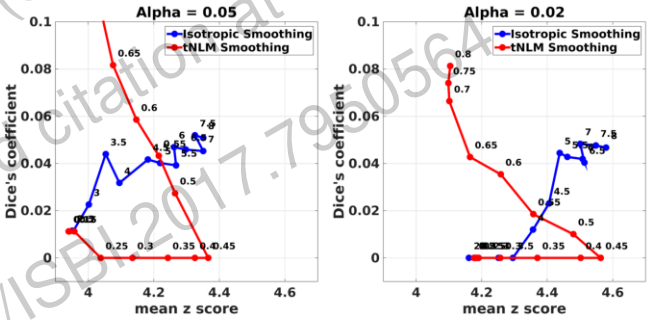

Fig. 6. Mean z score vs Dice's coefficient for different smoothing parameters and different α levels. The red curve represents the result under tNLM smoothing and the blue curve represents the result under isotropic smoothing. The parameter values for h and s are annotated above the curves.

3. DISCUSSION

The optimization-based method developed in this paper provides a means of systematically selecting the parameter for tNLM filtering when used as a preprocessing step for analyzing task-based or resting fMRI. The simulation produced a value of h close to that which optimized performance in terms of network identification when compared against ground truth. For experimental task fMRI, the optimal value coincided with the point at which we achieved maximum enhancement in SNR without blurring between distinct functional regions. In this work we assumed a specific form of the kernel. Further improvements may be gained by kernel-learning methods.

4. REFERENCES

[1] B Biswal, F Zerrin Yetkin, V. M Haughton, and J. S Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar mri," *Magnetic resonance in medicine*, vol. 34, no. 4, pp. 537–541, 1995.

[2] S. M Smith, P. T Fox, K. L Miller, D. C Glahn, P. M Fox, C. E Mackay, N Filippini, K. E Watkins, R Toro, A. R Laird, and C. F Beckmann, "Correspondence of the brain's functional architecture during activation and rest," *Proceedings of the National Academy of Sciences*, vol. 106, no. 31, pp. 13040–13045, 2009.

[3] R. L Buckner, P. A Bandettini, K. M OCraven, R. L Savoy, S. E Petersen, M. E Raichle, and B. R Rosen, "Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging," *Proceedings of the National Academy of Sciences*, vol. 93, no. 25, pp. 14878–14883, 1996.

[4] S. M Smith, C. F Beckmann, J Andersson, E. J Auerbach, J Bijsterbosch, G Douaud, E Duff, D. A Feinberg, L Griffanti, M. P Harms, et al., "Resting-state fmri in the human connectome project," *Neuroimage*, vol. 80, pp. 144–168, 2013.

[5] S Angenent, S Haker, A. R Tannenbaum, and R Kikinis, "On the laplace–beltrami operator and brain surface flattening," *Medical Imaging, IEEE Transactions on*, 1999.

[6] J. V Manjón, J Carbonell-Caballero, J. J Lull, G García-Martí, L Martí-Bonmatí, and M Robles, "Mri denoising using non-local means," *Medical image analysis*, vol. 12, no. 4, pp. 514–523, 2008.

[7] P Coupé, P Yger, S Prima, P Hellier, C Kervrann, and C Barillot, "An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images," *IEEE transactions on medical imaging*, vol. 27, no. 4, pp. 425–441, 2008.

[8] X.-N Zuo and X.-X Xing, "Effects of non-local diffusion on structural mri preprocessing and default network mapping: statistical comparisons with isotropic/anisotropic diffusion," *PLoS One*, vol. 6, no. 10, pp. e26703, 2011.

[9] M Bernier, M Chamberland, J.-C Houde, M Descoteaux, and K Whittingstall, "Using fmri non-local means denoising to uncover activation in sub-cortical structures at 1.5 t for guided hardi tractography," *Frontiers in human neuroscience*, vol. 8, pp. 715, 2014.

[10] N Wiest-Daesslé, S Prima, P Coupé, 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," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2008, pp. 171–179.

[11] A Buades, B Coll, and J.-M Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. IEEE, 2005, vol. 2, pp. 60–65.

[12] C Bhushan, M Chong, S Choi, A. A Joshi, J. P Haldar, H Damasio, and R. M Leahy, "Temporal non-local means filtering reveals real-time whole-brain cortical interactions in resting fmri," *PloS one*, vol. 11, no. 7, pp. e0158504, 2016.

[13] R. A Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507–521, 1915.

[14] J Shi and J Malik, "Normalized cuts and image segmentation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[15] W. M Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical association*, vol. 66, no. 336, pp. 846–850, 1971.

[16] H WU-Minn, "500 subjects+ meg2 data release: Reference manual," 2014.

[17] M. F Glasser, S. N Sotiropoulos, J. A Wilson, T. S Coalson, B Fischl, J. L Andersson, J Xu, S Jbabdi, M Webster, J. R Polimeni, et al., "The minimal preprocessing pipelines for the human connectome project," *Neuroimage*, vol. 80, pp. 105–124, 2013.

[18] D. M Barch, G. C Burgess, M. P Harms, S. E Petersen, B. L Schlaggar, M Corbetta, M. F Glasser, S Curtiss, S Dixit, C Feldt, et al., "Function in the human connectome: task-fmri and individual differences in behavior," *Neuroimage*, vol. 80, pp. 169–189, 2013.