Robust Anisotropic Diffusion to Produce Clear Statistical Parametric Map from Noisy fMRI

HAE YONG KIM[1], ZANG HEE CHO[2]

[1]USP - Escola Politécnica, Universidade de São Paulo,
Av. Prof. Luciano Gualberto, trav. 3, 158, CEP 05508-900, São Paulo, SP, Brasil.
hae@lps.usp.br
[2]UCI - Department of Radiological Sciences, University of California at Irvine,
Medical Sciences I, room B140, 92697-5000, Irvine, CA, USA.
zcho@uci.edu

Abstract. Functional magnetic resonance imaging (fMRI) uses MRI to noninvasively map areas of increased neuronal activity in human brain without the use of an exogenous contrast agent. Low signal-to-noise ratio of fMRI images makes it necessary to use sophisticated image processing techniques, such as statistical parametric map (SPM), to detect activated brain areas. This paper presents a new technique to obtain clear SPM from noisy fMRI data. It is based on the robust anisotropic diffusion. A direct application of the anisotropic diffusion to fMRI does not work, mainly due to the lack of sharp boundaries between activated and non-activated regions. To overcome this difficulty, we propose to calculate SPM from noisy fMRI, compute diffusion coefficients in the SPM space, and then perform the diffusion in fMRI images using the coefficients previously computed. These steps are iterated until the convergence. Experimental results using the new technique yielded surprisingly sharp and noiseless SPMs.

1. Introduction

The goal of functional neuroimaging is to map the activities of a living brain in space and time. The gold standard for measuring brain cell activities involves direct and invasive electrical recording of membrane potential of individual neurons. However, such measurements are limited to certain experimental conditions. For studies on human subjects, noninvasive methods, such as PET (positron emission tomography) or fMRI (functional magnetic resonance imaging), have to be applied.

From the early 80s PET dominated the field of functional neuroanatomy, but in the past ten years fMRI has developed into an alternative and powerful technique. Local increases in neural activity cause both a relative deoxygenation of blood and an increase in perfusion, that quickly reverses the deoxygenation, leading to an increase in oxygenation that endures for several seconds. fMRI is sensitive to the oxygenation of blood and has a spatio-temporal scale of about 1-3mm and one or more seconds. The lower limits on the effective resolution of fMRI are physiological and imposed by the spatio-temporal organization of evoked hemodynamic responses (2-5mm and 5-8 seconds).

By virtue of the relatively long half-life of the tracers used, PET can only measure responses summed over fairly long periods of time. In contrast, fMRI can be used in two ways: periodic (repeated stimuli) or event-related paradigms. In the periodic paradigm, the subject alternates between periods of stimulation and rest. In this case, each voxel of fMRI consists of a time-series that can be divided in epochs, an epoch being the period of time that corresponds to the activity followed by the rest. In the event-related paradigm, the subject executes the activity only during a short period of time, i.e., a trial.

Low signal-to-noise ratio in fMRI images compels us to use sophisticated image-processing techniques to detect activated brain areas. Raw fMRI images pass through many transformations until yielding the statistical parametric map (SPM). SPM is the spatial map of the statistical significance of an effect. To obtain SPM, the user specifies manually a design matrix and makes use of the general linear model (multiple linear regression) to compute the estimated parameters, i.e., to determine how well the time-series of each voxel fit the specified design matrix. These parameters are used to compute the statistical significance of an effect. These statistical tests, spatially disposed, form the SPM. In this paper, we will use indistinctly words “estimated parameters” and “SPM.” Course notes [SPM, 1997] is a good reference on fMRI processing. Even with all these image-processing apparatus, a noisy fMRI will yield a noisy SPM. Simple low-pass filters cannot be used because they blur the edges of activated areas. Traditional edge-preserving image filtering techniques cannot be used either because there is no clear edge between activated and non-activated areas.

In the literature, there are many works on attenuating noise and clustering activated regions in fMRI volumes

[Goutte et al., 1999; Ardekani and Kanno, 1998; Kershaw et al., 1999; Chuang et al., 1999; Friston et al., 1994]. In particular, Solé et al. [Solé et al., 2001] have recently proposed a technique named *anisotropic averaging*. This technique is inspired on the anisotropic diffusion, introduced by Perona and Malik [Perona and Malik, 1990]. Anisotropic averaging computes an initial set of clearly activated voxels by thresholding the correlation coefficient (estimated parameters). This set is then used to construct a complex "similarity measure" to compute the averaging coefficients. Despite the fact that Solé et al. try to explain the definition of their measure with intuitive arguments, we are impelled to ask if there are no simpler and more natural way to define the similarity measure. Solé et al. themselves state, "the key problem (...) becomes the design of the similarity measure." Moreover, their technique can be used only to process periodic fMRI tests, because their similarity measure is based on the discrete Fourier transformation of time-series of each voxel. In an event-related fMRI test, Fourier transformation does not make any sense.

In this paper, we present a new technique to obtain clear SPM from noisy fMRI data. Instead of defining a highly complex similarity measure based on the thresholded SPM, we use the gradient magnitude of SPM as the diffusion coefficients. We have also replaced the anisotropic averaging by the robust anisotropic diffusion [Black et al., 1998]. Our technique can be used to process both periodic and event-related fMRI tests. Simulated experiments using our technique produced surprisingly sharp and noiseless SPMs.

2. Robust Anisotropic Diffusion

Witkin introduced a clean formalism for the scale-space filtering [Witkin, 1983]. Let $I(x,y):\mathbb{R}^2 \to \mathbb{R}^+$ be a 2-D image in the continuous domain. The scale-space of this image is a 3-D image $I(x,y,t):\mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}^+$ that satisfies the following partial differential equation

$$\frac{\partial I(x,y,t)}{\partial t} = div(\nabla I),$$

using the original image $I(x,y,0)$ as the initial condition. Variable *t* is an artificial time parameter that specifies the image scale. Modifying the image according to this isotropic diffusion equation is equivalent to filtering the original image with a Gaussian filter, that is:

$$I(x,y,t) = G_t(x,y) * I(x,y),$$

where $G_t(x,y)$ is the Gaussian function with variance $\sigma^2 = 2t$. This linear scale-space has many nice mathematical properties. However, it blurs out image edges.

In order to keep sharp edges, while filtering noises and small details, Perona and Malik defined nonlinear anisotropic scale-space [Perona and Malik, 1990] by modifying the partial differential equation as follows

$$\frac{\partial I(x,y,t)}{\partial t} = div\left[g\left(\|\nabla I\|\right)\nabla I\right],$$

where $\|\nabla I\|$ is the gradient magnitude, and $g\left(\|\nabla I\|\right)$ is an "edge-stopping" function. They suggested two edge-stopping functions:

$$g\left(\|\nabla I\|\right) = \left[1 + \frac{\|\nabla I\|^2}{K^2}\right]^{-1} \text{ and}$$

$$g\left(\|\nabla I\|\right) = \exp\left[-\frac{\|\nabla I\|^2}{K^2}\right],$$

where *K* is a positive constant.

The choice of *g*(.) can greatly affect the extent to which discontinuities are preserved. So, recently, Black et al. [Black et al., 1998] proposed to use the robust estimation theory to define a better edge-stopping function (Tukey's biweight function):

$$g\left(\|\nabla I\|\right) = \begin{cases} \frac{1}{2}\left[1 - \left(\frac{\|\nabla I\|}{\sigma}\right)^2\right]^2, & \|\nabla I\| \le \sigma \\ 0, \quad \text{otherwise} \end{cases}$$

where $\sigma = K/\sqrt{2}$. Using Tukey's function, the diffusion process converges faster and yields sharper edges.

Perona and Malik discretized (spatio-temporally) their anisotropic diffusion equation as follows:

$$I(s,t+1) = I(s,t) + \frac{\lambda}{|\eta_s|}\sum_{p\in\eta_s} g\left(\nabla I_{s,p}\right)\nabla I_{s,p},$$

where $I(s,t)$ is a discretely sampled image, *s* denotes the pixel position in a discrete 2-D or 3-D grid, and *t* now denotes discrete time steps (iterations, $t \ge 0$). The constant $\lambda \in \mathbb{R}^+$ is a scalar that determines the rate of diffusion, η_s represents the spatial neighborhood of pixel *s*. $|\eta_s|$ is usually four for 2-D images and six for 3-D images, except at image boundaries. Perona and Malik approximated the image gradient magnitude in a particular direction as

$$\nabla I_{s,p} = I(p,t) - I(s,t), \quad p \in \eta_s.$$

The same discretization scheme can be used with Tukey's robust estimation function, yielding robust anisotropic diffusion.

3. General Linear Model

The general linear model is simply an equation that relates what one observes, to what one expected to see, by expressing the observations as a linear combination of expected components and some residual error. The general linear model comes in a number of guises, for example multiple linear regression, analysis of covariance or a simple *t* test. The general linear model can be written as:

$$Y = X\beta + \varepsilon,$$

where Y is the column vector of observations, ε is the column vector of error terms, and β is the column vector of parameters. X is the design matrix with one row per observation and one column per model parameter.

The following equation performs the least square estimation of parameters:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Different statistical tests (such as "are there any effects of interest?") can be computed using the estimated vector $\hat{\beta}$. SPM is the image of statistical tests. As we have said above, we will use indistinctly the words "image of estimated parameters" and "SPM."

Let us give a simple numerical example of parameter estimation in fMRI:

$$\begin{bmatrix} 50 \\ 51 \\ 60 \\ 62 \\ 51 \\ 52 \\ 62 \\ 63 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{bmatrix}$$

The first matrix Y is a time-series of values of a voxel. Let us suppose that values were taken in every 5 seconds. The second matrix X is the design matrix. The first column indicates that we would like to detect a periodic activation that repeats every 20 seconds and lasts 10 seconds. The second column is dummy. Estimating parameters, we obtain $\hat{\beta}_1 = 10.75$ and $\hat{\beta}_2 = 51.00$. High value of $\hat{\beta}_1$ indicates that very likely the voxel is activated.

4. Anisotropic Averaging

The anisotropic diffusion has been successfully applied to structural MRI [Gerig et al., 1992]. However, this technique cannot be applied directly to fMRI, mainly due to the lack of sharp boundaries between activated and non-activated regions. It also cannot be directly applied to SPM, because it is usually very noisy.

Smoothing fMRI (convolving the data with a smoothing kernel) generally increases the signal-to-noise ratio. However, it blurs the edges between activated and non-activated regions. Thus, it is desirable to perform only intra-region smoothing (avoiding inter-region smoothing).

Solé et al. [Solé et al., 2001] proposed this idea and called it anisotropic averaging, a technique motivated in part by the anisotropic diffusion. It consists on making a selective neighborhood averaging of the signal. Let I be an fMRI image, $I(p,n)$ denote the voxel value at spatial voxel position p and volume $n \in [1...N]$, and $I(p)$ denote the time-series signal at voxel p. Then, the selective neighborhood averaging consists on:

$$I(s,n,t+1) = \frac{1}{\sum_{p \in \eta_s} w(s,p)} \sum_{p \in \eta_s} w(s,p) I(p,n,t),$$

for $n \in [1...N]$ and $t \geq 0$.

The similarity measure Ψ determines the weights $w(s,p)$ using the time-series signal being averaged $I(s)$ and its corresponding neighbor $I(p)$:

$$w(s,p) = \Psi(I(s), I(p)).$$

This similarity measure allows us to distinguish activated voxels from non-activated ones and is able to perform a selective averaging, combining only signals of the same class.

Solé et al. propose to compute the initial set Ω of clearly activated voxels by thresholding correlation coefficients (estimated parameters $\hat{\beta}$). Then, the Fourier spectra of voxels in Ω are computed to define the similarity measure function Ψ. The Fourier spectrum of time-series of each voxel is also computed in order to evaluate the similarity between voxels. The whole procedure is highly complex and the readers are referred to [Solé et al., 2001] for further details. We would like to ask, "are all these calculations really necessary?" Moreover, "how can this technique be applied to event-related fMRI acquisition protocols?"

5. Proposed Method

We propose a different approach, directly motivated by the robust anisotropic diffusion, to filter fMRI. Our method is simpler, can be applied to event-related fMRI and has yielded surprisingly clear SPMs.

Given an fMRI image $I(s,n,0)$, estimate parameters

$\hat{\beta}(0)$ at iteration t=0 using the general linear model. The magnitude of gradient of SPM will be used as the argument of edge-stopping function g to calculate the diffusion coefficients. These coefficients are used to perform diffusion in fMRI data $I(s,n,0)$. Then, the diffused fMRI $I(s,n,1)$ is used to estimate new parameters $\hat{\beta}(1)$ at iteration t=1. These steps are repeated until the average of diffused value is below some threshold. It is also possible to specify the desired number of iterations, instead of defining the thresholding average diffused value. The following equation describes this process:

$$I(s,n,t+1) = I(s,n,t) + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g\left(\nabla \beta_{s,p}\right) \nabla I_{s,p} ,$$

for all $n \in [1...N]$ and for $t \geq 0$.

The best edge-stopping function g is the Tukey's biweight function, defined in section 2. Note that the correct choice of parameter σ of Tukey's function is essential to yield good results. The magnitude of this parameter depends on the average value of the estimated parameters $\hat{\beta}$.

6. Experimental Results

We have tested the method proposed in this paper on both simulated and real fMRI data. However, the experimental results on real fMRI images are inconclusive, because the actual activated region is unknown. Consequently, we will depict in this paper only simulated data with well-delimited activated areas. We aim to investigate further the experimental results on real fMRI images in a future work.

Figure 1 depicts part of a simple simulated fMRI phantom with 10×10×3 voxels per volume and 64 volumes. All voxel values are originally 500. Gaussian noise with standard deviation 10 was added to the original values. Volumes 3, 4, 7, 8, 11, 12, ... have a 6×6 square activated area in the center of the volume, with two non-activated holes with 4 voxels each. Activated voxels had their values increased by 20. Figure 2 depicts the image of estimated parameters using the general linear model. SPM obtained without filtering is presented in the left column. SPM obtained by filtering fMRI with the proposed method (using Tukey's edge-stopping function with $\sigma = 10$) is depicted in the right column. Note that the filtered SPM is completely noiseless, and the edges are perfectly preserved.

Figure 3 depicts real fMRI images with 79×95×68 voxels per volume and 12 volumes. The volumes were realigned in order to correct subject's head-movements. Then, the volumes were randomly shuffled in order to remove any activation signal that may be present in the original fMRI. The average value of voxels is roughly 450 and the average value eliminating the background is approximately 900. In volumes 3, 4, 7, 8, 11 and 12, an ellipsoidal area was artificially activated by adding 150 and Gaussian noise with standard deviation 10 to the original values. The left column of figure 3 depicts three slices of volume 1, with no activated voxels. The right column of figure 3 depicts three slices of volume 3, with artificially activated ellipsoid. Note that the activated area is hardly visible. Figure 4 depicts SPM obtained without filtering (left column) and the corresponding threshold images (right column). These images are very noisy and many non-activated areas were falsely detected as activated (and vice versa). Figure 5 depicts SPM obtained by filtering fMRI with the proposed method (left column) and corresponding threshold images (right column). Great part of noises was removed and the threshold images are perfect: there is neither non-activated area falsely detected as activated, nor activated voxel falsely detected as non-activated.

7. Conclusion

In this paper, we have presented a new technique to obtain clear SPM from noisy fMRI. It is directly inspired by the robust anisotropic diffusion. Experimental results on the simulated data show that surprisingly sharp and noiseless SPMs can be obtained.

8. Acknowledgements

One of the authors, H. Y. Kim, would like to express his gratitude to FAPESP and CNPq for the partial financial supports of this work under grants 2001/02400-9 and 300689/98-5, respectively.

9. References

[Ardekani and Kanno, 1998] B. A. Ardekani and I. Kanno, "Statistical Methods for Detecting Activated Regions in Functional MRI of the Brain," Magn. Reson. Imag., vol. 16, no. 10, pp. 1217-1225, 1998.

[Black et al., 1998] M. J. Black, G. Sapiro, D. H. Marimont and D. Hegger, "Robust Anisotropic Diffusion," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 421-432, Mar. 1998.

[Chuang et al., 1999] K. H. Chuang, M. J. Chiu, C. C. Lin, and J. H. Chen, "Model-Free Functional MRI Analysis Using Kohonen Clustering Neural Network and Fuzzy c-Means," *IEEE Trans. Med. Imag.*, vol. 18, pp. 1117-1128, Dec. 1999.

[Friston et al., 1994] K. J. Friston, P. Jezzard, and R. Turner, "The Analysis of Functional MRI Time-

Series," *Human Brain Mapping*, vol. 1, pp. 153-171, 1994.

[Gerig et al., 1992] G. Gerig, O. Kubler, R. Kikinis, and F. A Jolesz, "Nonlinear Anisotropic Filtering of MRI Data," *IEEE Trans. Med Imag.*, vol. 11, pp. 221-232, June 1992.

[Goutte et al., 1999] C. Goutte, P. Toft, E. Rostrup, F. A. Nielsen, and L. K. Hansen, "On Clustering fMRI Time Series," *NeuroImage*, vol. 9, no. 3, pp. 298-310, 1999.

[Kershaw et al., 1999] J. Kershaw, B. A. Ardekani, and I. Kanno, "Application of Bayesian Inference to fMRI Data Analysis," *IEEE Trans. Med. Imag*., vol. 18, pp. 1138-1153, Dec. 1999.

[Perona and Malik,1990] P. Perona and J. Malik, "Scale-Space and Edge Detection Using Anisotropic Diffusion," *IEEE. Trans. Patt. Anal. and Machine Intell.*, vol. 12, no. 7, pp 629-639, 1990.

[Solé et al., 2001] A. F. Solé, S. C. Ngan, G. Sapiro, X. P. Hu and A. López, "Anisotropic 2-D and 3-D Averaging of fMRI Signals," *IEEE Trans. Medical Imaging*, vol. 20, no. 2, pp. 86-93, Feb. 2001.

[SPM, 1997] K. J. Friston (ed.), "SPM Course – Short Course Notes," http://www.fil.ion.ucl.ac.uk/spm/course/notes97/, 1997.

[Witkin, 1983] A. P. Witkin, "Scale-Space Filtering," *Proc. 8th Int. Joint Conf. Art. Intelligence*, vol. 2, pp. 1019-1022, 1983.

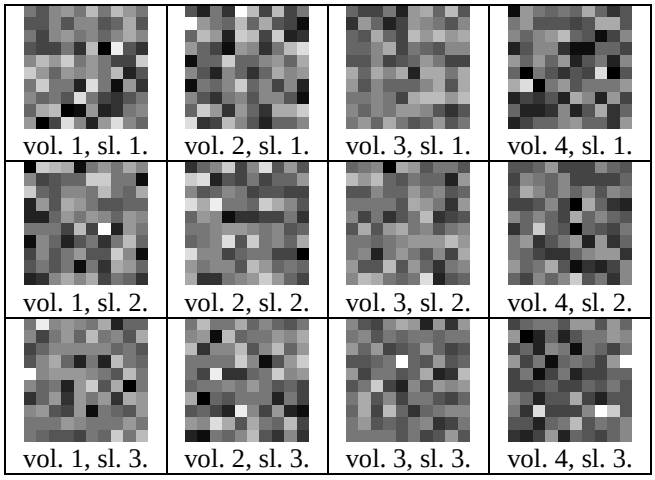

Fig. 1: A simple simulated fMRI phantom with 10×10×3 voxels per volume and 64 volumes. Only the first 4 volumes are depicted. All voxel values are 500 added with Gaussian noise with standard deviation 10. Volumes 3, 4, 7, 8, 11, 12, ... have some activated voxels where value 20 was added to the original value.

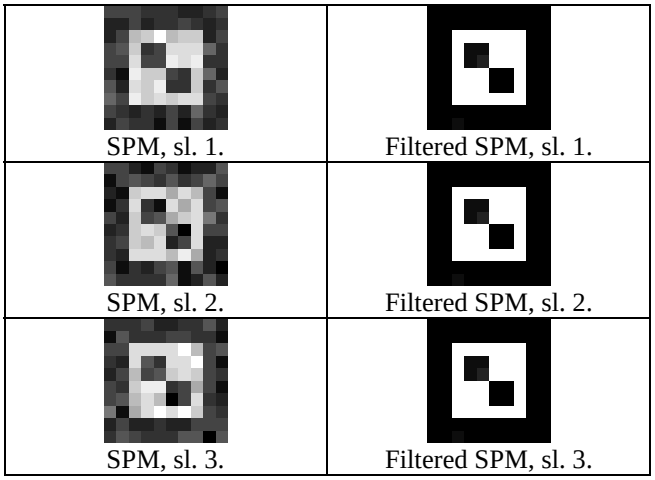

Fig. 2: Spatial map of estimated parameters $\hat{\beta}$ from fMRI depicted in figure 1. Left column: original parameters. Right column: parameters obtained filtering phantom with the proposed technique.

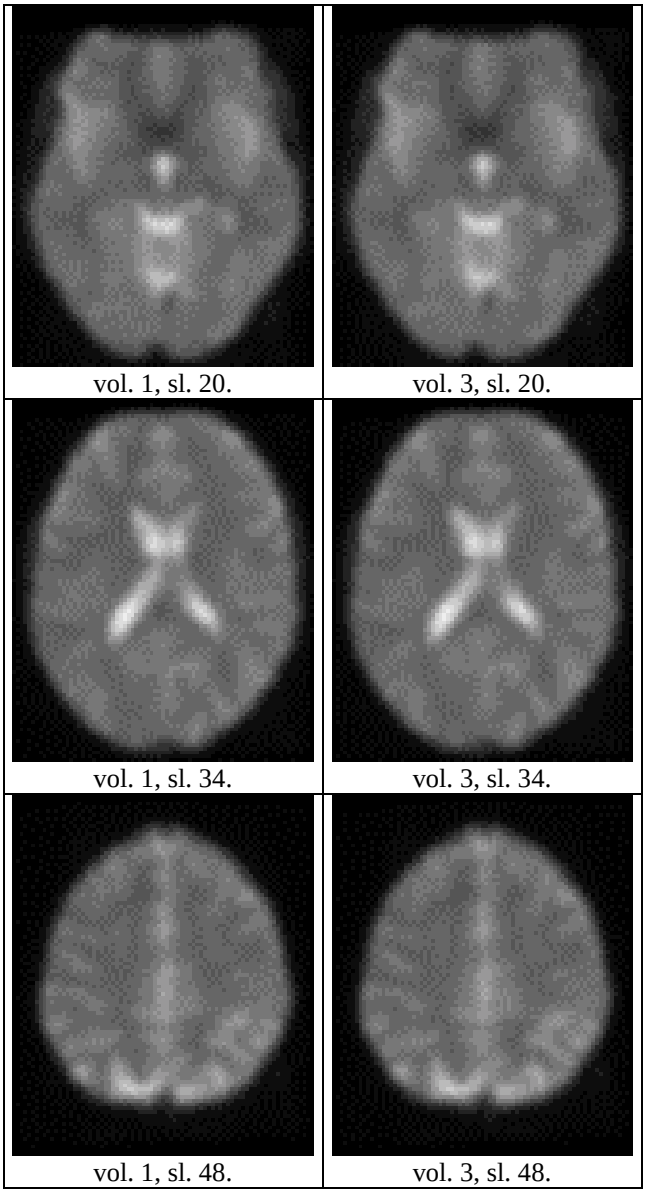

Fig. 3: Real fMRI images (79×95×68 voxels per volume, 12 volumes) with artificially activated area. The average value of voxels is roughly 450 and the average value eliminating the background is approximately 900. In volumes 3, 4, 7, 8, 11 and 12, an ellipsoidal area was artificially activated by adding 150 and Gaussian noise with standard deviation 10 to the original values.

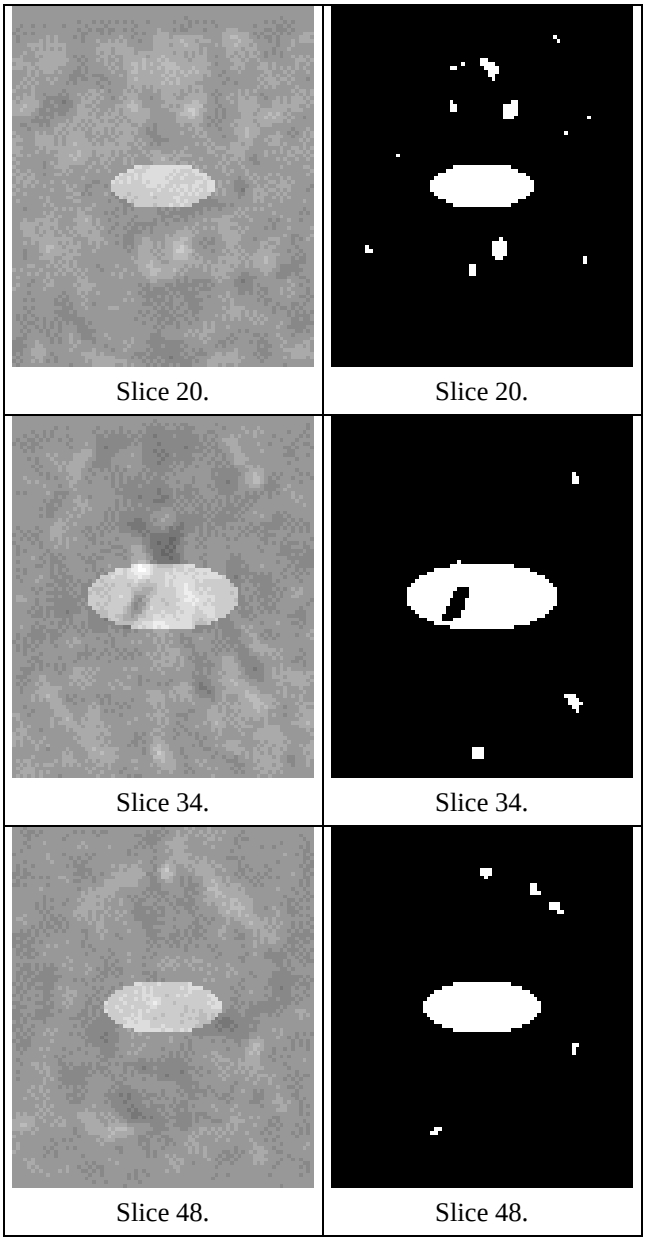

Fig. 4: Spatial map of estimated parameters $\hat{\beta}$ from fMRI depicted in figure 3. Left column: estimated parameters. Right column: activated area detected by thresholding the estimated parameters.

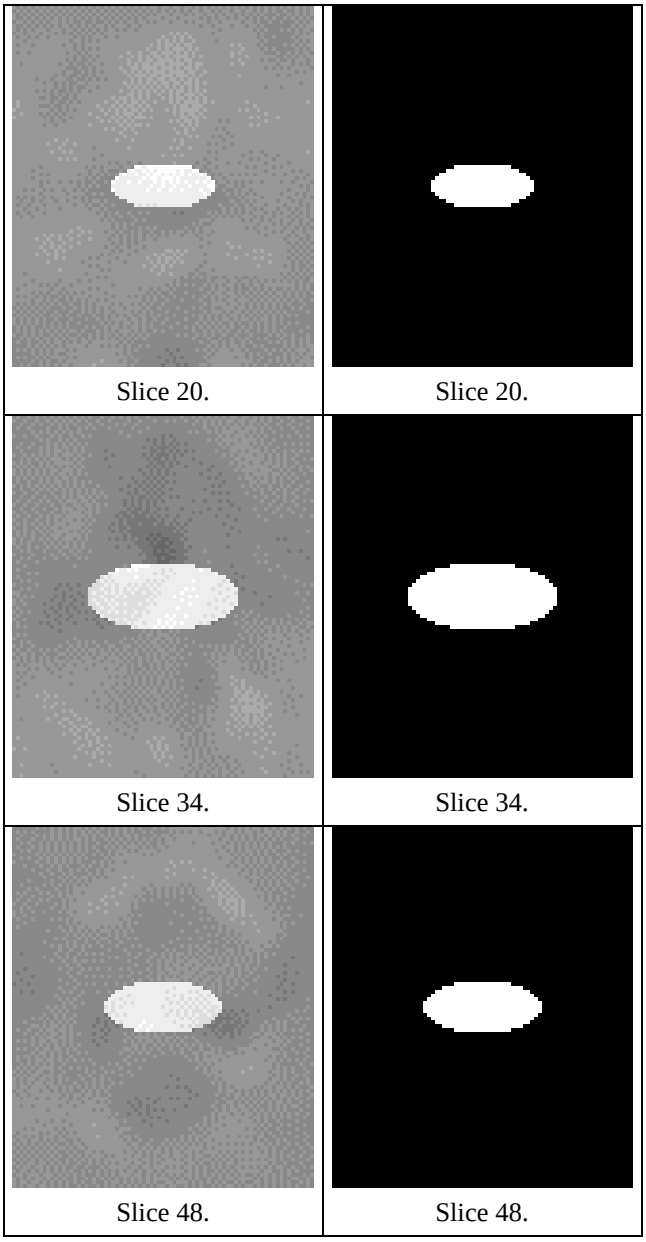

Fig. 5: Spatial map of estimated parameters $\hat{\beta}$ obtained using the proposed method. Left column: estimated parameters. Right column: activated area detected by thresholding the estimated parameters.