# Euler's Elastica Regularization for Voxel Selection of fMRI Data

Chuncheng Zhang
Institute of Automation, Chinese Academy of Sciences, Beijing, China
Email: chuncheng.zhang@ia.ac.cn

Zhiying Long
State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China
Email: friskying@163.com

*Abstract*—**Multivariate analysis methods have been widely applied to functional Magnetic Resonance Imaging (fMRI) data to reveal brain activity patterns and decode brain states. Among the various multivariate analysis methods, the multivariate regression models that take high-dimensional fMRI data as inputs while using relevant regularization were proposed for voxel selection or decoding. Although some previous studies added the sparse regularization to the multivariate regression model to select relevant voxels, the selected sparse voxels cannot be used to map brain activity of each task. Compared to the sparse regularization, the Euler's Elastica (EE) regularization that considers the spatial information of data can identify the clustered voxels of fMRI data. Our previous study added EE Regularization to Logical Regression (EELR) and demonstrated its advantages over the other regularizations in fMRI-based decoding. In this study, we further developed a multivariate regression model using EE in 3D space as constraint for voxel selection. We performed experimental tests on both simulated data and real fMRI data to investigate the feasibility and robustness of EE regression model. The performance of EE regression was compared with the Generalized Linear Model (GLM) and Total Variation (TV) regression in brain activity detection, and was compared with GLM, Laplacian Smoothed L0 norm (LSL0) and TV regression methods in feature selection for brain state decoding. The results indicated that EE regression possessed better sensitivity to detect brain regions specific to a task than did GLM and better spatial detection power than TV regression. Moreover, EE regression outperformed GLM, LSL0 and TV in feature selection.**

*Index Terms*—**fMRI, decoding, feature selection, Euler's elastica, multivariate regression**

## I. Introduction

Functional Magnetic Resonance Imaging (fMRI) is a noninvasive imaging technology that can reveal the neural mechanism underlying various cognitive processes by measuring the Blood-Oxygen-Level Dependent (BOLD) signal in the brain [1]-[3]. For fMRI data, it is very important to identify the voxels (features) that are relevant to each task because there are many irrelevant voxels that do not respond to the task. Therefore, voxel/feature selection is critical to both the estimation of activity patterns and the brain state decoding.

Conventionally, the univariate statistical methods have been widely applied to fMRI data to reveal the activated regions and select the useful voxels for brain state decoding. Among various univariate methods, the Generalized Linear Model (GLM) on a voxel-by-voxel basis is the most popular for estimating activity patterns and selecting the task-relevant voxels [4]-[7]. All univariate methods for voxel selection treat each voxel independently and does not consider the interaction among voxels.

In contrast to the univariate methods, multivariate methods are gaining increasing attention in fMRI data analysis. Some multivariate data-driven methods, such as independent component analysis, principal component analysis and partial least square, have been widely applied to identify task-related voxels or networks from fMRI data [8], [9]. Moreover, recursive feature elimination [10] and searchlight [11] algorithms have been applied to voxel selection of fMRI data.

Recently, multivariate regression models that take high-dimensional fMRI data as inputs have been used for voxel selection or decoding. Some regularizations should be added to the multivariate regression models to stabilize the estimation of regression coefficients and prevent overfitting, An iterative sparse representation-based method [12] and a SCoRS (survival count on random subsamples) method that used the L1 norm as the regulation [13] were proposed for voxel selection of fMRI data. In our previous study, we proposed a fast feature selection method that is based on a smoothed L0-norm using a Laplacian kernel (LSL0) [14]. Moreover, sparse logistic regression using a combination of L1 and L2 norm regularization was proposed to automatically select relevant voxels for the classification of fMRI data [15]. However, the above regulations did not consider the spatial structure of the image.

The spatial structure of the fMRI data is useful to the estimation of the multivariate regression model because

Manuscript received January 14, 2020; revised May 12, 2020.

©2020 Int. J. Sig. Process. Syst.
doi: 10.18178/ijsps.8.2.32-41

the voxels that respond to the external stimuli are generally distributed in clusters [16]-[18]. Michel et al. proposed an approach for regularized prediction based on Total Variation (TV) [19] that is mathematically defined as the norm of the image gradient regularization and demonstrated that TV regularization is well suited to the purpose of brain mapping while being a powerful tool for brain decoding. In our previous study, we added Euler's Elastica (EE) that is based on both the gradient and the curvature of images to Logical Regression (LR) model for fMRI-based decoding and demonstrated that EE regularized LR (EELR) showed better decoding performance than LR with TV regularization and LR with sparse regularization [20]. Although our previous study demonstrated that sparse regularization can be used to select sparse relevant voxels [14], the brain activity maps that are relevant to each task cannot be obtained by the selected sparse voxels. Because EE regularization considered the spatial information of data and showed better performance than TV regularization in fMRI-based decoding [20], it is essential to further investigate how to apply EE regularization to select relevant voxels and obtain brain activity mapping of a task.

In this study, we developed a multivariate regression model using EE in 3D space as constraint for voxel identification. To stabilize the estimation of weights, EE constraints on the weights were added to the multivariate regression model to obtain the solution with EE energy minimization. We investigated the performance of EE regression method from two aspects. First, the estimated weights were used to identify the brain regions that were activated by each task. The performance of EE regression was compared with GLM using both the simulated and real fMRI data. Second, we selected the voxels with high weights as the important features and inputted the features to the support vector machine (SVM) to decode the brain states. The decoding accuracies of EE regression were compared with the other feature selection methods that include TV regression, LSL0, and GLM using both simulated and real fMRI data. The results demonstrated that EE regression is more sensitive to detect the regions that are specific to a task than GLM and exhibited better classification accuracy than the other feature selection methods in most cases.

## II. Theory

We proposed EE regularized multinomial logistic regression (EELR) algorithm in 3D voxel space for multi-class classification in our pervious study [20]. Because the theory of EE regularization was described in our previous study, we only introduced EE regularized 3D linear regression (EE regression) for voxel selection of fMRI data.

The 3D linear regression model in voxel space can be expressed in (1):

$$y = x \vdots w \tag{1}$$

In this model, $x \in \mathrm{R}^{\mathrm{T}\times \mathrm{V}_1\times \mathrm{V}_2\times \mathrm{V}_3}$ is 4D observed fMRI data, $\mathrm{y} = \{y_1, y_2, \dots, y_T\} \in R^T$ is the reference function of a task, and $w \in \mathrm{R}^{\mathrm{V}_1\times \mathrm{V}_2\times \mathrm{V}_3}$ is a 3D weight matrix. $V_1, V_2$ *and* $V_3$ are the number of voxels in the three spatial dimensions respectively. The operation $\vdots$ is defined as a kind of sum of production between $x$ and $w$ satisfying (2).

$$\mathrm{y}^{\mathrm{t}} = \sum_{\mathrm{i}=1}^{\mathrm{V}_1}\sum_{\mathrm{j}=1}^{\mathrm{V}_2}\sum_{\mathrm{k}=1}^{\mathrm{V}_3} \mathrm{x}_{\mathrm{ijk}}^{\mathrm{t}} * \mathrm{w}_{\mathrm{ijk}} \tag{2}$$

where $t \in [1, \dots, T]$ represents time point.

We performed numerical discretization in 3D space using the method in [20] for EE regression model. The task of EE regression method is to find a solution $w \in R^{V_1\times V_2\times V_3}$ for (1) that best fits the regression model and has EE energy minimization. In the discrete space, the solution of EE regression can be obtained by solving the optimization problem in (3) [21], [22].

$$\min \int_\Omega \left(\mathrm{a} + \mathrm{b}\left|\nabla \cdot \frac{\nabla \mathrm{u}}{|\nabla \mathrm{u}|}\right|^2\right)|\nabla \mathrm{u}| + \frac{\eta}{2}\|\mathrm{y} - \mathrm{x} \vdots \mathrm{u}\|^2 \tag{3}$$

where $\Omega$ represents the discrete 3D voxel space, $u$ represents the 3D weight matrix, and $\nabla \cdot \frac{\nabla u}{|\nabla u|}$ represents the curvature, the parameters $a, b$ are related to the elastic energy and the ratio between them controls the connectivity and smoothness of the level lines, and $\eta$ is the penalty of the loss function of the regression model [22].

The minimization of (3) is equivalent to the minimization problem (4) [20, 23].

$$\min_{\mathrm{v,u,m,p,n}} \int_\Omega (\mathrm{a} + \mathrm{b}|\nabla \cdot \mathrm{n}|^2)|\mathrm{p}| + \frac{\eta}{2}\|\mathrm{y} - \mathrm{x} \vdots \mathrm{v}\|^2 \tag{4}$$

where four variables $v$, $\boldsymbol{p}$, $\boldsymbol{m}$, and $\boldsymbol{n}$ are introduced and satisfied (5).

$$\mathrm{v} = \mathrm{u}, \mathrm{p} = \nabla \mathrm{u}, \mathrm{n} = \mathrm{m}, |\mathrm{p}| = \mathrm{m} \cdot \mathrm{p} \tag{5}$$

According to our previous study [20], the constrained minimization problem in (5) can be transformed into minimizing the augmented Lagrangian functional that is defined in (6).

$$\begin{aligned}\mathcal{L}(v, w, \boldsymbol{m}, \boldsymbol{p}, \boldsymbol{n}; \lambda_1, \boldsymbol{\lambda_2}, \lambda_3, \boldsymbol{\lambda_4}) = \int_\Omega (a + b(\nabla \cdot \boldsymbol{n})^2)|\boldsymbol{p}| + \frac{\eta}{2}\|y - x \vdots v\|^2 + r_1 \int_\Omega (|\boldsymbol{p}| - \boldsymbol{m} \cdot \boldsymbol{p}) + \int_\Omega \lambda_1(|\boldsymbol{p}| - \boldsymbol{m} \cdot \boldsymbol{p}) + \frac{r_2}{2}\int_\Omega |\boldsymbol{p} - \nabla w|^2 + \int_\Omega \boldsymbol{\lambda_2} \cdot (\boldsymbol{p} - \nabla w) + \frac{r_3}{2}\int_\Omega (v - w)^2 + \int_\Omega \lambda_3(v - w) + \frac{r_4}{2}\int_\Omega |\boldsymbol{n} - \boldsymbol{m}|^2 + \int_\Omega \boldsymbol{\lambda_4} \cdot (\boldsymbol{n} - \boldsymbol{m}) + \delta_{\mathcal{R}}(\boldsymbol{m})\end{aligned} \tag{6}$$

where $\lambda_1, \boldsymbol{\lambda_2}, \lambda_3$ *and* $\boldsymbol{\lambda_4}$ are Lagrange multipliers, and $r_1, r_2, r_3$ *and* $r_4$ are positive penalty parameters associated with Lagrange multipliers. The parameter $r_2$ controls the amount of diffusion of $w$. The parameters $r_3$ *and* $r_4$ control the closeness between $v, w$ and $\boldsymbol{n}, \boldsymbol{m}$, respectively.

The problem of minimizing (6) can be transformed into the following five sub-problems (7-11) [20].

$$\mathcal{E}_1(\mathrm{v}) = \frac{\eta}{2}\int_\Gamma \|\mathrm{y} - \mathrm{x} \vdots \mathrm{v}\|^2 + \int_\Omega \frac{\mathrm{r}_3}{2}(\mathrm{v} - \tilde{\mathrm{w}}^{\mathrm{k}-1})^2 + \lambda_3 \mathrm{v} \tag{7}$$

$$\begin{aligned}\mathcal{E}_2(\mathrm{w}) = \int_\Omega \frac{\mathrm{r}_2}{2}\left|\tilde{\mathrm{p}}^{\mathrm{k}-1} - \nabla \mathrm{w}\right|^2 - \lambda_2 \cdot \nabla \mathrm{w} + \frac{\mathrm{r}_3}{2}\left(\tilde{\mathrm{v}}^{\mathrm{k}} - \mathrm{w}\right)^2 + \lambda_3(-\mathrm{w})\end{aligned} \tag{8}$$

$$\mathcal{E}_3(\mathrm{m}) = \delta_{\mathcal{R}}(\mathrm{m}) + \frac{\mathrm{r}_4}{2}\int_\Omega \left|\tilde{\mathrm{n}}^{\mathrm{k}-1} - \mathrm{m}\right|^2 - \lambda_4 \cdot \mathrm{m} - (\mathrm{r}_1 + \lambda_1)\mathrm{m} \cdot \tilde{\mathrm{p}}^{\mathrm{k}-1} \quad (9)$$

$$\mathcal{E}_4(\boldsymbol{p}) = \int_\Omega \left(a + b\left(\nabla \cdot \tilde{\boldsymbol{n}}^{\boldsymbol{k-1}}\right)^2\right)|\boldsymbol{p}| + (r_1 + \lambda_1)(|\boldsymbol{p}| - \tilde{\boldsymbol{m}}^k \cdot \boldsymbol{p}) + \frac{r_2}{2}|\boldsymbol{p} - \nabla\tilde{w}^k|^2 + \boldsymbol{\lambda_2} \cdot \boldsymbol{p} \quad (10)$$

$$\mathcal{E}_5(\mathrm{n}) = \int_\Omega \mathrm{b}(\nabla \cdot \mathrm{n})^2 \left|\tilde{\mathrm{p}}^{\mathrm{k}}\right| + \frac{\mathrm{r}_4}{2}\left|\mathrm{n} - \tilde{\mathrm{m}}^{\mathrm{k}}\right|^2 + \lambda_4 \cdot \mathrm{n} \quad (11)$$

The updates of Lagrange multipliers and the minimization of sub-problems (8-11) are performed in the same way as our previous study [20]. Because the minimization of sub-problem (7) of EE regression is different from that of EELR in our previous study [20], we present the solution of sub-problem (7) in the Appendix materials.

## III. Materials and Methods

In this section, two simulated experiments and a real fMRI experiment were conducted. The first simulated experiment was used to determine the optimal parameters of EE and TV algorithms and investigate the accuracy of weight estimation using both EE and TV models. The second simulated experiment is a human fMRI resting data-based simulation to investigate the feasibility of EE regression method in both the brain activity detection and the feature selection for decoding. Moreover, the real fMRI experiment was performed to verify the effectiveness and reliability of applying EE regression to identify relevant voxels from fMRI data. The performance of EE regression was compared with TV regression and GLM in the detection of brain activity, and compared with TV regression, LSL0 and GLM methods in the feature selection for decoding.

Although there are seven parameters in EE regression algorithm, we set the parameters $a, b, r_1, r_2, r_4$ as $\{a = 1, b = 20, r_1 = 1, r_2 = 200, r_4 = 300\}$ according to our previous study [20]. Moreover, we determined the optimal value of the parameters $\eta$ and $r_3$ ($\eta = 100, r_3 = 60$) by using the simulated data because the sub-problem $\mathcal{E}_1(v)$ in equation (7) of this study is different from our previous study [20]. Meanwhile, the penalty weight λ of TV regression algorithm was optimized using the same simulated data as EE regression and was set to $\lambda = 0.05$. The same LSL0 algorithm as our previous study [14] was used in this study. All the codes of the LSL0 and EE regression algorithms were performed on MATLAB 8.3.0.532 (R2014a). EE regression code was written based on the code of Tai's study [23]. TV code was written based on the code of an open-source package (http://nilearn.github.io/decoding/space_net.html).

### A. Simulated 3D Experiment

This simulated 3D experiment was conducted to determine the optimal parameters and explore the accuracy of weight estimation for EE and TV regression model. Moreover, the performance of the estimation accuracy of EE regression was compared with that of TV regression.

#### 1) Generation of simulated 3D data

The simulated data in this experiment was generated in the same way as Michel's study [19]. Each dataset was generated according to equation (12).

$$y^t = \sum_{i=1}^{V_1}\sum_{j=1}^{V_2}\sum_{k=1}^{V_3} x_{ijk}^t * w_{ijk} + e^t \quad (12)$$

where $t$ refers the $t^{th}$ time point. The $x_{ijk}^t$ refers observed signal at $t^{th}$ time point at the voxel $(i, j, k)$. The signal of each voxel was assumed as following Gaussian distribution, that is $x_{ijk}^t \sim N(0, 1)$. The $w_{ijk}$ represents the weights of the $(i, j, k)$ voxel. The $y^t$ refers the simulated target for the $t^{th}$ time point. The error term $e^t \in N(0, r)$ is Gaussian noise whose standard deviation $r$ was varied to generate data with different signal-to-noise ratios (SNR) levels. The SNR value is calculated use (13).

$$\mathrm{SNR} \equiv -20\log_{10}\mathrm{r}^2 \quad (13)$$

The simulated datasets with four SNRs levels (1, 2, 5, 10dB) were generated in the simulation. Each SNR level contained 100 datasets and each dataset contained 100 time points. The spatial size of each dataset is 12×13×14 voxels. Two square Regions of Interest (ROI) (size 5×5×5) were generated to simulate 3D weight vector. The weight of each voxel in ROIs was set according to the equation (14).

$$w_{ijk} = b * exp\left(\frac{(i-c_i)^2+(j-c_j)^2+(k-c_k)^2}{3}\right) \quad (14)$$

where the point $(c_i, c_j, c_k)$ represents the center of ROI, and $b \in [-1, 1]$ represents the sign of the weights. The weights of the voxels outside ROIs were set to zero. The spatial distribution of the weights is shown in Fig. 1. For each dataset, the weights of 50% voxels that were randomly selected from each ROI were set to zero in order to simulate inter-dataset variability.

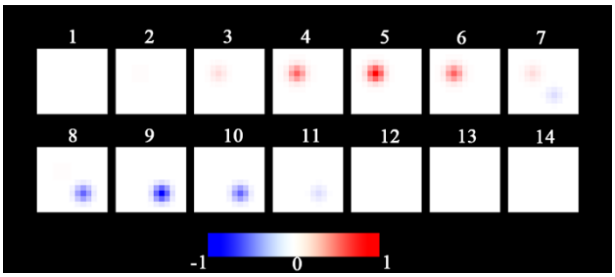


Figure 1. Two-dimensional slices of the source of three-dimensional volume of simulated 3D experiment.

#### 2) Parameter optimization

The simulated datasets with SNR equal to 5dB were used to determinate optimal values of the parameter $\eta$ and $r_3$ for EE and the parameter $\lambda$ for TV. We let the parameter $\eta$ vary in the range $\{1,10,100,200,500\}$, $r_3$ vary in the range $\{40,50,60,70,80\}$, and $\lambda$ vary in the range $\{0.001,0.01,0.05,0.1,0.5\}$. As a result, there are 25 parameter combinations of $\eta$ and $r_3$ for EE regression. For each dataset with SNR equal to 5dB, EE regression with different parameter combinations ($\eta$ and $r_3$) and TV regression with different $\lambda$ were applied separately to

estimate the weight of the regression model. Meanwhile, the error between the true and the estimated weight was calculated by using (15).

$$error = \frac{\|\hat{w}-w\|_2^2}{\|w\|_2^2} \quad (15)$$

where $\hat{w}$ is the estimated weight and $w$ is the true weight. The mean errors of all subjects' datasets were calculated for each parameter combinations ($\eta$ and $r_3$) of EE model and each $\lambda$ of TV model. The value that yielded the lowest error was selected as the optimal value for $\eta$, $r_3$ and $\lambda$. The optimal values of $\eta$, $r_3$ and $\lambda$ were used in the following simulated and real fMRI experiments.

*3) Weight estimation*

After the optimal values of the parameters were determined, EE regression and TV regression using the optimal parameters were applied to the simulated datasets at the four SNR levels (1, 2, 5 and 10 dB) to estimate the regression weights. The datasets that were used in weight estimation are independent with the datasets that were used in parameter optimization. For each SNR level, the mean error of each method across 100 datasets was calculated.

### *B. Human fMRI Resting Data-Based Simulation*

This simulation was designed to examine the performance of EE regression in brain activity estimation and feature selection. The activity patterns that were estimated by EE regression were compared with TV regression and GLM. Moreover, the classification accuracies of the SVM using the feature selection methods of EE regression, TV regression, LSL0 and GLM were compared.

The simulated datasets used in this simulation were the same as those in our previous study [20]. We only described some major points of the simulated data and did not present the detailed data generation process in this study.

*1) Preprocessing*

The imaging parameters of the resting fMRI data and the participant information were presented in the previous study [20]. All the participants gave written constant form. The experiment was approved by the Institutional Review Board (IRB) of the State Key Laboratory of Cognitive Neuroscience and Learning in Beijing Normal University.

The resting fMRI data of 12 participants underwent the motion correction and spatial normalization in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/). The normalized fMRI data were resliced into 3×3×4 $mm^3$ voxels.

*2) Data generation*

The simulated datasets were generated based on the pre-processed resting fMRI datasets of 12 subjects. For each subject, three types of simulated datasets that included two-task dataset, four-task dataset and simulated dataset with corrupted regions were generated. The two-task datasets and four-task dataset were used to compare the activation detection power of EE regression, TV regression and GLM as well as the decoding performance of SVM using EE regression, TV regression, LSL0 and GLM as the feature selection methods. Moreover, the datasets with corrupted regions was used to demonstrate the advantages of EE regression over TV regression.

Each simulated dataset of each subject included two runs. For the two-task simulated datasets, the two tasks occurred in an ABBABAAB sequence in each run. The simulated datasets with corrupted regions included two tasks in each run. The ROIs that were activated by each task are shown in Fig. 3 (a-b). We suppose that the red ROIs were corrupted by larger noises than the yellow ROIs. For simplicity, the inter-subject variations and signal percent changes were not considered in the simulation. The signal percent change of the yellow ROIs was set to 2% while that of the red ROIs was set to 1.2% for all the subjects.

*3) Brain activity analysis*

For each subject, the linear drift was removed from each run using the spm_detrend function in SPM8 [24]. The time series of each voxel was normalized to zero mean and unit variance.

EE regression, TV regression and GLM were applied to all the preprocessed two-task datasets of 12 subjects. For EE/TV regression, the reference function of each task that was derived from the task paradigm and the HRF was used as the vector y in equation (7). After the regression weight vector $w$ of each task was estimated by EE/TV regression, the weight vector was transformed into Z-scores.

After weight estimation, one-sample T-test was applied to the weight vectors of all the subjects to test the voxels that significantly responded to each task. The voxel-wise threshold was set as $p \leq 0.001$. All statistical results were corrected by a topological false discover rate (FDR) method on peaks at an overall (corrected) alpha level of 0.01.

### *C. Real fMRI Experiment*

*1) Subjects*

The real fMRI data used in this study were the same as our previous study [25]. For readability, only some main points were described here. The detailed imaging parameters can be found in the previous study [25].

Twenty-seven right-handed college participants including fifteen females and twelve males (aged 22.67±2.96 years) from Beijing Normal University took part in the experiment. All participants gave written consent according to the guidelines that were set by the MRI center of Beijing Normal University. The experiment was approved by the Institutional Review Board (IRB) of the State Key Laboratory of Cognitive Neuroscience and Learning in Beijing Normal University.

*2) Experiment design*

Forty shapes with three disparity levels (+30 arcmin, -30 arcmin and 0 arcmin) were used in the experiment. The whole experiment consisted of 120 stimuli with three different disparity levels. The zero disparity level (0 arcmin) corresponds to zero disparity (ZD), the positive disparity level (30 arcmin) corresponds to uncrossed disparity (UD) and the negative disparity level (-30 arcmin) corresponds to crossed disparity (CD).

The experiment used block design and included two runs. Each run contained three tasks that were viewing visual stimuli with CD, UD, and ZD levels. For each run, each task corresponded to four task blocks and twelve 24-s task blocks alternated with twelve 12-s resting blocks. Each task block contained twelve stimuli with a specific disparity level. Each stimulus was presented 1.5 s and the inter-stimuli interval was 0.5 s. When the two continuous stimuli were different, participants were required to press the button with their right index finger in each task block. Otherwise, they were required to press the button with their left index finger. Participants were required to fixate a cross at the center of the screen in the resting blocks,

*3) Preprocessing*

The fMRI data were preprocessed using SPM8. The first three volumes of each run were removed from the data to remove the instability of the initial scanning. The functional images of each participant were first realigned and then spatially normalized into the standard MNI template. The voxel size of the normalized images was set to 3×3×4 $mm^3$.

*4) Data processing*

The linear drift was removed from each run of each participant using the spm_detrend function in SPM8 [24]. The time course of each voxel was normalized to zero mean and unit variance.

*a) Brain activity analysis*

For each subject, the first run was used for brain activity analysis. EE regression and TV regression were applied in the same way as the simulated experiment to estimate the regression weight of each task. For GLM analysis, the four training runs of each subject were modeled in one regression model that included the three regressors (three tasks) and one constant term. For each method, one-sample T-test was applied to the weight of each task for a group of subjects to detect the brain regions that were significantly activated by each task after weight estimation. All statistical results were corrected by a topological false discover rate (FDR) method on peaks at an overall (corrected) alpha level of 0.01.

*b) Feature selection and classification*

Two-fold cross-validation was performed to decode the brain states of the three tasks datasets. During the first/second fold, we used the first run as the training/testing data and the second run as the testing/training data. Four feature selection methods that included EE regression, TV regression, LSL0 and GLM were applied to the training data of each participant separately to identify the useful features for classification.

EE regression, TV regression, LSL0 and GLM were applied to the training data to select the voxels relevant to each task in the same way as the simulated experiment. For the three/two-class classification, the union set of the selected voxels relevant to each of the four/three/two tasks constituted the final features of the training data.

For each classification, the SVM classifier was trained using the selected features of the training data and was applied to each scan of the testing run to judge the types of visual stimuli viewed by each participant. The testing data used the same features as the training data. For each selection method, the mean accuracy of each classification across the twenty-seven participants was calculated. Moreover, the mean accuracy across the two folds was further calculated for each classification.

## IV. Results

### A. Simulated 3D Experiment

*1) Parameter optimization*

Fig. 2 (a, b) shows the mean estimation error of EE regression and TV regression using different parameters. It shows that the mean error is minimum when the parameter$\eta$, and $r_3$were 100 and 60 respectively for EE regression and the parameter $\lambda$ was 0.05 for TV regression. The simulated and real fMRI experiments in this study used the optimal values of parameters $\eta$, $r_3$ and $\lambda$.

*2) Weight estimation*

The mean estimation error of EE and TV regressions at all the SNR levels are presented in Fig. 2 (c). It can be seen that EE regression yielded significantly lower estimation errors than TV regression at all SNRs.

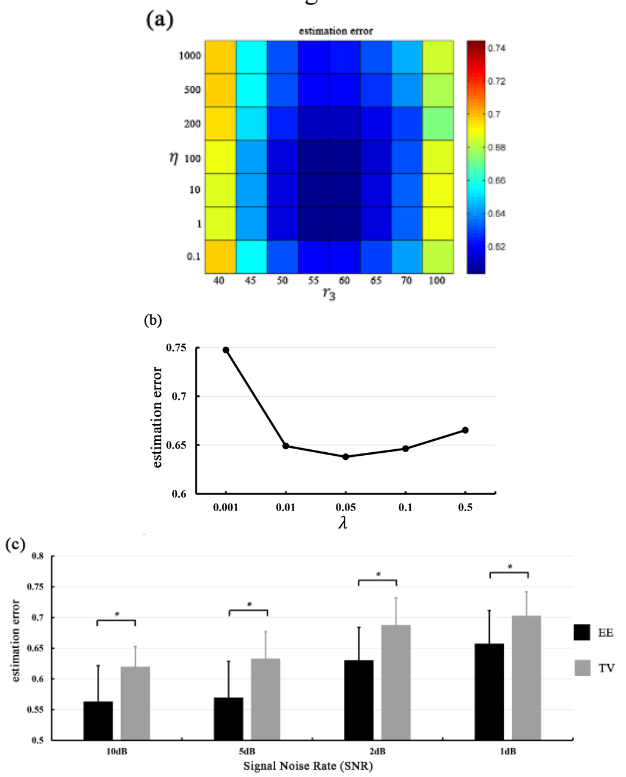


Figure 2. Results of Simulated 3D experiment. (a) The variation of estimation error with different parameters for EE regression. (b) The variation of estimation error with different parameters for TV regression. (c) The variation of estimation error with SNRs for EE and TV regression.

### B. Human fMRI Resting Data-Based Simulation

Fig. 3 (c-f) shows the group-level activation patterns that were estimated by EE regression and TV regression for the simulated datasets with corrupted regions. It can be seen that EE regression detected both yellow and red regions in Fig. 3 (a-b) while TV regression only detected

the yellow regions in Fig. 3 (a-b) and failed to detect the red regions that were corrupted with high noises.

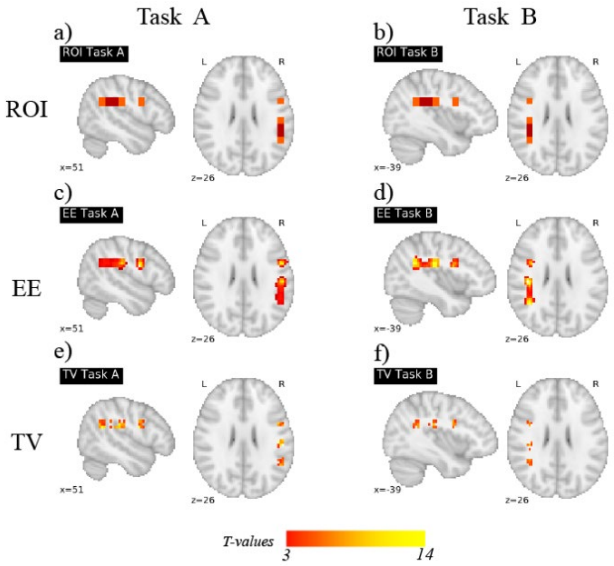


Figure 3. (a-b) The ROIs of the simulated data with corrupted regions of the human fMRI resting data-based simulation. (c-f) The group activity pattern for the corrupted regions simulation. Group activity pattern of EE regression (e-d) and TV regression (e-f).

### C. Real fMRI Experiment

#### 1) Brain activity analysis

Fig. 4 displays the group activation pattern of each task for EE regression, TV regression and GLM methods. The local maxima coordinates of the activation of EE and TV regression and GLM are reported in Table I. The activated areas in retinotopic areas V1, V2d, V3d, V3A, V7, V2v, V3v, V4v and LOC were identified using Caret atlas (http://brainvis.wustl.edu/wiki/index.php/Caret:Download ).

For the CD task, EE regression mainly detected activity in the bilateral V2v, V3A, V3d, V7 and LOC. TV regression mainly detected activity in the bilateral V4v, V7 and LOC. GLM method mainly detected activity in bilateral V1, V2v, V3v, V4v, V7 and LOC.

For the UD task, EE regression mainly detected activity in the bilateral V1, V3v, V3A, V7 and LOC. TV regression mainly detected activity in the bilateral V1, V4v and LOC. GLM method mainly detected activity in bilateral V1, V2v, V3v, V4v and LOC.

For the ZD task, EE regression mainly detected activity in the bilateral V1, V2v, V3v, V4v, TV regression mainly detected activity in bilateral V4v and LOC. The GLM method mainly detected activity in bilateral V1, V2v, V3v and LOC.

It can be seen that the activated regions of UD/CD showed little overlap with those of ZD for both EE regression and TV regression and large overlap for GLM method. Moreover, EE regression detected larger activated regions than TV regression.

#### 2) Feature selection and classification

The accuracy of the two-class and three-class classifiers using the four voxel selection methods are shown in Fig. 5. It can be seen that EE regression showed the highest accuracy among the four feature selection methods in all cases. EE regression showed significantly higher accuracy than TV regression for CD vs. UD and CD vs. UD vs. ZD. EE regression showed significantly higher accuracy than LSL0 in most cases, except in the case of CD vs. UD. Moreover, EE regression showed significantly higher accuracy than GLM in most cases, except in the case of CD vs. UD.

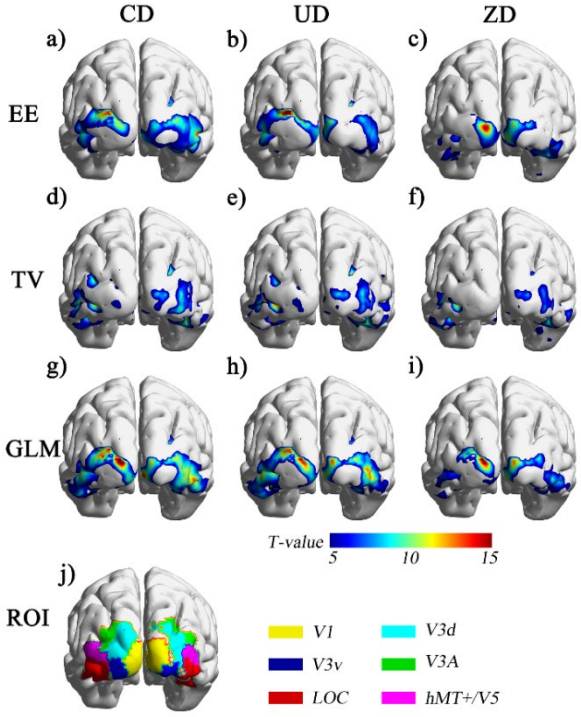


Figure 4. (a-i) The group activity pattern of stereoscopic fMRI experimental data. (a–c) Activated regions of each task for EE regression. (d–f) Activated regions of each task for TV regression. (g-i) Activated regions of each task for GLM. (j) The activated areas in retinotopic areas V1, V2d, V3d, V3A, V7, V2v, V3v, V4v and LOC of one subject.

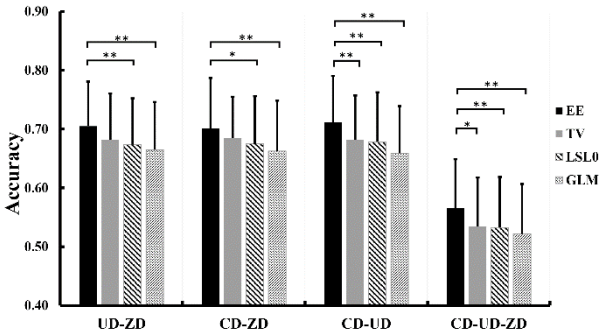


Figure 5. The classification accuracies of stereoscopic fMRI experiment. The classification accuracies using EE regression, TV regression, LSL0 and GLM feature selection methods. *p<0.1, **p<0.05.

## V. Discussion

In this study, we proposed EE regression method that incorporates EE constraints in 3D space into the multivariate regression model for fMRI data analysis.

TABLE I. ACTIVITY FOCI (MNI COORDINATES) OF EACH DISPARITY LEVEL DETECTED BY EE AND TV REGRESSION AND GLM

| Method | | Region | | BA | Peak coordinate | | | t-value |
|---|---|---|---|---|---|---|---|---|
| | | | | | x | y | z | |
| *EE* | CD | Fusiform gyrus | L | 37 | -21 | -79 | -6 | 7.39 |
| | | Superior occipital gyrus | L | 19 | -21 | -94 | 30 | 10.20 |
| | | Middle temporal gyrus | L | 39 | -45 | -70 | 10 | 7.39 |
| | | Occipital gyrus | R | 19 | 9 | -85 | 6 | 6.78 |
| | | Superior occipital gyrus | R | 19 | 27 | -88 | 30 | 10.26 |
| | | Middle temporal gyrus | R | 39 | 51 | -58 | 6 | 9.61 |
| | UD | Occipital gyrus | L | 19 | -9 | -97 | 22 | 8.20 |
| | | Superior occipital gyrus | L | 19 | -24 | -91 | 30 | 8.81 |
| | | Middle temporal gyrus | L | 37 | -48 | -76 | 18 | 6.98 |
| | | Superior occipital gyrus | R | 19 | 18 | -94 | 30 | 10.12 |
| | | Middle temporal gyrus | R | 37 | 48 | -61 | 6 | 5.93 |
| | ZD | Superior occipital gyrus | L | 19 | -15 | -91 | 6 | 7.76 |
| | | Occipital gyrus | R | 18 | 9 | -91 | 10 | 7.25 |
| *TV* | CD | Middle occipital gyrus | L | 37 | -30 | -79 | 10 | 8.27 |
| | | Middle occipital gyrus | R | 37 | 27 | -79 | 6 | 10.17 |
| | UD | Middle occipital gyrus | L | 37 | -33 | -79 | 18 | 9.13 |
| | | Superior occipital gyrus | R | 19 | 36 | -85 | 22 | 10.01 |
| | ZD | Middle temporal gyrus | L | 37 | -42 | -58 | 10 | 8.67 |
| | | Middle occipital gyrus | L | 19 | -33 | -76 | 6 | 7.82 |
| | | Middle temporal gyrus | R | 37 | 42 | -49 | 14 | 7.90 |
| | | Middle occipital gyrus | R | 19 | 27 | -79 | 6 | 9.45 |
| *GLM* | CD | Superior occipital gyrus | L | 18 | -15 | -100 | 18 | 10.08 |
| | | Middle occipital gyrus | L | 19 | -39 | -82 | 18 | 8.26 |
| | | Occipital gyrus | R | 17 | 24 | -79 | 2 | 11.11 |
| | | Middle temporal gyrus | R | 37 | 51 | -67 | 2 | 9.63 |
| | UD | Occipital gyrus | L | 18 | -12 | -100 | 18 | 10.61 |
| | | Middle occipital gyrus | L | 19 | -24 | -79 | 2 | 8.47 |
| | | Middle temporal gyrus | L | 37 | -51 | -64 | 6 | 8.10 |
| | | Occipital gyrus | R | 19 | 15 | -100 | 22 | 9.10 |
| | | Middle occipital gyrus | R | 19 | 33 | -88 | 30 | 8.72 |
| | | Middle temporal gyrus | R | 37 | 48 | -67 | 2 | 7.98 |
| | ZD | Superior occipital gyrus | L | 18 | -12 | -91 | 6 | 7.97 |
| | | Middle occipital gyrus | L | 19 | -24 | -82 | 2 | 8.60 |
| | | Occipital gyrus | R | 18 | 24 | -79 | 2 | 9.70 |
| | | Middle temporal gyrus | R | 37 | 42 | -67 | 2 | 7.21 |

The robustness and feasibility of EE regression in detecting brain activity pattern and selecting feature for decoding was investigated. The results from the simulated data and real fMRI data demonstrated that EE regression tended to detect regions that are selective to each task and detect larger regions than TV regression. Moreover, classification based on the set of features selected by EE regression presented higher accuracy than TV regression, LSL0 and GLM in most cases.

In the simulated 3D experiment, EE regression produced significantly lower estimation error of the weight than TV regression at all the SNR levels, which suggest that EE regression had better robustness to noises than TV regression (see Fig. 2 (c)). Fig. 3 (c-f) shows the group activation patterns estimated by EE regression and TV regression for the simulated datasets with corrupted regions. EE regression detects all the predefined regions while TV regression failed to detect the red regions with large noise in Fig. 3 (a-b). The virtue of EE model in (6) is that the regularization using EE energy [26], [27] penalizes the integral of the square of the curvature along edges. Consequently, the model can reconnect contours along large distances while recovering the curvature of objects. Moreover, image-inpainting researches also demonstrated reconnection of noise broken parts by EE model [28], [29]. As a result, under the role of EE constraint, EE regression model tended to recover all the activated voxels within a cluster although the noises corrupted some activated voxels. Our results further demonstrated that EE regularization had better robustness to noises than TV regularization, which is consistent with our previous study [20].

For the real fMRI experiment, the activity patterns of the three disparity levels detected by GLM showed much larger overlap than those estimated by EE regression and TV regression (see Fig. 4). Moreover, EE regression detected activity in V3A, V3d, V3v, V7 and LOC for CD and UD instead of ZD, and TV regression detected activity in V4v and LOC for CD and UD instead of ZD. It was reported that V3A, V3d, V3v, V7 and LOC were relevant to the disparity processing, especially V3A and V7 [30]-[33]. In contrast to GLM, EE regression detected more regions relevant to the disparity processing for CD/UD. Moreover, EE regression detected stronger activation than TV regression for UD, CD and ZD. The results further suggested that EE regression tended to detect the regions that were relevant to each task and EE regression had stronger detection power than TV regression.

In terms of brain state decoding, the classification accuracy of EE regression was higher than TV regression. LSL0 and GLM in most cases (see Fig. 5). The results may suggest that the features that were selected by EE regression contained more discriminative information than the other feature selection methods. Because EE regression, TV regression and LSL0 mainly selected the voxels that are relevant to each task rather than the voxels jointly participating in all tasks, EE regression, TV regression and LSL0 showed better decoding performance than GLM. However, there were large

differences between EE constraint and L0 constraint. LSL0 removes the correlated voxels and selects an extremely small number of discriminatory voxels with a sparse pattern. In contrast, EE regression keeps all the correlated and relevant voxels and revealed distributed clusters that were engaged in each task. Our results suggest that additional correlated voxels also carried the discriminative information and are helpful to the classification, which is consistent with some previous studies [15]. Moreover, EE regression considers the voxels of the whole brain in the analysis without any prior feature selection, whereas LSL0 requires the initial feature selection steps. In contrast to TV regression, EE regression considers more spatial information (the curvature of images) as regulations, which contributes to the better decoding performance of EE regression.

EE regression algorithm includes three parameters ($\eta, a\ and\ b$) that are relevant to EE regression model and the four penalty parameters ($r_1, r_2, r_3\ and\ r_4$) that are relevant to EE energy minimization function. Our previous demonstrated that the parameters $a, b, r_1, r_2, r_4$ can be set to the fixed value ($a = 1, b = 20, r_1 = 1, r_2 = 200, r_4 = 300$) [34]. Because the sub-problem $\mathcal{E}_1$(v) of this study is different from our previous study [20], we reset the optimal value of the parameters $\eta$ and $r_3$. In this study, the optimal values of the parameters $\eta$ and $r_3$ were determined by the simulated experiments. Both our simulated and real fMRI experiments demonstrated that the optimal parameters ($\eta = 100, r_3 = 60$) worked well in the fMRI data analysis.

## VI. Conclusions

In this study, we demonstrated the feasibility and robustness of the multivariate regression using EE in 3D voxel space as constraint for fMRI data analysis. The performance of EE regression in brain activity analysis and feature selection were investigated using both simulated and real fMRI experiments. The results demonstrated that EE regression showed the better detection power than TV regression. Therefore, EE regression has the potential to become a powerful method to reveal the neural mechanism underlying various cognitive processes by selecting relevant voxels.

## Appendix

### A. Sub-problem of Minimizing $\mathcal{E}_1(v)$

The minimization problem $\mathcal{E}_1(v)$ can be solved as regularized least squares problem using iteration method.

$$\mathcal{E}_1(v) = \frac{\eta}{2}\int_{\Gamma} \|y - x : v\|^2 + \int_{\Omega} \frac{r_3}{2}(v - \widetilde{w}^{k-1})^2 + \lambda_3 v \quad (16)$$

We rewrite $\mathcal{E}_1(v)$ in a new 2-D space as following:

$$\mathcal{E}_1(v_1) = \frac{\eta}{2}\int_{\Gamma} \|A \cdot v_1 - y\|^2 + \int_{\Omega'} \frac{r_3}{2}(v_1 - \widetilde{w_1}^{k-1})^2 + \lambda_3 v_1 \quad (17)$$

In the new space, the 3-D spatial space is mapped into 1-D voxel space and the temporal space is left unchanged. Specifically, $\Omega' \in R^1$ refers the 1-D voxel space, $A \in R^{T\times(V_1\times V_2\times V_3)}$ refers fMRI data, $v_1, w_1 \in R^{(V_1\times V_2\times V_3)\times 1}$ refers the mapped $v$ and $u$.

We initial the iteration by assuming an ideal condition of $A^T A = I$.

Based on $\widetilde{w_1}^{k-1}$, the minimizer of $\mathcal{E}_2$(w) at $(k-1)^{th}$ iteration, we get minimization square solution of $A \cdot w_1^{(0)} = $ y.

$$w_1^{(0)} = \widetilde{w_1}^{k-1} + A^T(AA^T)^{-1}\left(y - A \cdot \widetilde{w_1}^{k-1}\right) \quad (18)$$

Replacing $y$ with $A \cdot u_1^{(0)}$, the temporal integral becomes

$$\int_{\Gamma} \|A \cdot v_1 - y\|^2 = \left(v_1 - w_1^{(0)}\right)^T A^T A\left(v_1 - w_1^{(0)}\right) \quad (19)$$

It should be noticed that in the voxel space

$$\int_{\Omega'}\left(v_1 - w_1^{(0)}\right)^2 = \left(v_1 - w_1^{(0)}\right)^T I\left(v_1 - w_1^{(0)}\right) \quad (20)$$

Using (3-5), the approximation of $\mathcal{E}_1(v_1)$ can be formulated as

$$\mathcal{E}_1'(v_1) = \frac{\eta}{2}\int_{\Omega'}\left(v_1 - w_1^{(0)}\right)^2 + \int_{\Omega'}\frac{r_3}{2}\left(v_1 - \widetilde{w_1}^{k-1}\right)^2 + \lambda_3 v_1 \quad (21)$$

The minimizer of $\mathcal{E}_1'(v_1)$ is

$$\widetilde{v_1}^{(0)} = \frac{\eta w_1^{(0)} + r_3 u_1}{\eta + r_3} \quad (22)$$

where $u_1 = \widetilde{w_1}^{k-1} - \frac{\lambda_3}{r_3}$.

If the ideal condition of $A^T A = I$ is satisfied, we get the solution $\tilde{v}^k$ by mapping $\widetilde{v_1}^{(0)}$ back into the 3-D spatial space. If not, the gradient descent method is applied to get the final solution.

Inviting $\widetilde{v_1} = \widetilde{v_1}^{(0)}$ back to (2), we get (8).

$$\mathcal{E}_1(\widetilde{v_1}) = \frac{\eta}{2}\int_{\Gamma} \|A \cdot \widetilde{v_1} - y\|^2 + \int_{\Omega'}\frac{r_3}{2}(\widetilde{v_1} - u_1)^2 \quad (23)$$

We apply gradient descent method to find minimizer of $\mathcal{E}_1(\widetilde{v_1})$.

For $j = 1, \ldots, J$, we apply (9) and (10)

$$v_1^j = v_1^{j-1} - 2^{-j}\left(A^T A v_1^{j-1} - 2A^T y + v_1^{j-1} - 2u_1\right) \quad (24)$$

$$v_1^j = v_1^j + A^T(AA^T)^{-1}\left(y - A \cdot v_1^j\right) \quad (25)$$

where the initial is set as $v_1^0 = \widetilde{v_1}^{(0)}$, the minimizer is $\widetilde{v_1} = v_1^J$.

Finally, on the $k^{th}$ minimizing step, we get the solution $\tilde{v}^k$ by mapping $\widetilde{v_1}$ back into the 3-D spatial space.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Zhiying Long conducted the research; Chuncheng Zhang analyzed the data; Zhiying Long and Chuncheng Zhang wrote the paper.

## ACKNOWLEDGMENT

This work is supported by the Key Program of National Natural Science Foundation of China (61731003), China Postdoctoral Science Foundation (2019M650893), the National Natural Science Foundation of China (61671067 and 61473044), and the Interdiscipline Research Funds of Beijing Normal University and the Fundamental Research Funds for the Central Universities (2017XTCX04).

## REFERENCES

[1] E. Zarahn, G. K. Aguirre, and M. D'Esposito, "Empirical analyses of BOLD fMRI statistics," *NeuroImage*, vol. 5, no. 3, pp. 179-197, 1997.
[2] G. Aguirre, E. Zarahn, and M. D'esposito, "The variability of human, BOLD hemodynamic responses," *NeuroImage*, vol. 8, no. 4, pp. 360-369, 1998.
[3] Y. Chai, *et al.*, "Visual temporal frequency preference shows a distinct cortical architecture using fMRI," *NeuroImage*, vol. 197, pp. 13-23, 2019.
[4] K. J. Friston, *et al.*, "Analysis of fMRI time-series revisited," *Neuroimage*, vol. 2, no. 1, pp. 45-53, 1995.
[5] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of functional MRI time‐series," *Human Brain Mapping*, vol. 1, no. 2, pp. 153-171, 1994.
[6] K. J. Friston, A. P. Holmes, K. J. Worsley, J. Poline, C. D. Frith, and R. S. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, vol. 2, no. 4, pp. 189-210, 1994.
[7] J. Soch, A. P. Meyer, J. D. Haynes, and C. Allefeld, "How to improve parameter estimates in GLM-based fMRI data analysis: Cross-validated bayesian model averaging," *NeuroImage*, vol. 158, pp. 186-195, Sep. 2017.
[8] R. Viviani, G. Grön, and M. Spitzer, "Functional principal component analysis of fMRI data," *Human Brain Mapping*, vol. 24, no. 2, pp. 109-129, 2005.
[9] A. Krishnan, L. J. Williams, A. R. McIntosh, and H. Abdi, "Partial least squares (PLS) methods for neuroimaging: A tutorial and review," *NeuroImage*, vol. 56, no. 2, pp. 455-475, 2011.
[10] F. D. Martino, G. Valente, N. Staeren, J. Ashburner, R. Goebel, and E. Formisano, "Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns," *NeuroImage,* vol. 43, no. 1, pp. 44-58, 2008.
[11] N. Kriegeskorte, R. Goebel, and P. A. Bandettini, "Information-based functional brain mapping," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 10, pp. 3863-3868, 2006.
[12] Y. Li, P. Namburi, Z. Yu, C. Guan, J. Feng, and Z. Gu, "Voxel selection in fMRI data analysis based on sparse representation," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 10, pp. 2439-2451, 2009.
[13] J. Rondina, *et al.*, "SCoRS-a method based on stability for feature selection and mapping in neuroimaging," *IEEE Trans. Med. Imaging*, vol. 33, no. 1, pp. 85-98, 2013.
[14] C. Zhang, S. Song, X. Wen, L. Yao, and Z. Long, "Improved sparse decomposition based on a smoothed L0 norm using a laplacian kernel to select features from fMRI data," *Journal of Neuroscience Methods*, vol. 245, pp. 15-24, 2015.
[15] S. Ryali, K. Supekar, D. A. Abrams, and V. Menon, "Sparse logistic regression for whole-brain classification of fMRI data," *NeuroImage*, vol. 51, no. 2, pp. 752-764, 2010.
[16] V. Michel, E. Eger, C. Keribin, and B. Thirion, "Adaptive multi-class bayesian sparse regression-an application to brain activity classification," in *Proc. MICCAI 2009: fMRI Data Analysis Workshop-Medical Image Computing and Computer Aided Intervention*, 2009, p. 1.
[17] V. Michel, E. Eger, C. Keribin, J. B. Poline, and B. Thirion, "A supervised clustering approach for extracting predictive information from brain activation images," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010, pp. 7-14.
[18] J. Pinto, *et al.*, "Improved 7 tesla resting-state fMRI connectivity measurements by cluster-based modeling of respiratory volume and heart rate effects," *NeuroImage*, vol. 153, pp. 262-272, Jun 2017.
[19] V. Michel, A. Gramfort, G. Varoquaux, E. Eger, and B. Thirion, "Total variation regularization for fMRI-based prediction of behavior," *IEEE Transactions on Medical Imaging*, vol. 30, no. 7, pp. 1328-1340, 2011.
[20] C. Zhang, L. Yao, S. Song, X. Wen, X. Zhao, and Z. Long, "Euler elastica regularized logistic regression for whole-brain decoding of fMRI data," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 7, pp. 1639-1653, 2017.
[21] D. Mumford, *Elastica and Computer Vision*, New York: Springer, 1994, pp. 491-506.
[22] T. Lin, H. Xue, L. Wang, and H. Zha, "Total variation and euler's elastica for supervised learning," *Computer Science*, 2012.
[23] X. C. Tai, J. Hahn, and G. J. Chung, "A fast algorithm for euler's elastica model using augmented lagrangian method," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 313-344, 2011.
[24] J. Tanabe, D. Miller, J. Tregellas, R. Freedman, and F. G. Meyer, "Comparison of detrending methods for optimal fMRI preprocessing," *NeuroImage*, vol. 15, no. 4, pp. 902-907, 2002.
[25] Y. Li, C. Zhang, C. Hou, L. Yao, J. Zhang, and Z. Long, "Stereoscopic processing of crossed and uncrossed disparities in the human visual cortex," *Bmc Neuroscience*, vol. 18, no. 1, p. 80, 2017.
[26] R. Levien, *The Elastica: A Mathematical History*, Electrical Engineering and Computer Sciences University of California at Berkeley, 2008.
[27] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity*, Cambridge University Press, 2013.
[28] E. Bae, J. Shi, and X. C. Tai, "Graph cuts for curvature based image denoising," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1199-1210, 2011.
[29] L. Ambrosio and S. Masnou, "A direct variational approach to a problem arising in image reconstruction," *Interfaces and Free Boundaries*, vol. 5, no. 1, pp. 63-82, 2003.
[30] B. R. Cottereau, S. P. Mckee, and A. M. Norcia, "Bridging the gap: Global disparity processing in the human visual cortex," *Journal of Neurophysiology*, vol. 107, no. 9, pp. 2421-2429, 2012.
[31] T. J. Preston, S. Li, Z. Kourtzi, and A. E. Welchman, "Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain," *The Journal of Neuroscience*, vol. 28, no. 44, pp. 11315-11327, 2008.
[32] M. L. Patten and A. E. Welchman, "fMRI activity in posterior parietal cortex relates to the perceptual use of binocular disparity for both signal-in-noise and feature difference tasks," *Plos One*, vol. 10, no. 11, 2015.
[33] N. Goncalves, H. Ban, R. M. Sanchezpanchuelo, S. T. Francis, D. Schluppeck, and A. E. Welchman, "7 tesla FMRI reveals systematic functional organization for binocular disparity in dorsal visual cortex," *The Journal of Neuroscience*, vol. 35, no. 7, pp. 3056-3072, 2015.
[34] J. A. Swets, *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*, Psychology Press, 2014.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

**Chuncheng Zhang** was born in Tianjin, China in 1986. He received the PhD degree in computer engineering from Information Science and Technology school of Beijing Normal University, Beijing, China in 2018. He is currently working as post-doc in Institute of Automation Chinese Academy of Sciences, Beijing, China. He is interested in research areas of cognitive neural science and medical image processing using machine learning methods.

**Zhiying Long** received the B.S. degree in electronics and the Ph.D. degree in psychology from Beijing Normal University, Beijing, China, in 1998 and 2005, respectively. Her research interests include brain data mining and brain state decoding.