

# Group MRF for fMRI Activation Detection

Bernard Ng, Rafeef Abugharbieh  
The University of British Columbia  
2329 West Mall, Vancouver, BC V6T 1Z4  
bernardn@ece.ubc.ca, rafeef@ece.ubc.ca  
<http://bisicl.ece.ubc.ca>

Ghassan Hamarneh  
Simon Fraser University  
8888 University Drive, Burnaby, BC V5A 1S6  
hamarneh@cs.sfu.ca  
<http://mial.fas.sfu.ca>

## Abstract

*Noise confounds present serious complications to accurate data analysis in functional magnetic resonance imaging (fMRI). Simply relying on contextual image information often results in unsatisfactory segmentation of active brain regions. To remedy this, we propose a novel Group Markov Random Field (Group MRF) that extends the neighborhood system to other subjects to incorporate group information in modeling each subject's brain activation. Our approach has the distinct advantage of being able to regularize the states of both intra- and inter-subject neighbors without having to create a stringent one-to-one voxel correspondence as in standard fMRI group analysis. Also, our method can be efficiently implemented as a single MRF, hence enabling activation maps of a group of subjects to be simultaneously and collaboratively segmented. We validate on both synthetic and real fMRI data and demonstrate superior performance over standard analysis techniques.*

## 1. Introduction

Image noise remains a major challenge to accurate segmentation. In scenarios with only a moderate level of sparsely distributed noise, regularization approaches that exploit contextual image information, such as Markov Random Fields (MRF), typically perform well in segmenting the objects of interest [1]. However, if the majority of image voxels are corrupted by strong noise, MRFs may not be effective due to the lack of reliable neighbors. The heart of the problem is that, in certain cases, there may simply be insufficient information within a single image to enable reliable segmentation. Additional information is thus needed in these situations. If multiple images containing objects of the same class are available, one can exploit common features shared among the images to enhance the segmentation of each image. This general idea of incorporating group information is widely used in many areas including recognition [2], registration [3], and reconstruction [4]. A popular approach is to build a model

or template from a set of training images that is then used to segment new images [2]. Alternatively, one may use multiple templates to generate candidate labels that are subsequently reconciled using e.g. majority voting to fuse these labels into a single segmentation [5]. Generating representative models, however, can be quite difficult for certain applications, such as medical imaging where even experts oftentimes cannot provide consistent, accurate ground truth segmentations. Under such circumstances, unsupervised approaches are needed.

A few recent unsupervised approaches have alluded to the importance of aggregating information over a set of candidate segmentations. A notable approach called “ensemble clustering” that combines outputs from multiple clustering algorithms has shown superior performance over using each algorithm separately [6]. In [7], Ward and Hamarneh proposed extracting skeletons from a *group* of images and applied majority voting to prune each skeleton with robust results demonstrated. In this work, we extend this fundamental idea of drawing consensus across images to the analysis of functional magnetic resonance imaging (fMRI) data, where the highly complex noise structure and the lack of ground truth segmentation render group-wise unsupervised approach particularly attractive.

In a typical fMRI experiment, multiple subjects (usually 10-15) are recruited to perform a certain task, e.g. finger tapping, while three dimensional (3D) MRI scans of their brains are acquired at regular time intervals. Neurons within brain regions involved will fire in response to task stimulus, which alters the oxygenation level in nearby tissues, resulting in an intensity change in the MR signal. This induced signal contrast is referred to as blood oxygen level dependent (BOLD) contrast, which is widely used as a basis for inferring brain activation. However, confounds such as scanner noise, head motion, and oxygen level changes arising from the cardiac and respiratory cycles have perplexing effects on the BOLD signal, which greatly complicates accurate analysis of fMRI data.

The standard way of analyzing fMRI data involves examining the intensity changes of each voxel over time and statistically comparing each voxel's intensity time course against a hypothesized response. A general linear model (GLM) is typically used for this purpose, where statistics reflecting the degree of similarity between the

stimulus and voxel time courses are generated and assembled into an activation statistics map [8]. A threshold is then applied to identify the activated voxels. Since each voxel is analyzed independently, voxel interactions are ignored despite that each voxel is unlikely to function in isolation. Therefore, some have proposed modeling activation statistics maps as MRFs to incorporate contextual image information [9], while others employed Bayesian approaches to directly integrate neighborhood information into the activation statistics estimates [10]. Encouraging neighboring voxels to have similar labels (e.g. active or non-active) helps reduce the number of isolated voxels falsely declared as active. However, solely relying on contextual image information is often inadequate to deal with the inherently low signal-to-noise (SNR) of fMRI data. Thus, pooling additional information from other subjects may prove beneficial [11].

In most fMRI studies, the standard approach for aggregating information across subjects is to first non-rigidly register the anatomical image of each subject onto a common brain template with the same transform applied to the activation statistics map. Statistical testing is then applied to activation statistics assembled across subjects to delineate the commonly active voxels. The resulting group map is then taken as representative of all subjects. The implicit assumption is that a one-to-one correspondence exists between the active voxels across subjects and that the voxels are perfectly aligned in the template space. However, due to the large anatomical inter-subject variability, rarely can the anatomical images be perfectly aligned to the template. Hence, registration errors pose major limitations to the standard analysis approach. In fact, even if one manages to align the anatomical images perfectly, whether a one-to-one functional correspondence exists between voxels is debatable. The considerable functional inter-subject variability observed in past studies [12] suggests that such voxel correspondence does not likely exist. Nevertheless, active voxels within brain regions involved with the experimental tasks are often consistently found across subjects, though the exact locations are usually different [12]. The question thus becomes whether the observed functional variability is due to noise or true inter-subject differences. If the observed variability mainly arises from noise, incorporating group information to segment each subject’s activation statistics map may better reveal inter-subject commonalities.

In this paper, we propose exploiting group information in fMRI analysis by modeling the activation map of each subject as an MRF with the neighborhood system extended to other subjects. Our approach thus jointly accounts for the states of both intra- and inter-subject neighbors in estimating the state of each voxel. Also, by incorporating group information in segmenting each subject’s activation map, as opposed to estimating a group map, both inter-subject commonalities as well as differences can be

studied. In our proposed formulation, MRFs of all subjects can be jointly and cooperatively solved as a single MRF. We thus refer to our proposed method as ‘Group MRF’.

## 2. Proposed method

Our proposed method mimics a neurologist’s expertise deployed in deciding if a voxel is active. Specifically, to determine if voxel  $j$  in subject  $i$  is active, one may first examine the spatial neighbors of voxel  $j$ . However, if the neighboring voxels are found to be too noisy or unreliable, one may have to consider voxels of other subjects located in proximity to voxel  $j$ . Since only voxels proximal to each other are hypothesized to be in a similar state, the stringent requirement for a one-to-one voxel correspondence is alleviated. Instead, a relaxed condition of requiring only the brain structures of the subjects to be approximately aligned is in place. Based on this intuition, we propose to model each subject’s activation statistics map as an MRF whose neighborhood system extends to other subjects (Fig. 1). Since the vast anatomical variability in real fMRI data renders accurate whole-brain registration difficult, we employ an alternative approach, where we define regions of interest (ROIs) and perform alignment at the regional level. This approach ensures that no brain structures will be mistakenly declared as part of another structure as often encountered with whole-brain registration.

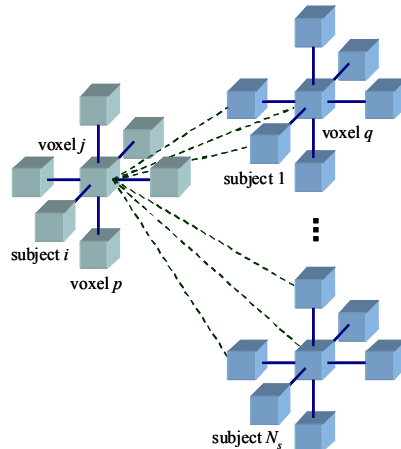


Fig. 1. Proposed Group MRF. Each subject’s ROI activation map is modeled as an MRF with its neighborhood system extended to other subjects to incorporate group information.

### 2.1. Group MRF

Let  $G_i = (V_i, E_i)$  be a graph, where  $E_i = \varepsilon_i \cup \varepsilon_{-i}$ ,  $V_i$  and  $\varepsilon_i$  are the sets of voxels and edges within subject  $i$ ’s ROI, and  $\varepsilon_{-i}$  is the set of edges between subject  $i$  and other subjects. Also, let  $jp \in \varepsilon_i$  be the intra-subject edge between voxels  $j$  and  $p$  and  $jq \in \varepsilon_{-i}$  be the inter-subject edge between voxel  $j$  and  $q$ . Delineation of subject  $i$ ’s ROI activation statistics map can be formulated as the following MRF energy minimization problem:

$$E_i(x^m) = \sum_{j \in \mathcal{V}_i} \theta_j^m(x_j^m) + \alpha \left( \sum_{jp \in \mathcal{E}_i} \theta_{jp}^m(x_j^m, x_p^m) + \gamma \sum_{jq \in \mathcal{E}_{-i}} \theta_{jq}^m(x_j^m, x_q^m) \right), \quad (1)$$

where  $x_r^m \in L = \{1 \dots K\}$  is the label assigned to voxel  $r$ ,  $m$  indicates the experimental task,  $\theta_j^m(x_j^m)$  denotes a unary potential, and  $\theta_{jp}^m(x_j^m, x_p^m)$  and  $\theta_{jq}^m(x_j^m, x_q^m)$  are intra- and inter-subject pairwise potentials.  $\alpha$  is used to adjust the contribution of the pairwise potentials while  $\gamma$  balances the contribution of intra- and inter-subject pairwise potentials.

## 2.2. Unary potential

Let  $t_j^m$  be the activation statistic of voxel  $j$ . If voxel  $j$  has a low probability of being activated during task  $m$ , we penalize labeling  $x_j$  as active, and vice versa. Our unary potential is thus defined as:

$$\theta_j^m(x_j^m = k) = 1 - p(x_j^m = k | t_j^m), \quad (2)$$

where  $p(x_j^m = k | t_j^m)$  is the probability of  $x_j^m$  being assigned label  $k$  given  $t_j^m$ , estimated using a constrained Gaussian mixture model (CGMM) [13], as detailed in Section 2.4.

## 2.3. Pairwise potentials

The proposed extended neighborhood system consists of an intra- and an inter-subject component. For the *intra-subject* component, we use the Pott's model potential:

$$\theta_{jp}^m(x_j^m, x_p^m) = 1 - \delta(x_j^m - x_p^m), \quad (3)$$

where  $\delta(y) = 1$  if  $y = 0$  and 0 otherwise, and edges are added between the 4- and 6-connected spatial neighbors for the 2D and 3D case, respectively. For the *inter-subject* component, we also use the Pott's model potential:

$$\theta_{jq}^m(x_j^m, x_q^m) = 1 - \delta(x_j^m - x_q^m). \quad (4)$$

Edges are added between voxel  $j$  of subject  $i$  and its  $c$  nearest inter-subject neighbors for every subject pair.

## 2.4. Algorithm implementation

The proposed method can be implemented as a single MRF by treating the voxels of all subjects as a single set and adding edges within and between all subjects in the manner described in Section 2.3. Since we are labeling the voxels as either active or non-active (i.e.  $K = 2$ ) as in [13], globally optimal labels for this Group MRF's energy function (5) can be estimated using binary graph cut [14].

$$E = \sum_{i=1}^{N_s} \sum_{j \in \mathcal{V}_i} \theta_j^m(x_j^m) + \alpha \left( \sum_{jp \in \mathcal{E}_i} \theta_{jp}^m(x_j^m, x_p^m) + 0.5\gamma \sum_{jq \in \mathcal{E}_{-i}} \theta_{jq}^m(x_j^m, x_q^m) \right), \quad (5)$$

where  $N_s$  is the number of subjects and 0.5 is introduced to avoid penalizing each inter-subject edge twice. By using

this implementation strategy, activation statistics maps of all subjects can be jointly and collaboratively segmented.

We set  $\alpha$  in (5) to  $1/N_s$  to scale the cumulative contribution of the subjects to a level similar to that of a single subject, and thus ensures that the pairwise potentials will not dominate over the unary potential. The choice of  $\gamma$  depends on the degree of inter-subject variability, which is unknown a priori. We have thus set  $\gamma$  to 0.5 to weight subject  $i$ 's own information more than that of other subjects. Nevertheless, varying  $\gamma$  by  $\pm 0.2$  (i.e.  $\pm 40\%$ ) did not significantly affect the results. To compute  $p(x_j^m = k | t_j^m)$  in (2), we first estimate  $t_j^m$  using a GLM [8]:

$$y_j = X\beta_j + \omega_j, \quad (6)$$

$$t_j^m = \beta_j^m / se(\beta_j^m), \quad (7)$$

where  $y_j$  is the intensity time course of voxel  $j$ ,  $\omega_j$  is assumed to be white Gaussian noise,  $\beta_j$  is a vector containing the estimated effects  $\beta_j^m$  for the various tasks, and  $se(\beta_j^m)$  is the standard error of  $\beta_j^m$  [15].  $X$  is a design matrix with boxcar functions (time-locked to task stimuli) convolved with the hemodynamic response (HDR) as regressors [8]. Based on  $t_j^m$ ,  $p(x_j^m = k | t_j^m)$  can be estimated using a CGMM [13], as shown graphically in Fig. 2.

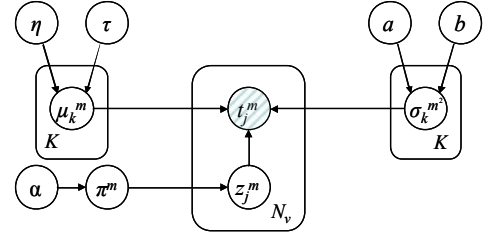


Fig. 2. Graphical depiction of constrained GMM for computing the unary potential  $\theta_j^m(x_j^m)$ .

In CGMM,  $t_j^m$  is assumed to be generated from a mixture of  $K$  Gaussian distributions with mixing coefficients  $\pi_k^m$ , means  $\mu_k^m$ , and variance  $\sigma_k^{m^2}$ . Conjugate priors are used to constrain the model parameters  $\mu_k^m$ ,  $\sigma_k^{m^2}$ , and  $\pi_k^m$ :

$$t_j^m \sim \sum_{k=1}^K \pi_k^m N(\mu_k^m, \sigma_k^{m^2}), \quad (8)$$

$$\mu_k^m \sim N(\eta, \tau^2), \quad (9)$$

$$\sigma_k^{m^2} \sim IG(a, b), \quad (10)$$

$$\pi^m \sim Dir(\alpha), \quad (11)$$

where  $IG(a, b)$  and  $Dir(\alpha)$  denote inverse Gamma and Dirichlet distributions. Adding priors mitigates the singularity problem in maximum likelihood solutions, i.e.  $\mu_k^m$  collapsing onto  $t_j^m$  with  $\sigma_k^{m^2}$  approaching 0, thus assigning infinite probability at  $t_j^m$ . Also, having priors allows us to encode our knowledge of  $\mu_k^m$  into the model. Specifically, we know that  $t_j^m$  of non-active voxels should theoretically be 0 and the threshold used for delineating the active voxels given  $t_j^m$  is roughly between 3 and 4 based on

Gaussian Random Field (GRF) theory [8]. This prior knowledge is encoded into  $\eta$  with  $\tau^2$  set to 1. As for  $\sigma_k^{m^2}$ , we use an uninformative prior by setting both  $a$  and  $b$  to 0.5 [13], since little is known about this parameter.  $\alpha$  is set to  $1/K$  assuming equal prior probabilities of being assigned any one of the  $K$  labels. Gibbs sampling is employed to compute  $p(x_j^m = k | t_j^m)$  [13]:

$$p(\pi^m | z^m, \alpha) = \text{Dir}\left(\alpha_1 + \sum_{j=1}^{N_V} z_j^{m,1}, \dots, \alpha_K + \sum_{j=1}^{N_V} z_j^{m,K}\right), \quad (12)$$

$$p(\mu_k^m | t^m, z^m) = \text{N}\left(\frac{\sum_{j=1}^{N_V} z_j^{m,k} t_j^m \tau^2 + \eta \sigma_k^{m^2}}{\tau^2 \sum_{j=1}^{N_V} z_j^{m,k} + \sigma_k^{m^2}}, \frac{\tau^2 \sigma_k^{m^2}}{\tau^2 \sum_{j=1}^{N_V} z_j^{m,k} + \sigma_k^{m^2}}\right), \quad (13)$$

$$p(\sigma_k^{m^2} | t^m, z^m, a, b) = \text{IG}\left(a + 0.5 \sum_{j=1}^{N_V} z_j^{m,k}, b + 0.5 \sum_{j=1}^{N_V} z_j^{m,k} (t_j^m - \mu_k^m)^2\right), \quad (14)$$

$$p(z_j^m = k | t_j^m, \mu_k^m, \sigma_k^{m^2}, \pi_k^m) \propto \pi_k^m \text{N}(\mu_k^m, \sigma_k^{m^2}), \quad (15)$$

where  $N_V$  is the number of voxels in subject  $i$ 's ROI and  $z_j^{m,k}$  is an indicator variable set to 1 if voxel  $j$  during task  $m$  is estimated to be in state  $k$ , and 0 otherwise. For the inter-subject pairwise potential, we have empirically set  $c$  to 3, but increasing  $c$  to 4 or 5 had little effects on the results.

### 3. Synthetic data experiments

To validate our proposed method, we generated 500 synthetic datasets, each consisting of ten subjects. The activation pattern comprised three clusters (Fig. 3). The signal intensity of the active voxels (marked with circles in Fig. 3) was set to decrease exponentially as a function of distance from the respective activation centroid of each cluster. The smaller cluster was added to test whether Group MRF can detect spatially isolated active voxels that are *consistently present* across subjects. The synthetic time courses of active voxels were generated by convolving a box-car function, having the same stimulus timing as in our experiment (Section 4.1), with a canonical HDR [8] and adding Gaussian noise. We introduced additional inter-subject variability to emulate two potential cases. In the first case, which we refer to as the common cluster location (CCL) case (Fig. 3(a)-(d)), the location of the clusters were fixed but the placement of the activation centroids (i.e. location away from which the signal of the active voxels decrease) were varied across subjects and permitted to be anywhere within the clusters. This case emulates the situation where the active regions completely overlap between subjects but the location at which the BOLD signal concentrates varies. Hence, it will *appear* that there is little overlap between the active regions of the subjects. In the second case, which we refer to as the

varying cluster location (VCL) case (Fig. 3(e) & (f)), we randomly varied the locations of the two larger clusters across subjects. This case models the situation where the active regions of the subjects do not completely overlap. The maximum cluster misalignment was set to two voxels ( $\sim 30\%$  of the cluster width) in both the vertical and horizontal directions. The maximum SNR was set as 0.5 for both cases in Fig. 3, i.e. voxels near the activation centroids had an SNR close to 0.5, whereas voxels away from the activation centroids had an SNR below 0.5.

For comparisons, we also tested the following methods: (i) GLM with Gaussian smoothing and a threshold based on GRF theory for a p-value of 0.05 [8], (ii) MRF separately applied to each subject's  $t_j^m$  map, and (iii) second level GLM (i.e. substituting spatially smoothed  $\beta_j^m$  of all subjects into  $y_j$  in (6) and setting  $X$  as a column of ones) with a GRF threshold [8]. We refer to methods (i), (ii), and (iii) as individual GLM (iGLM), individual MRF (iMRF), and group GLM (gGLM), respectively.

#### 3.1. Common cluster location

Fig. 3(a)-(d) shows the qualitative results for the CCL case. Only results from four of the ten subjects in one of the synthetic datasets are plotted due to space limitations. Using iGLM missed most of the active voxels away from the activation centroids, whereas using iMRF resulted in more active voxels identified, although active voxels with  $t_j^m$  below 3 were largely undetected. Also, the smaller cluster was missed due to lack of intra-subject active neighbors. Using gGLM detected more mildly activated voxels, but the spatial smoothing required for using GRF threshold resulted in many false positives. Using Group MRF detected most of the active voxels, including those with  $t_j^m$  well below 3. Also, Group MRF consistently detected the smaller cluster in all subjects. To quantify the performance, we computed the average Dice similarity coefficient (DSC) over the 500 synthetic datasets (Fig. 4).

$$\text{DSC} = \frac{2TP}{2TP + FP + FN}, \quad (16)$$

where TP, FP, and FN denote the number of true positives, false positives, and false negatives, respectively.

As evident from the low DSC for iGLM and iMRF, a single subject's information may be insufficient to obtain satisfactory segmentation at low SNR, even when contextual information is considered. This result confirms our argument for incorporating additional information. Solely relying on group information, however, can also be problematic as apparent from the gGLM results, where increasing SNR reduced DSC. This counter-intuitive finding can be explained as follows. Since the locations of the activation centroids are randomly varied across subjects,  $\beta_j^m$  at each voxel location will thus be different between subjects. Considering that the range of  $\beta_j^m$  values is governed by the SNR, increasing SNR will increase

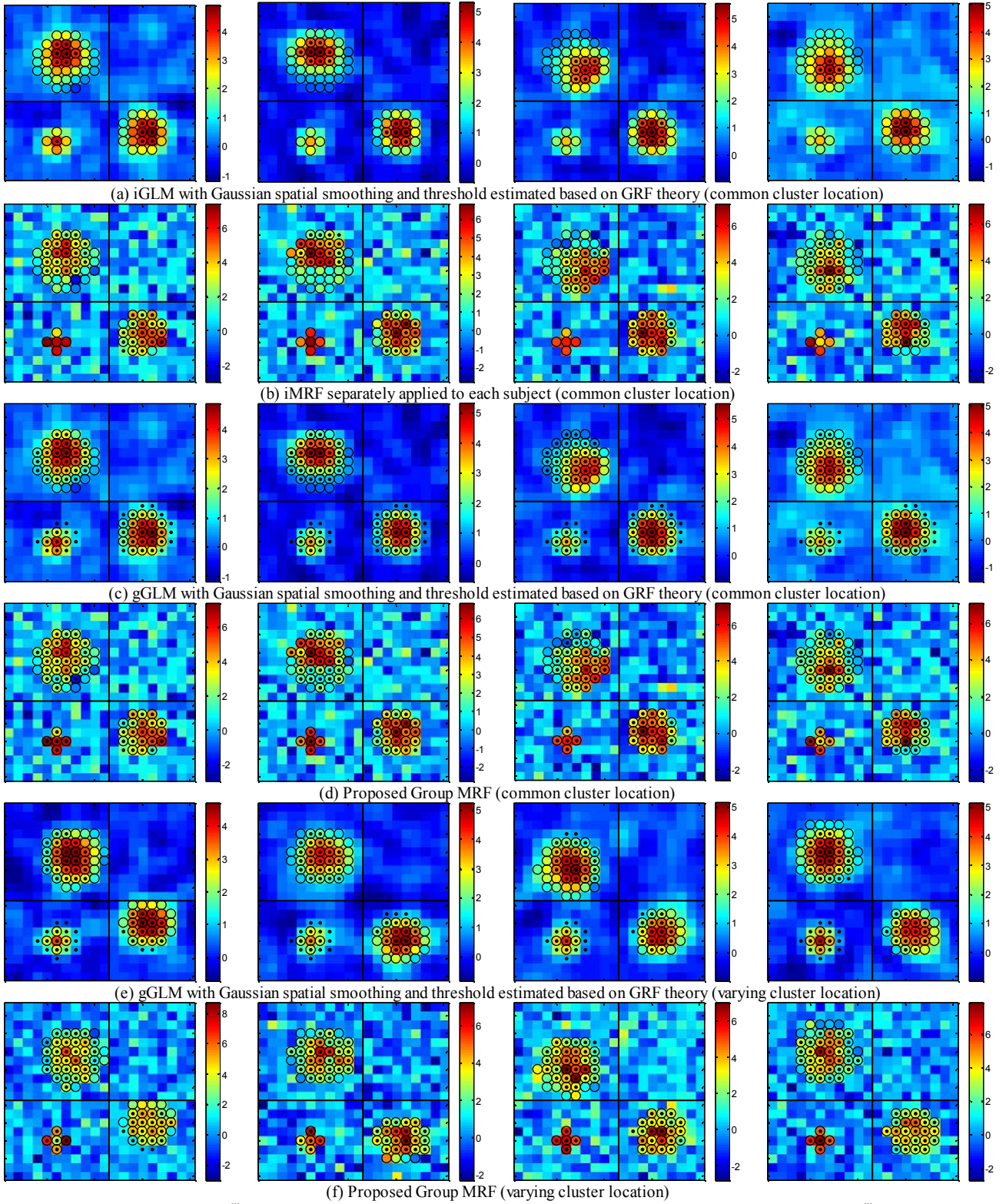


Fig. 3. Synthetic data results.  $t_j^m$  plotted with dots and circles indicating detected and ground truth active voxels.  $t_j^m$  maps along each column correspond to the same subject's activation map for each test case, but appears different due to spatial smoothing employed in GLM. For the CCL case, Group MRF (d) detected the majority of active voxels, whereas iGLM (a) and iMRF (b) failed to obtain such delineation. gGLM (c) also detected most of the active voxels, but included many false positives. For the VCL case, Group MRF (f) was able to adapt to the cluster variations, whereas the single group map approach of gGLM (e) does not facilitate such adaptation.

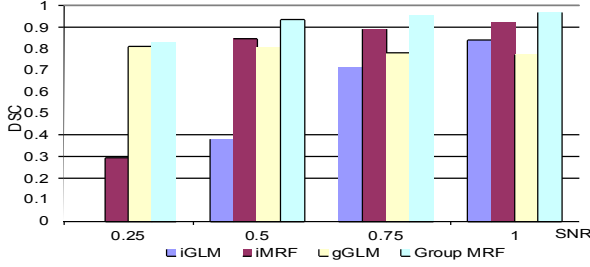


Fig. 4. DSC of the four contrasted methods for the CCL case at various SNR levels. Group MRF consistently outperformed all the other methods. Note: DSC of iGLM = 0 at SNR = 0.25.

inter-subject variability in  $\beta_j^m$ , which decreases  $TP$ , hence a reduction in DSC. In contrast, using Group MRF resulted in higher DSC for all SNR levels compared to the other methods. Also, unlike gGLM, the DSC of Group MRF increased with increasing SNR. The reason is that Group MRF does not directly compare  $t_j^m$  across subjects. Instead, group information is integrated by first assigning a label to each voxel and then estimating the state of a voxel based on consensus of labels between the intra- and inter-subject neighbors. Note that at higher SNR where more reliable information is available within each subject, adding group information still improved results over using iGLM and iMRF alone. Moreover, increasing the number of subjects increased DSC as shown in Fig. 5.

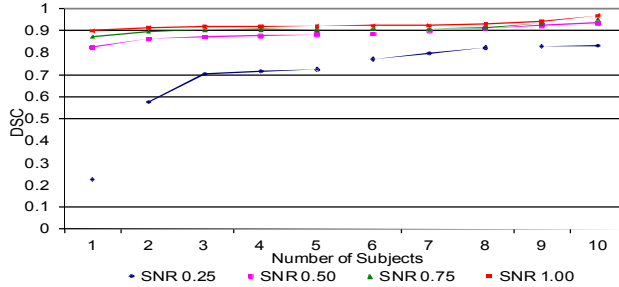


Fig. 5. DSC vs. number of subjects for the CCL synthetic test using Group MRF at varying levels of SNR. Increasing the number of subjects consistently increased DSC.

### 3.2. Varying cluster location

Qualitative results for the VCL test case are shown in Fig. 3(e)-(f). iGLM and iMRF resulted in similar performance as in the CCL case, since neither method depends on group information. Therefore, we omitted these results from Fig. 3. Using gGLM, many active voxels near the cluster borders were missed due to cluster misalignments. In contrast, Group MRF was able to adapt to the inter-subject differences with more active voxels identified and fewer false positives declared. This adaptability arises from the fact that, unlike gGLM, Group MRF does not solely rely on group consensus but also uses subject-specific information to segment each subject's  $t_j^m$  map. Thus, the more pronounced inter-subject differences were preserved. Quantitative results show that Group MRF again

outperformed all other methods (Fig. 6). We note that increasing the number of subjects did not increase DSC initially (Fig. 7), since when the number of subject is low, it is difficult to draw consensus with the added cluster misalignments. Nevertheless, after adding seven subjects, DSC began and continued to increase, thus demonstrating Group MRF's robustness to inter-subject variability.

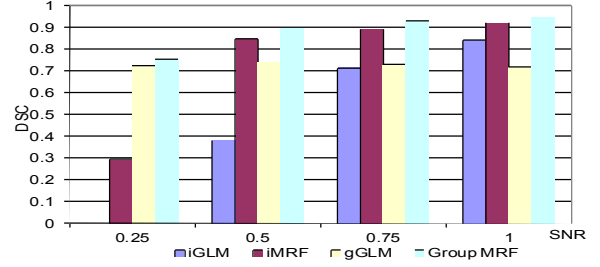


Fig. 6. DSC of the four contrasted methods for the VCL case at various SNR levels. Group MRF again outperformed all other methods. Note: DSC of iGLM = 0 at SNR = 0.25.

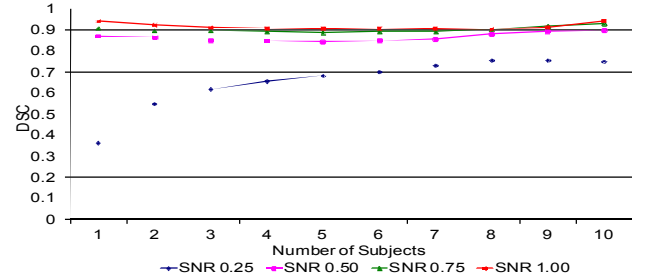


Fig. 7. DSC vs. number of subjects for VCL case using Group MRF. DSC decreased initially due to cluster misalignments, but eventually increased when enough subjects were deployed.

## 4. Real data analysis

### 4.1. Materials

After obtaining informed consent, fMRI data were collected from 10 healthy subjects (3 men, 7 women, mean age  $57.4 \pm 14$  years). Each subject used their right hand to squeeze a bulb with sufficient pressure to maintain a bar shown on a screen within an undulating pathway. The pathway remained straight during baseline periods and became sinusoidal at a frequency of 0.25 Hz (slow), 0.5 Hz (medium) or 0.75 Hz (fast) during time of stimulus. Each session lasted 260 s, alternating between baseline and stimulus of 20 s duration. Functional MRI was performed on a Philips Gyroscan Intera 3.0 T scanner (Philips, Best, Netherlands) equipped with a head-coil. T2\*-weighted images with BOLD contrast were acquired using an echo-planar (EPI) sequence with an echo time of 3.7 ms, a repetition time of 1985 ms, a flip angle of  $90^\circ$ , an in plane resolution of  $128 \times 128$  pixels, and a pixel size of  $1.9 \times 1.9$  mm. Each volume consisted of 36 axial slices of 3 mm thickness with a 1 mm gap. A 3D T1-weighted image consisting of 170 axial slices was further acquired to facilitate anatomical localization of activation. Slice timing



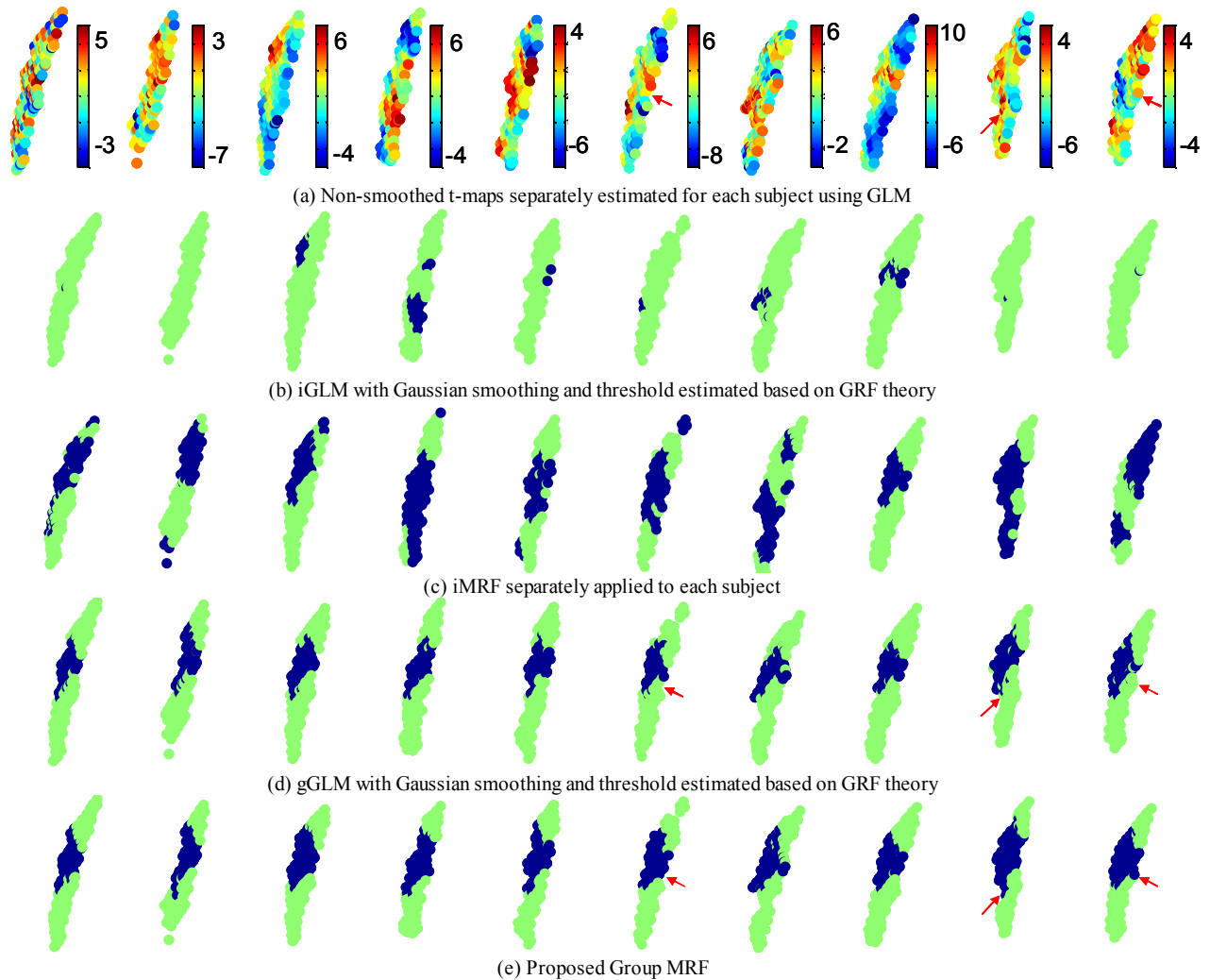


Fig. 9. Real data results for all 10 subjects. Blue in (b)-(d) indicates detected active voxels in the left M1. iGLM(b) failed to detect the hand region shown in Fig. 8, whereas iMRF(c) detected the hand region but also falsely declared adjacent areas as active. gGLM (d) detected the hand region, but neglected several adjacent voxels that were moderately active. Examples of such voxels (that can be easily identified in the ROI orientation shown) are indicated with a red arrow. Group MRF (e) correctly detected the hand region including the moderately active voxels neglected by gGLM, thus demonstrating Group MRF's ability to adapt to inter-subject differences.

- [8] K.J. Friston, A.P. Holmes, K.J. Worsley, J.B. Poline, C.D. Frith, and R.S.J. Frackowiak. Statistical Parametric Maps in Functional Imaging: A General Linear Approach. *Hum. Brain Mapp.*, 2(4):189-210, 1995.
- [9] X. Descombes, F. Kruggel, and D.Y. von Cramon. Spatio-Temporal fMRI Analysis Using Markov Random Fields. *IEEE Trans. Med. Imaging*, 17:1028-1039, 1998.
- [10] W.D. Penny, N.J. Trujillo-Barreto, and K.J. Friston. Bayesian fMRI Time Series Analysis with Spatial Priors. *Neuroimage*, 24:350-362, 2005.
- [11] D.R. Bathula, L.H. Staib, H.D. Tagare, X. Papademetris, R.T. Schultz, and J.S. Duncan. Multi-Group Functional MRI Analysis Using Statistical Activation Priors. In: *Proc. Med. Img. Computing and Computer Assit. Intervention fMRI Data Analysis Workshop*, 2009.
- [12] M.W.G. Vandenbroucke, R. Goekoop, E.J.J. Duschek, J.C. Netelenbos, J.P.A. Kuijer, F. Barkhof, Ph. Scheltens, and S.A.R.B. Rombouts. Interindividual Differences of Medial Temporal Lobe Activation during Encoding in an Elderly Population Studied by fMRI. *Neuroimage*, 21:173-180, 2004.
- [13] S. Makni, P. Ciuciu, J. Idier, and J.B. Poline. Joint Detection-Estimation of Brain Activity in Functional MRI: A Multichannel Deconvolution Solution. *IEEE Trans. Sig. Processing*, 53(9):3488-2502, 2005.
- [14] Y. Boykov, O. Veksler, and R. Zabih. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Trans. Pat. Anal. Mach. Intel.*, 23(11):1222-1239, 2001.
- [15] W.W. Daniel. *Biostatistics: A Foundation for Analysis in the Health Sciences*, 9<sup>th</sup> Ed. Wiley, 2008.
- [16] R. Liao, J.L. Krolik, and M.J. McKeown. An Information-theoretic Criterion for Intrasubject Alignment of fMRI Time Series: Motion Corrected Independent Component Analysis. *IEEE Trans. Med. Imaging*, 24(1):29-44, 2005.