fMRI Binary Detection of Brain Activated Regions with Graph-Cuts

Joana Coelho[1,2], João Sanches[1,2] and Martin H. Lauterbach[3,4] (MD)

Abstract— The *functional Magnetic Resonance Imaging* (fMRI) is a technique with increasing applications in studying the brain function. The *blood-oxygenation-level-dependent* (BOLD) is a fMRI method that allows the detection of brain activated regions after the application of an external stimulus, e.g., visual or auditive. This technique is based on the assumption that the metabolism increases in activated areas as well as the oxygen uptake. Analising this information is a challenging problem because the BOLD signal is very noisy and its changes due to the application of a stimulus are very weak. Therefore, the detection of temporal correlations with the applied stimulus requires sophisticated statistical algorithms to understand if the changes on the BOLD signal are pure noise or are related with the applied stimulus, called *paradigm* in the fMRI scope.

The traditional approach to detect activated regions is based on the *general linear model* (GLM) to describe the BOLD signal using statistical inference techniques to infer the activation. Unfortunately, this technique requires the tunning of parameters by a clinician which makes it impossible to be completely automatic.

In this paper we propose a new technique, here called SPM-GC, designed in a Bayesian framework where the *explanatory variables* (EV's) that characterize the activation status of a given region are considered to be binary. In this approach the classification task is modeled as a huge combinatorial optimization task that is optimally solved by using a recent optimization technique based on *Graph-Cuts*.

Exhaustive tests using synthetic data are presented and the error probability, P_e, of the algorithm is characterized. Examples using real data are also presented to illustrate the application and performance of the algorithm.

Index Terms— functional MRI, Estimation, Denoising, Bayes, Graph-Cuts.

I. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) is a new technique for studying the dynamic processes occurring in the brain of living beings, namely humans. This technique has an increasing number of clinical applications such as the characterization and mapping of functional areas in the damaged brain, defining mechanisms of reorganization or compensation from injury and also helping in brain surgical planning [1].

fMRI is based on the assumption that the metabolism increases in activated areas as well as the oxygen uptake. This oxygen increasing is measured by the *blood-oxygenation-level-dependent* (BOLD) signal, represented in the functional magnetic resonance images as displayed in Fig. 4. The BOLD signal is very noisy and its changes due the application of stimuli are very weak. Therefore, the detection of temporal correlations with the applied stimuli requires sophisticated statistical algorithms to assess if the changes on the BOLD signal are pure noise or are related with applied stimulus, called *paradigm* in the fMRI scope.

The usual approach to detect correlations with the paradigms is based on the *General Linear Model* (GLM) [2], [3], [1] where the BOLD signal is modeled as a noisy version of the linear combination of the stimuli signals convolved with a *hemodynamic*

response function (HRF). The *explanatory variables* (EVs) are the coefficients of the linear combination, estimated by linear regression. The most used algorithm is the SPM-GLM [3] which is based on a classical inference statistical test, e.g., T or F, where a p-value threshold is used to attribute a statistical significance to each coefficient and, therefore, assess if a given area was activated or not by the correspondent stimulus. Other approaches have been also proposed based on the *Principal Component Analysis* (PCA) [4], *Independent Component analysis* (ICA) [5] or Bayesian approach [6], [7].

An additional difficulty is the *hemodynamic response function*. Usually, the HRF is unknown and can be different for each voxel [8], [9]. A single approximation model may be used for the whole brain, usually based on gamma functions [3] or it may be estimated for each voxel [10].

In this paper, a new Bayesian algorithm is proposed, here called SPM-GC, where the estimation and inference of the *explanatory variables* (EV) are performed together instead of the usual approach of doing it independently. Here, for each stimulus, the EV of each voxel is considered to be binary. The algorithm jointly estimates these variables and the HRF at each voxel. The prior used to estimate the HRF is physiologically supported as proposed in [11].

The estimation of the EVs in a Bayesian framework corresponds to a huge combinatorial optimization problem that may be optimally solved by using the algorithm proposed in [12] based on *Graph-Cuts*. This algorithm is fast and is able to find out the global minimum of the cost function.

The advantage of this algorithm is being parameter free which allows to circumvent the subjectivity associated with the algorithms that depend on tunning parameters defined by the clinician.

Monte Carlo tests using synthetic data are used to characterize the robustness of the algorithm from an error probability point of view. Experiments using real data are also presented to illustrate its application in real situations. Furthermore, the results obtained with real data are compared with the ones obtained by the medical doctor with the *BrainVoyager* software [13]. This comparison shows that the proposed algorithm provides similar results with the ones obtained with *BrainVoyager* without the need of any parameter adjustment removing the subjectivity associated to the results.

This paper is organized as follows. Section II formulates the problem from a mathematical point of view and in section III the experimental results are presented. Section IV concludes the paper.
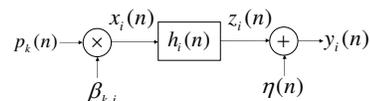

Fig. 1. BOLD signal generation model.

II. PROBLEM FORMULATION

Let $\mathbf{y}_i = \{y_i(n)\}$ with $0 \leq n \leq L-1$ be the BOLD signal at i^{th} voxel generated according Fig.1 where $\mathbf{h}_i = \{h_i(n)\}$ is the HRF associated with the ith voxel, $\eta(n) \sim \mathcal{N}(0, \sigma_y^2)$ is *additive white*

Gaussian noise (AWGN), $p_k(n)$ is the kth stimulus and $\beta_{k,i}$ is the binary *explanatory variable* associated with it. The binary variable $\beta_{k,i}$ is 1 when the ith voxel is activated by the kth stimulus and 0 otherwise. When N stimuli are applied simultaneously

$$y_i(n) = \overbrace{h_i(n) * \underbrace{\sum_{k=1}^{N} \beta_{k,i} p_k(n)}_{x_i(n)}}^{z_i(n)} + \eta(n). \quad (1)$$

The signal $\mathbf{x}_i = \{x_i(n)\}$ (see Fig.1) may be expressed as $\mathbf{x}_i = \theta \mathbf{b}_i$ where $\mathbf{b}_i = \{\beta_{1,i}, \beta_{2,i}, ..., \beta_{N,i}\}$ and

$$\theta = \begin{pmatrix} p_1(1) & p_2(1) & p_3(1) & ... & p_N(1) \\ p_1(2) & p_2(2) & p_3(2) & ... & p_N(2) \\ p_1(3) & p_2(3) & p_3(3) & ... & p_N(3) \\ \vdots & \vdots & \vdots & ... & \vdots \\ p_1(L) & p_2(L) & p_3(l) & ... & p_N(L) \end{pmatrix} \quad (2)$$

The output of $h_i(n)$, $z_i(n) = h_i(n) * x_i(n)$, may be obtained by $\mathbf{z}_i = \mathbf{H}_i \mathbf{x}$ or by $\mathbf{z}_i = \mathbf{\Phi}_i \mathbf{h}$ where $\mathbf{H}_i$ and $\mathbf{\Phi}_i$ are the following $L \times L$ Toeplitz matrices (it is assumed that $\mathbf{h}$ has the same length of $\mathbf{p}_k = \{p_k(n)\}$),

$$\mathbf{H}_i = \begin{pmatrix} h_i(1) & 0 & 0 & 0 & 0 & 0 \\ h_i(2) & h_i(1) & 0 & 0 & 0 & 0 \\ h_i(3) & h_i(2) & h_i(1) & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ h_i(L) & ... & h_i(3) & h_i(2) & ... & h_i(1) \end{pmatrix} \quad (3)$$

$$\mathbf{\Phi}_i = \begin{pmatrix} x_i(1) & 0 & 0 & 0 & 0 \\ x_i(2) & x_i(1) & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 \\ x_i(L) & x_i(L-1) & ... & ... & x_i(1) \end{pmatrix}. \quad (4)$$

The vector $\mathbf{y}_i$ may be, therefore, obtained by the following two ways

$$\mathbf{y}_i = \mathbf{\Psi}_i \mathbf{b}_i + \mathbf{n} \quad (5)$$
$$\mathbf{y}_i = \mathbf{\Phi}_i \mathbf{h}_i + \mathbf{n} \quad (6)$$

where $\mathbf{\Psi}_i = \mathbf{H}_i \theta$ and $\mathbf{n} = \{\eta(n)\}$.

The *Maximum a Posteriori* (MAP) estimation is obtained by minimizing the following energy function

$$E(\mathbf{y}_i, \mathbf{b}_i, \mathbf{h}_i) = E_y(\mathbf{y}_i, \mathbf{b}_i, \mathbf{h}_i) + E_h(\mathbf{h}_i) \quad (7)$$

where the *data fidelity term* is $E_y(\mathbf{y}_i, \mathbf{b}_i, \mathbf{h}_i) = -\log(p(\mathbf{y}_i|\mathbf{z}_i(\mathbf{b}_i, \mathbf{h}_i)))$ and the prior term associated to $\mathbf{h}_i$ is $E_h(\mathbf{h}_i) = -\log(p(\mathbf{h}_i))$.

This prior incorporate the *a priori* knowledge about $\mathbf{h}_i$ which is, according with [11]:
i) the HRF starts and ends at 0 and
ii) HRF is smooth.

The smoothness [14] of $\mathbf{h}$ may be imposed by assuming that $p(\mathbf{h}_i)$ is a *Gibbs* distribution with quadratic potential functions,

$$p(\mathbf{h}_i) = \frac{1}{Z_h} e^{-\alpha \sum_{n=2}^{N} (h_i(n) - h_i(n-1))^2} \quad (8)$$

which leads to $E_h(\mathbf{h}_i) = -\log(p(\mathbf{h}_i)) = \alpha(\mathbf{\Delta}\mathbf{h}_i)^T(\mathbf{\Delta}\mathbf{h}_i) + C_h$ where α is a parameter that tunes the smoothing degree for $h_i(n)$, C_h is a constant, Z_h is a partition function and $\mathbf{\Delta}$ is the following difference operator

$$\mathbf{\Delta} = \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 & -1 \\ -1 & 1 & 0 & ... & 0 & 0 & 0 \\ 0 & -1 & 1 & ... & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & ... & -1 & 1 & 0 \\ 0 & 0 & 0 & ... & 0 & -1 & 1 \end{pmatrix} \quad (9)$$

Assuming AWGN (after the pre-whitening pre-processing [1]) over each time course the energy function to be minimized may be written in the following two ways,

$$E(\mathbf{y}_i, \mathbf{b}_i, \mathbf{h}_i) = \frac{1}{2\sigma_y^2} (\mathbf{\Psi}_i \mathbf{b}_i - \mathbf{y}_i)^T (\mathbf{\Psi}_i \mathbf{b}_i - \mathbf{y}_i) + D_i (10)$$

$$= \frac{1}{2\sigma_y^2} (\mathbf{\Phi}_i \mathbf{h} - \mathbf{y}_i)^T (\mathbf{\Phi}_i \mathbf{h}_i - \mathbf{y}_i) + \alpha \mathbf{h}_i^T (\mathbf{\Delta}^T \mathbf{\Delta}) \mathbf{h}_i + C_i \quad (11)$$

where C_i and D_i are constants. The minimization of $E(\mathbf{y}_i, \mathbf{b}_i, \mathbf{h}_i)$ with respect to $\mathbf{b}_i$ and $\mathbf{h}_i$ is obtained by solving $\nabla_{\mathbf{b}_i} E = 0$ and $\nabla_{\mathbf{h}_i} E = 0$.

The first minimization is performed by minimizing the energy function (10) which may be written as folllows

$$E(\mathbf{y}_i, \mathbf{b}_i, \mathbf{h}_i) = \frac{1}{2\sigma_y^2} \sum_{k=1}^{N} \sum_{n} (\Psi_i(n,k)\beta_{k,i} - y_i(n))^2 + D_i \quad (12)$$

where each term is

$$\sum_{n} (\Psi_i(n,k)\beta_{k,i} - y_i(n))^2 = \begin{cases} \sum_{n} y_i^2(n) & \beta_{i,k} = 0 \\ \sum_{n} (\Psi_i(n,k) - y_i(n))^2 & \beta_{i,k} = 1 \end{cases} \quad (13)$$

Therefore, the binarization of $\beta_{k,i}$ that leads to the minimization of (10) is the following

$$\beta_{k,i} = \begin{cases} 1 & \text{if } \sum_{n} \left[(\Psi_i(n,k) - y_i(n))^2 - y_i^2(n)\right] \leq 0 \\ 0 & otherwise \end{cases} \quad (14)$$

The second minimization is performed by minimizing the energy function (11) which leads to

$$\hat{\mathbf{h}}_i = \left(\mathbf{\Phi}_i^T \mathbf{\Phi}_i + \lambda(\mathbf{\Delta}^T \mathbf{\Delta})\right)^{-1} \mathbf{\Phi}_i^T \mathbf{y}_i \quad (15)$$

where $\lambda = 2\alpha\sigma_y^2$.

The estimation of $\mathbf{b}_i$ and $\mathbf{h}_i$ for each voxel is iteratively performed and the overall estimation algorithm is developed according to the prototyping algorithm described in Table I.

1. ▷ Initialization of $\beta_{k,i} = 0.5$ and $\mathbf{h}_i = \mathbf{g}$, where $\mathbf{g}$ is gamma function [3].
2. ▷ set $t = 1$
3. ▷ set $i = 1$,
4. ▷ set $k = 1$
5. ▷ binarize the $\beta_{k,i}$ by minimizing (10)
6. ▷ estimate $\hat{\mathbf{h}}_i^t$ according (15)
7. ▷ increment k and return to step 5 if $k <$ number of stimuli
8. ▷ increment i and return to step 4 if $i <$ number of voxels
9. ▷ increment t and return to step 3 while $\sum_{k,i} |\ell_{k,i}^t - \ell_{k,i}^{t-1}| \neq 0$
10. ▷ binarize the $\beta_{k,i} \to \ell_{k,i}$ for all voxels taking into account spatial correlation by using Graph-Cuts

TABLE I
PROTOTYPE ALGORITHM

The binarization procedure at step 10) is performed independently for each stimulus k, $\beta_{k,i}$, by using the algorithm proposed in [12] where spatial correlation is taken into account. In fact, this is

the only step where the correlation among neighbors is considered. Before, the estimation procedure of $\mathbf{h}_i$ and $\mathbf{b}_i$, associated with each voxels, is performed independently of the neighbors in a *time-course* basis.

Let $\mathbf{B}_k$ be a 3D matrix volume containing the binary EV, $\beta_{k,i}$, at each voxel location with respect to the single k^{th} stimulus obtained at step 5) of the iterative algorithm and $\mathcal{L}_k = \{\ell_{k,i}\}$ a spatial correlated version of $\mathbf{B}_k$ where $\ell_{k,i} \in \{0, 1\}$. In step 10) each $\mathbf{B}_k$ is processed, in a slice-by-slice basis, by solving the following optimization problem

$$\hat{\mathcal{B}}_k = \arg\min_{\mathcal{B}_k} E(\mathbf{B}_k, \mathcal{L}_k) \tag{16}$$

where the energy function is

$$\begin{aligned}
E(\mathbf{B}_k, \mathcal{L}_k) &= \sum_i |\ell_{k,i} - \beta_{k,i}| \tag{17}\\
&+ \alpha \sum_i \left[V(\ell_{k,i}, \ell_{k,i_h}) + V(\ell_{k,i}, \ell_{k,i_v}) \right] / \tilde{g}_i
\end{aligned}$$

and ℓ_{k,i_τ} are the causal neighbors of $\ell_{k,i}$ at each slice, α is a parameter to tune the strength of smoothness, $\tilde{g}_i$ is the normalized ($\epsilon \leq \tilde{g}_k \leq 1$) gradient of $\mathbf{B}_k$ at the ith node and $\epsilon = 10^{-2}$ is a small number to avoid division by zero. $V(\ell_1, \ell_2)$ is a penalization function defined as follows

$$V(\ell_1, \ell_2) = \begin{cases} 0 & \ell_1 = \ell_2 \\ 1 & \ell_1 \neq \ell_2 \end{cases} \tag{18}$$

The energy function (18) is composed by two terms: the first called *data term* and the second called *regularization term*. The first forces the classification to be $\ell_{k,i} = \beta_{k,i}$. The second term forces the uniformity of the solution because the cost associated with uniform labels is smaller than non uniform ones (see equation (18)). However, in order to preserve transitions the terms are divided by the normalized gradient magnitude of $\mathbf{B}_k$ at i^{th} location, $\tilde{g}_i$. Therefore, when the gradient magnitude increases the regularization strength is reduced at that location.

The minimization task of (18), formulated in (16), is a huge combinatorial optimization problem in the $\{0, 1\}^M$ high dimensional space where M is the number of voxels in each 3D volume.

In [12] it is shown that several energy minimization problems in high dimensional discrete spaces can be efficiently solved by using *Graph-Cuts* (GC) based algorithms. The authors have designed a very fast and efficient algorithm to compute the global minimum of the energy function. However, the algorithm is not completely general which means that some energy functions can not be minimized with the proposed method. In [15] the authors present a wide class of energy functions that may be minimized with the GC method. Fortunately, the function (16) belongs to that class.

III. EXPERIMENTAL RESULTS

A. Synthetic Data

In this section, Monte Carlo tests of the SPM-GC algorithm are presented in order to characterize its robustness. Two synthetic binary images of 128×128 pixels were generated representing the regions activated by two complementary stimuli. However, only one stimulus is presented here where the white voxels represent the activated regions. The BOLD signal, $y(n)$, was generated using the model previously presented in Fig.1. The paradigm was generated in a block-design basis of 4 epochs, 20 seconds each (10 seconds of activation and 10 seconds of rest). The HRF signal was a basic gamma function known by its physiological meaning [11]. To evaluate the performance of the algorithm several noise levels

were tested in the range $\sigma_y = [0; 5]$ which can also be compared to the BOLD signal energy level by the *signal-to-noise ratio* (SNR). This generated synthetic data, composed by $2 \times 128 \times 128 = 32768$ independent $y(n)$ time courses is equivalent to perform 32768 runs of Monte Carlo tests, were used to compute the error probability, $P_e(\sigma) = \frac{1}{NM} \sum_{i=1}^{NM} \left| \hat{b}_i - b_i \right|$ where $N = M = 128$. Fig. 2 displays the algorithm error probability with (red) and without (blue) the post-processing *Graph-Cut* binarization at step 10 (see Table I). The error probability is always smaller when using the post-processing step. Even for high levels of noise ($SNR > -25$dB), the algorithm with post-processing manage to detect the correct activated regions without misclassification. Fig.3 shows an example of the activation detection for a SNR=-24.5 time-course ($\sigma_y = 3.4$) with (left) and without (right) post-processing with *Graph-Cuts*.
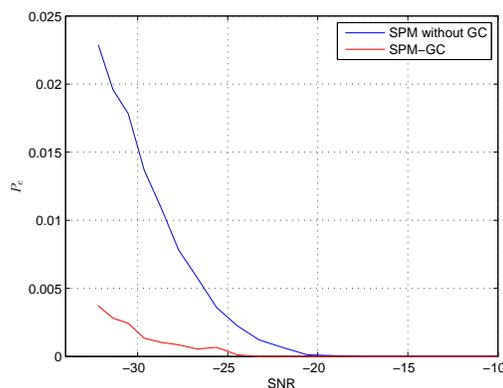

Fig. 2. Graphic with the computed error probability for each noise level of each algorithm (SPM-GC and SPM proposed algorithm without the step 10).
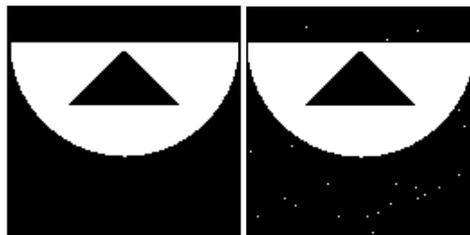

Fig. 3. Activated regions of a synthetic binary image with SNR=-24.5 for one of the paradigm: *left* - Activated regions using SPM-GC where $P_e = 0$; *right* - Activated regions using the algorithm without the previously mentioned step 10 where $P_e = 0.0012$.

B. Real Data

Two volunteers participated on stimulated verbal and motor activity during a fMRI data acquisition on a *Philips Intera Achieva Quasar Dual 3T* whole-body system with a 8 channel head-coil. T_2*-weighted echo-planar images (EPI) 23cm square field of view with 128×128 matrix size resulting in an in-plane resolution of $1, 8 \times 1, 8mm$ for each 4 mm slice. echo time=33ms, flip angle=20^0 were acquired with *TR*=300ms. The paradigms were all structured on the same block-design, with 20 samples per epoch (meaning 10 samples of stimulus followed by 10 samples of baseline, summing up to 60s time per epoch) and a total of 4 epochs. The fMRI data was preprocessed with the standard procedures implemented in the *BrainVoyager* software [13] for motion correction, registration, whitening and spatial smoothing. This data was then statistically processed by the *BrainVoyager* SPM-GLM algorithm and by the

SPM-GC algorithm. The SPM-GLM brain maps depend on the p-value tuned by the clinician. A neurologist provided the results of SPM-GLM giving a *reference* result (the one he considered to be the correct one) and he also provided two other results which he considered to be *loose* and *restricted*. The activated regions of SPM-CG algorithm are coded with color intensity gradient, inversely proportional to the energy function of (11). This applied colormap gives an important perception of the confidence of this results since the low intensity regions correspond to a higher value of the energy function (see 11) that is being minimized. Visual inspections of the results in Figs. 4 and 5 show some resemblance between the reference result of the SPM-GLM brain maps and the ones obtained by the SPM-GC algorithm. Although the SPM-GC also detects some regions not present in the reference result, those can be found in the loose result provided by the neurologist. It should be noticed that those regions represent a less confident result given its colormap intensity.
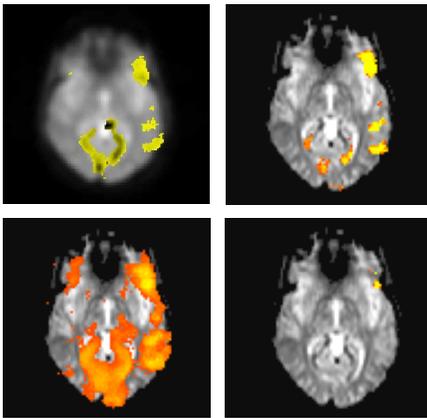

Fig. 4. Real data activated regions of a verb generation paradigm: *up, left* - Result of the parameter free SPM-GC algorithm.; *up, right* - Reference result given by the SPM-GLM algorithm.; *down, left* - Loose result given by the SPM-GLM algorithm.; *up, right* - Restricted result given by the SPM-GLM algorithm.;
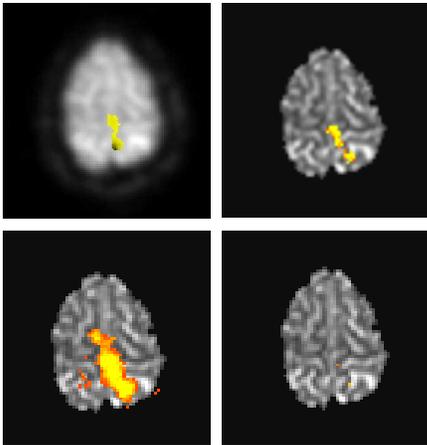

Fig. 5. Real data activated regions of a motor paradigm: *up, left* - Result of the parameter free SPM-GC algorithm.; *up, right* - Reference result given by the SPM-GLM algorithm.; *down, left* - Loose result given by the SPM-GLM algorithm.; *up, right* - Restricted result given by the SPM-GLM algorithm.;

IV. CONCLUSIONS

In this paper an algorithm parameter free to detect brain activated areas in fMRI is described. The traditional estimation and inference

steps are joint together where the *explanatory variables* (EVs) are considered binary and jointly estimated with the *hemodynamic response function* (HRF) in a space varying basis. The detection problem is formulated in a Bayesian framework where an energy function is minimized and where a physiological based prior for the HRF is used in order to force its smoothness. The estimation of the binary EVs is performed by using a *Graph-Cuts* based thresholding algorithm that takes into account the spatial correlation among neighbors in order to remove spurious activation foci generated by the noise which reduces the error probability. Monte Carlo tests with synthetic data are presented to characterize the robustness of the algorithm in terms of error probability. Examples using real data are also presented and the results obtained with the proposed algorithm are compared with the ones obtained with the *Brainvoyager* commercial software. These comparisons shows that the proposed algorithm leads to similar results obtain with the *Brainvoyager* software without need of any parameter tunned by the medical doctor, removing the subjective nature of the results.

REFERENCES

[1] P. Jezzard, P. M. Matthews, and S. M. Smith, *Functional magnetic resonance imaging: An introduction to methods.* Oxford Medical Publications, 2006.

[2] K. J. Friston, "Analyzing brain images: Principles and overview," in *Human Brain Function*, R.S.J. Frackowiak and K.J. Friston and C. Frith and R. Dolan and J.C. Mazziotta, Ed. Academic Press USA, 1997, pp. 25–41.

[3] K. J. Friston and A. P. Holmes and K. J. Worsley and J. B. Poline and C. Frith and R. S. J. Frackowiak, "Statistical Parametric Maps in Functional Imaging: A General Linear Approach," *Human Brain Mapping*, vol. 2, pp. 189–210, 1995.

[4] R. Baumgartner, L. Ryner, W. Richter, R. Summers, M. Jarmasz, and R. Somorjai, "Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis," *Magn Reson Imaging*, vol. 18, no. 1, pp. 89–94, Jan 2000, comparative Study.

[5] F. A. Nielsen, "Bibliography on independent component analysis in functional neuroimaging," 2007. [Online]. Available: http://www2.imm.dtu.dk/ ˜ fn/bib/Nielsen2001BibICA/Nielsen2001BibICA.html

[6] K. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: Theory," *NeuroImage*, vol. 16, pp. 465–483, 2002.

[7] W. Penny, N. Trujillo-Barreto, and K. Friston, "Bayesian fMRI time series analysis with spatial priors," *NeuroImage*, vol. 24, no. 2, pp. 350–362, 2005.

[8] Y. Lu, A. P. Bagshaw, C. Grova, E. Kobayashi, F. Dubeau, and J. Gotman, "Using voxel-specific hemodynamic response function in eeg-fmri data analysis," *NeuroImage*, vol. 32, no. 1, pp. 238–247, August 2006.

[9] G. K. Aguirre, E. Zarahn, and M. D'esposito, "The variability of human, BOLD hemodynamic responses," *Neuroimage*, vol. 8, no. 4, pp. 360–369, Nov 1998, clinical Trial.

[10] P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, C. Pallier, and H. Benali, "Unsupervised robust non-parametric estimation of the hemodynamic response function for any fmri experiment," *IEEE Trans. Med. Imaging*, vol. 22, no. 10, pp. 1235–1251, 2003.

[11] G. Marrelec, H. Benali, P. Ciuciu, M. Pélégrini-Issac, and J. B. Poline, "Robust bayesian estimation of the hemodynamic response function in event-related bold fmri using basic physiological information." *Hum Brain Mapp*, vol. 19, no. 1, pp. 1–17, May 2003. [Online]. Available: http://dx.doi.org/10.1002/hbm.10100

[12] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, 2001.

[13] "Brainvoyager software." [Online]. Available: http://www.brainvoyager.com/

[14] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing.* Prentice-Hall, 2000.

[15] V. Kolmogorov and R. Zabih, "What energy functions can be minimizedvia graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, 2004.