January 2008

A novel image analysis method based on bayesian segmentation for event-related functional MRI

Lejian Huang

Mary L. Comer

Thomas M. Talavage

Huang, Lejian; Comer, Mary L.; and Talavage, Thomas M., "A novel image analysis method based on bayesian segmentation for event-related functional MRI" (2008). *Department of Electrical and Computer Engineering Faculty Publications.* Paper 55.
http://dx.doi.org/http://dx.doi.org/10.1117/12.774977

A Novel Image Analysis Method Based on Bayesian Segmentation for Event-related Functional MRI

Lejian Huang[1], Mary L. Comer[1], Thomas M. Talavage[1,2,3]

[1]School of Electrical and Computer Engineering, Purdue University
[2]Weldon School of Biomedical Engineering, Purdue University
[3]Department of Radiology, Indiana University School of Medicine

ABSTRACT

This paper presents the application of the expectation-maximization/maximization of the posterior marginals (EM/MPM) algorithm to signal detection for functional MRI (fMRI). On basis of assumptions for fMRI 3-D image data, a novel analysis method is proposed and applied to synthetic data and human brain data. Synthetic data analysis is conducted using two statistical noise models (white and autoregressive of order 1) and, for low contrast-to-noise ratio (CNR) data, reveals better sensitivity and specificity for the new method than for the traditional General Linear Model (GLM) approach. When applied to human brain data, functional activation regions are found to be consistent with those obtained using the GLM approach.

Keywords: fMRI, EM/MPM algorithm, posterior probability map, white noise model, AR(1) model

1. INTRODUCTION

Functional MRI (fMRI) is a form of neuroimaging, based on changes in blood flow and blood oxygenation, which can permit identification of those parts of the brain activated by sensation, cognition or activity. The conventional fMRI analysis approaches, like general linear model (GLM) [1] or cross-correlation [2], are univariate in the spatial domain. In other words, they determine activated regions through matching of time-series images to a given reference waveform voxel by voxel. Although the resulting mapping image will be analyzed spatially, some valuable spatial information may have been lost during the voxel-by-voxel matching. The proposed EM/MPM method for event-related fMRI identifies the activated region based on spatial and temporal information together. Firstly, it estimates the peak time by fitting the time series data to a given hemodynamic response model. Secondly, it segments the activation regions in an averaged peak-time image which preserves spatial information.

It is assumed that there are two states for every voxel in time series: activated or non-activated. Without considering the effect of noise and anatomical differences in the human brain, the intensity of those non-activated voxels in time-series images should be comparable across states, but those activated voxels will exhibit intensity

Computational Imaging VI, edited by Charles A. Bouman, Eric L. Miller, Ilya Pollak,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6814, 68140D, © 2008 SPIE-IS&T · 0277-786X/08/$18

variation corresponding to the hemodynamic response. Therefore, for every image in which activated and non-activated voxels coexist, the image may be deemed to be made up of two different phases. The signal detection problem now is changed into an image segmentation problem. So we can identify activated regions with only two fMRI images (an activated image + a resting image), at least in theory. Considering the low contrast-to-noise rate (CNR) for fMRI data, it is improbable with only those two images. We need to find a way to increase CNR for segmentation. Averaging the peak-time images in multiple events may be a solution for increasing CNR. But it is dependent on the fMRI noise in time series. Thus, we need to discuss how two different temporal noise models affect CNR in section 2. In section 3, we will discuss EM/MPM algorithm. Finally, we will apply this new method to synthetic data and human brain data and compare the results to those derived from GLM.

2. CONTRAST-TO-NOISE RATIO AND NOISE MODELS

2.1 Contrast-to-Noise Ratio

The contrast for one activated image is defined as the difference of signal intensity between activated and non-activated voxels if the activated image is only composed of these two types of voxels. The noise is defined as the standard deviation of non-activated voxels. If the noise in activated images is assumed unvaried, the peak-time image has the largest contrast-to-noise ratio in all image of one response. In our method, the peak time is estimated through matching fMRI data to a Gamma-variate model [3][4]. Usually, the CNR of one peak-time image is still not large enough for segmentation. In order to improve CNR, we average peak-time images over multiple events. In the following two subsections, we will discuss the relationship between CNR and averaging peak-time images over events.

2.2 White Noise Model

If the fMRI noise in the temporal domain is a white noise, each voxel in a time series is represented as

$$Y_i = H_i + \varepsilon_i \tag{1}$$

where Y_i is the measured fMRI voxel value at the i*th* time point; H_i is hemodynamic response signal at the i*th* time point; ε_i is a random error term with Gaussian distribution N(0, σ^2); ε_i and ε_j ($i \neq j$) are uncorrelated. Let K be the number of events, each producing a hemodynamic response assumed to be of length L. If the peak time t_p for every event-related response is unvaried, then $\overline{Y} = \sum_i Y_i$ ($i = t_p , t_{p+L} , \ldots , t_{p+L(K-1)}$) is a Gaussian distribution with mean H_{t_p} and variance σ^2 /K. So the CNR of $\overline{Y}$ is $\sqrt{K}$ times the CNR of Y_{t_p}.

2.3 AR(1) Noise Model

A more appropriate model for the noise observed in an fMRI time-series is an autoregressive model of order 1 [5]. Each voxel in the time series is then represented as

$$Y_i = H_i + w_i + \eta_i \tag{2}$$

where

$$w_i = \rho w_{i-1} + \varepsilon_i \tag{3}$$

Y_i and H_i are same as in the white noise model. w_i is the AR(1) physiological noise with correlation coefficient ρ and ε_i is a zero-mean white Gaussian noise with variance σ_ε^2. η_i is the scanner noise, which is assumed to be N(0, σ_η^2).

We have $E\{w_i\}=0$, $\mathrm{var}\{w_i\}=\mathrm{var}\{w_{i+n}\}=\frac{\sigma_\varepsilon^2}{1-\rho^2}$ and $\mathrm{cov}(w_{i+n},w_i)=\frac{\sigma_\varepsilon^2}{1-\rho^2}\rho^{|n|}$ (n is time point interval) so that

$E\{Y_i\} = H_i$, $E\{Y_{i+n}\} = H_{i+n}$, $\mathrm{var}\{Y_i\}=\frac{\sigma_\varepsilon^2}{1-\rho^2}+\sigma_\eta^2$ and $\mathrm{cov}(Y_{i+n},Y_i)=\frac{\sigma_\varepsilon^2}{1-\rho^2}\rho^{|n|}$. So the CNR of Y_{t_p} is given by the following equation:

$$CNR_{Y_{t_p}} = \frac{H_{t_p}}{\sqrt{\frac{\sigma_\varepsilon^2}{1-\rho^2}+\sigma_\eta^2}} \tag{4}$$

Let K be the number of trials, each producing a hemodynamic response assumed to be of length L. If the peak time t_p for every event-related response is unvaried, then $\bar{Y} = \sum_i Y_i$ is also follows a Gaussian distribution with $E\{\bar{Y}\} = H_{t_p}$ and

$$\mathrm{var}\{\bar{Y}\} = \frac{1}{K}\left(\frac{\sigma_\varepsilon^2}{1-\rho^2}+\sigma_\eta^2\right) + \frac{2}{K^2}\sum_{i=1}^{K-1}(K-i)\frac{\sigma_\varepsilon^2}{1-\rho^2}\rho^{i\times L} \tag{5}$$

Note that the first term of equation (5), $\frac{1}{K}\left(\frac{\sigma_\varepsilon^2}{1-\rho^2}+\sigma_\eta^2\right)$, is equal to $\frac{1}{k}\mathrm{var}\{Y_{t_p}\}$. Let us substitute typical parameters' value into equation (5) to see the ratio ζ between the second term of equation (5), $\frac{2}{K^2}\sum_{i=1}^{K-1}(K-i)\frac{\sigma_\varepsilon^2}{1-\rho^2}\rho^{i\times L}$, and the first term of equation (5) in term of physiological noise correlation coefficient ρ. The result is shown in figure 1.

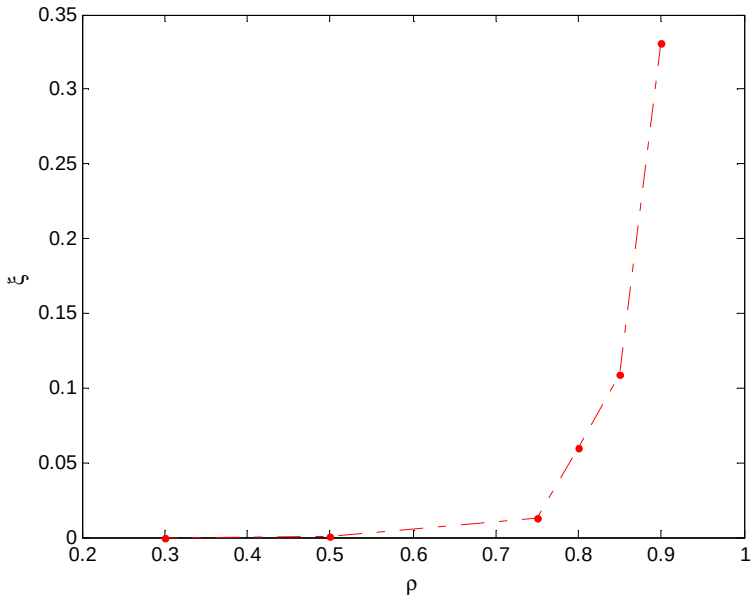


Figure 1. The ratio of the second part of Eq. (5) to the first part in terms of physiological noise correlation coefficient (K=15, L=15, $\sigma_\eta = 1$, $\sigma_\varepsilon = \frac{\sqrt{7}}{4}$)

From figure 1, it is shown that when $\rho \leq 0.8$, the second term of Eq. (5) will be significantly less than the first term. So, $\mathrm{var}\{\bar{Y}\}$ will be approximated by $\frac{1}{K}(\frac{\sigma_\varepsilon^2}{1-\rho^2} + \sigma_\eta^2)$ and CNR of $\bar{Y}$ by $\sqrt{K} \times CNR_{Y_{t_p}}$, which is consistent with the result derived from white noise model. Thus, we conclude that averaging the peak-time images over events can significantly improve the CNR, which will make it possible to segment and identify the activated regions using fewer time-series images.

3. THE EM/MPM ALGORITHM

Before we discuss the EM/MPM algorithm, some notations which will be used in the following discussion are written as follows:

S: set of lattice points and $S=\{(i,j)|0 \leq i \leq N_1 - 1, 0 \leq j \leq N_2 - 1\}$, where N_1 and N_2 are the number of voxels in the x and y directions;

s: a lattice point, $s \in S$;

η: a neighborhood system and $\eta=\{\eta_s|s \in S\}$, where each η_s is a neighborhood of s satisfying (i) $s \notin \eta_s$ and (ii) $s \in \eta_t$ implies $t \in \eta_s$;

c: a clique, a set of points, which are all neighbors of each other;

C: the collection of all cliques in S;

N: the number of voxel in one image = $N_1 \times N_2$;

X: the label field, containing the classification of each voxel to the activated or non-activated state;

Y: the observed image (average peak-time image);

X_s: the element in **X** at s, a random variable;

Y_s: element in **Y** at s, a random variable;

$\boldsymbol{\theta}$: parameter vector and $\boldsymbol{\theta}$ is defined as $[\mu_0, \sigma_0^2, \mu_1, \sigma_1^2]$ in our method.

3.1 2-D Discrete Markov Random Field

We assume that activated voxels will have different statistical characteristics from non-activated voxels, allowing us to convert the signal detection problem into a segmentation problem. Depending on its statistical properties and its neighbors, each voxel in the activated fMRI image must be assigned to one of two labels, activation ($X_s = 1$) or non-activation ($X_s = 0$), forming a two-dimensional label field. In this paper, the Markov random field model is used for the label field. By the Hammersley-Clifford Theorem, the probability mass function of **X** takes the form of a Gibbs distribution over the collection of all cliques, *C*:

$$p_X(x) = \frac{1}{Z} \exp\left(-\beta \sum_{\{r,s\} \in C} \delta(x_r \neq x_s) \right) \tag{6}$$

where β is the spatial interaction parameter; Z is the normalizing constants for the density; $\delta(x_r \neq x_s)=0$ if $x_r = x_s$ and $\delta(x_r \neq x_s)=1$ if $x_r \neq x_s$. This is actually a special case of the Gibbs distribution known as the ising model.

3.2 Observed Image Model

The label state X_s for every voxel is not observed directly and must be estimated from the observed image **Y**. It is assumed that random variables $Y_1, Y_2, \ldots, Y_N$ are conditionally independent given the label field **X**. it is also assumed that the conditional probability density of Y_s depends only on X_s, and is independent of any other element of **X.** We model Y_s to be conditionally Gaussian given X_s. This leads to the following expression for the conditional probability density function of **Y**:

$$f_{Y|X}(y \mid x, \theta) = \prod_{s=1}^{N} f_{Y_s|X}(y_s|x,\theta) = \prod_{s=1}^{N} f_{Y_s|X_s}(y_s|x_s,\theta) = \prod_{s=1}^{N} \frac{1}{\sqrt{2\pi\sigma_{x_s}^2}} \exp\left(-\frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2} \right), \tag{7}$$

where $\theta = [\mu_0, \sigma_0^2, \mu_1, \sigma_1^2]$. The posterior probability mass function of **X** given **Y** is obtained as follows:

$$p_{X|Y}(x \mid y, \theta) = \frac{f_{Y|X}(y|x,\theta)\, p_X(x)}{f_Y(y|\theta)} =$$

$$\frac{1}{Zf_Y(y|\theta)}\left[\prod_{s=1}^{N}\frac{1}{\sqrt{2\pi\sigma_{x_S}^2}}\right]\exp\left(-\sum_{s=1}^{N}\frac{(y_S-\mu_{x_S})^2}{2\sigma_{x_S}^2}-\beta\sum_{\{r,s\}\in C}\delta(X_r\neq X_S)\right). \tag{8}$$

3.3 MPM Algorithm

MPM is one type of Bayesian estimation, which is to choose the estimate $\hat{x}=\hat{x}(y)$ by minimizing the expected cost E(C(X, $\hat{x}$(Y))) taken with respect to the posterior probability distribution $P_{X|Y}$, where C(X, $\hat{x}$(Y)) is the cost of $\hat{x}$ when x is the correct answer. Maximum *a posterior* (MAP) is considered too conservative because it treats all errors equally. MPM is a more reasonable approach, which assigns a cost proportional to the number of error voxels and its cost function is defined as

$$C(x,\hat{x})=\sum_{s\in S}\left[1-\delta(x_s-\hat{x}_s)\right] \tag{9}$$

The general criterion for optimal class labeling is to minimize the following expected cost function:

$$E\left[C(X,\hat{x}(Y))\right]=\sum_{x,y}C(x,\hat{x}(Y))P(x\mid y)P(y)=\sum_{y}E\left[C(X,\hat{x}(Y))\mid Y=y\right]p(y) \tag{10}$$

For MPM, we substitute Eq. (9) to Eq. (10). Thus, for each y

$$E\left[C(X,\hat{x}(Y))\mid Y=y\right]=N-\sum_{s\in S}\left[\sum_{x_{s-}}P(x_s=\hat{x}_s,x_{s-}\mid y)\right] \tag{11}$$

where *N* is the number of voxel in one image, x_{s-} = {x_r: r∈S and r ≠ s} and $\sum_{x_{s-}}P(x_s=\hat{x}_s,x_{s-}\mid y)$ is a marginal probability of P(x|y) for x_{s-} with $x_s=\hat{x}_s$. For each voxel s∈S, the estimator is to find the value of k (k = 1 for activation, k = 0 for nonactivation) which maximizes

$$P(X_s=k\mid Y=y)=\sum_{x\in S_{k,s}}p_{X|Y}(x\mid y,\theta), \tag{12}$$

where $S_{k,s}$ = {**x**: x_s = k}. But computing the above marginal probability mass functions is very time-consuming and infeasible. In [6], by using a Markov chain Monte Carlo (MCMC) sampling process which converges in distribution to a random field with probability mass function given by Eq. (8), the authors approximate the marginal probability mass function with the fraction of time the Markov chain spends in state k at voxel s, for each k and s. A pseudocode is shown below to explain how to count the fraction of time the Markov chain spends in state k = 1 (activation) at each voxel s in the MPM algorithm:

$N_{act} \leftarrow$ count_act_num( ) {
  for m = 1, …, MPM_iteration_number {
    $N_{act} = 0$; //Initialize to 0
    posterior $\leftarrow \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\sum_{s=1}^{N} \frac{(y_s-\mu_1)^2}{2\sigma_1^2} - \beta \sum_{\{r,s\}\in C} \delta(X_r \neq (X_s=1)) \right)$;
    random_num $\leftarrow$ random( ); // random( ): a uniform random function
    If (random_num $\leq$ posterior) {
      $N_{act} = N_{act} + 1$;
    }
   return N_{act};
  }
}
fraction= N_{act}/MPM_iteration_number;

3.4 EM Algorithm

In order to approximate the above marginal probabilities, the value of parameter vector **θ** must be estimated. We use the Expectation-Maximization (EM) algorithm to estimate the parameters $\left(\mu_k, \sigma_k^2\right)$ in every iteration *p*. Here we only give the final equations to update the parameters in EM algorithm. The detailed discussion is included in [7].

$$\mu_k^{(p+1)} = \frac{1}{N_k^{(p+1)}} \sum_{s=1}^{N} y_s \, p_{X_s|Y}(k|y,\theta^{(p)}) \tag{13}$$

$$\sigma_k^{2(p+1)} = \frac{1}{N_k^{(p+1)}} \sum_{s=1}^{N} (y_s-\mu_k^{(p+1)})^2 \, p_{X_s|Y}(k|y,\theta^{(p)}) \tag{14}$$

$$N_k^{(p+1)} = \sum_{s=1}^{N} p_{X_s|Y}(k|y,\theta^{(p)}) \tag{15}$$

3.5 EM/MPM Algorithm

So, the whole EM/MPM algorithm procedure is described as follows:

Step1: Initialize estimate of parameter vector **θ(0)**.

Step2: For iteration p = 1, …, P

i) perform a certain number of iteration of the MPM algorithm based on Eq. (12) and the estimate of parameter vector **θ(p-1)**;

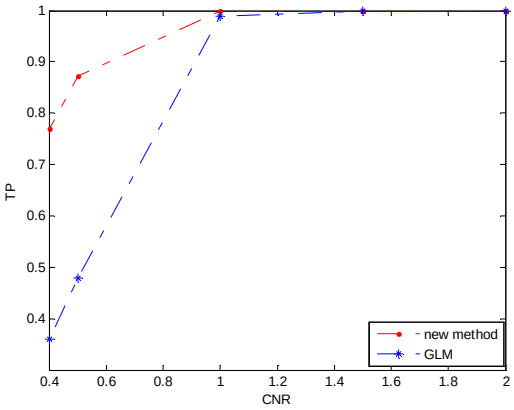
ii) Approximate the estimate of the marginal probability mass function and then use EM update Eqs (13-15) to obtain the estimate of parameter vector **θ(p)**.

4. RESULTS

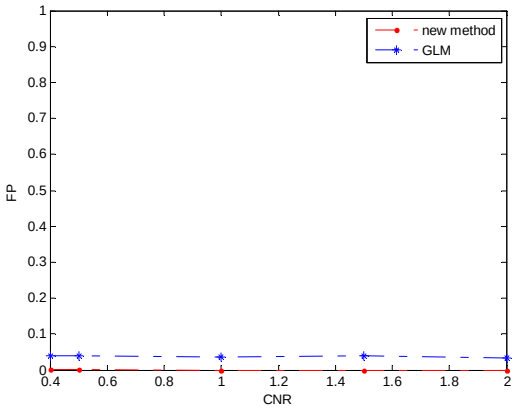
4.1 Synthetic Data Results

The EM/MPM method produces a posterior probability map (PPM) from which the activation region can be identified with its statistical significance.

Using simulation conditions similar to those defined in [4][5] where σ_η=1 , σ_ε=$\frac{\sqrt{7}}{4}$ and ρ=0.75 , true positive (TP) and false positive (FP) rates were determined and are shown in figure 4 ((α = 0.05, *p value* $\leq$ 0.05 $\Rightarrow$ activation for GLM; posterior probability $\geq$ 0.95 $\Rightarrow$ activation for the new method).



(a) Relation between True Positive (TP) rate and Contrast-to-noise (CNR) for new method and GLM



(b) Relation between False Positive (FP) rate and Contrast-to-noise (CNR) for new method and GLM

Figure 4. True Positive(FP) and False Positive(FP) rate comparisons between new method and GLM

From figure 4(a) and (b), we can see that when the CNR is high ($\geq$ 1), both the proposed and popular GLM methods can correctly detect almost all the activated voxels, however the proposed method maintains a lower false positive rate. When CNR is lower (<1), the proposed method can significantly increase detection performance, exhibiting both a higher true positive rate (greater sensitivity) and lower false positive rate (greater specificity).

When we set the CNR at 0.7 and use posterior probability and p value as its discrimination threshold for the EM/MPM method and GLM, respectively, the receiver-operator-characteristic curves are generated as in figure 5. From this figure, we conclude that the new method will have higher detection performance than that derived from GLM, i.e., higher true positive rate and lower false positive rate.

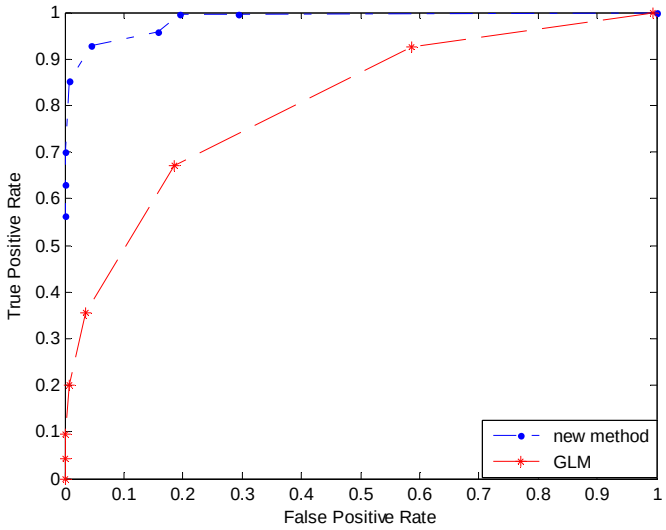
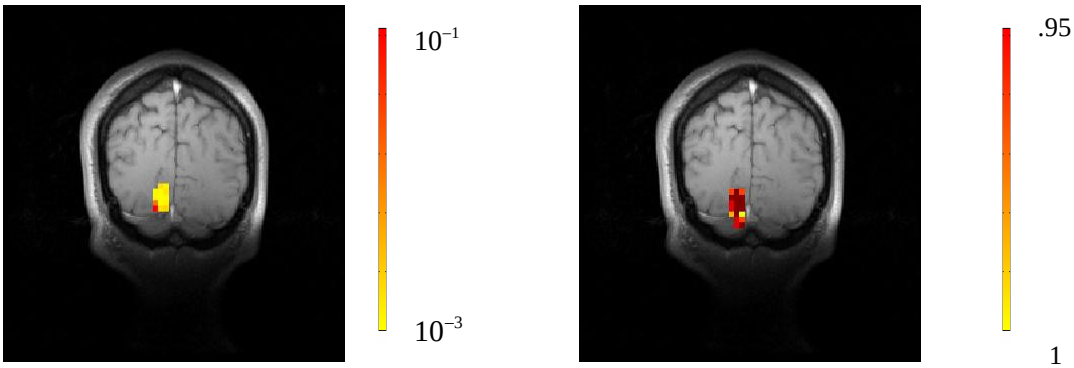


Figure **5.** ROC curves of the new method and GLM when CNR=0.7

4.2 Human Brain Data Results

The EM/MPM method is applied to the same human brain data as in [4] to assess consistency with the GLM approach. The new method uses posterior probability to mark the activation location and statistical significance. The GLM approach is used here as the "gold standard" based on its wide acceptance within the fMRI community. From figure 6, we can observe that, as expected based on the synthetic data analyses, results of the proposed algorithm are consistent with those derived from the GLM approach (α =0.05, *p value* $\leq$ 0.05 $\Rightarrow$ activation for GLM; posterior probability $\geq$ 0.95 $\Rightarrow$ activation for the new method).



(a) Functional map of GLM (p value threshold = .05)

(b) Functional map of new method (posterior marginal probability threshold = 0.95)

Figure **6.** Functional images for GLM (α=0.05) and the new method. The left and yellow-to-red colorscale corresponds the p value for GLM and the posterior probability for the new method, respectively.

5. CONCLUSION

We have demonstrated that the accuracy of signal detection in fMRI can be significantly improved. Our results show that the proposed EM/MPM method can achieve greater sensitivity and specificity than the popular GLM approach, particularly at lower CNR values. When EM/MPM method is applied to human brain data, the detected functional activation region is consistent with that of the GLM approach, suggesting that performance is at least as effective.

6. REFERENCES

[1] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *Human Brain Mapping*, 2:189-210, 1995.

[2] P. A. Bandettini, A. Jesmanowicz, E. C. Wong, J. S. Hyde, "Processing Strategies for Time-Course Data Sets in Functional MRI of the Human Brain," *Magn Reson Med*, vol.30, pp.161- 173, 1993.

[3] A. M. Dale, R. L. Buckner, "Selective averaging of rapidly presented individual trials using fMRI," *Human Brain Mappping*, 5:329–340, 1997

[4] A. Rao. Model Based Estimation And Detection of Hemodynamic Response in Event-Related FMRI. *Ph.D. dissertation*, Purdue University, 2005.

[5] P. L. Purdon, V. Solo, R. M. Weiskoff, and E. N. Brown. Locally regularized spatiotemporal modeling and model comparison for functional MRI. *Neuroimage*, 14:912-923, 2001.

[6] J. Marroquin, S. Mitter, T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Statist. Assoc.*, vol. 82, pp. 76-89, Mar. 1987.

[7] M. Comer and E. Delp, "The EM/MPM Algorithm for segmentation of textured images: analysis and further experimental results," *IEEE Trans. on Imaging Processing*, Vol. 9, No. 10, 1731-1744, Oct. 2000