Incorporating Spatial Priors into an Information Theoretic Approach for *f*MRI Data Analysis*

Junmo Kim[1], John W. Fisher III[1,2], Andy Tsai[1], Cindy Wible[3],
Alan S. Willsky[1], and William M. Wells III[2,3]

[1] Massachusetts Institute of Technology,
Laboratory for Information and Decision Systems,
Cambridge, MA, USA
{junmo, atsai, willsky}@mit.edu
[2] Massachusetts Institute of Technology,
Artificial Intelligence Laboratory,
Cambridge, MA, USA
{fisher, sw}@ai.mit.edu
[3] Harvard Medical School,
Brigham and Women's Hospital,
Department of Radiology,
Boston, MA, USA
cindy@bwh.harvard.edu

Abstract. In previous work a novel information-theoretic approach was introduced for calculating the activation map for *f*MRI analysis [Tsai *et al* , 1999]. In that work the use of mutual information as a measure of activation resulted in a nonparametric calculation of the activation map. Nonparametric approaches are attractive as the implicit assumptions are milder than the strong assumptions of popular approaches based on the general linear model popularized by Friston *et al* [1994]. Here we show that, in addition to the intuitive information-theoretic appeal, such an application of mutual information is equivalent to a hypothesis test when the underlying densities are unknown. Furthermore we incorporate local spatial priors using the well-known Ising model thereby dropping the implicit assumption that neighboring voxel time-series are independent. As a consequence of the hypothesis testing equivalence, calculation of the activation map with local spatial priors can be formulated as mincut/maxflow graph-cutting problem. Such problems can be solved in polynomial time by the Ford and Fulkerson method. Empirical results are presented on three *f*MRI datasets measuring motor, auditory, and visual cortex activation. Comparisons are made illustrating the differences between the proposed technique and one based on the general linear model.

---

* J. Kim, J. Fisher, A. Tsai, and A. Willsky supported in part by AFOSR grant F49620-98-1-0349, subcontract GC123919NGD from BU under the AFOSR Multidisciplinary Research Program, and ONR grant N00014-00-1-0089. C. Wible was supported in part under NIMH grants MH40799 and MH52807. W. Wells was supported in part by the Whitaker foundation and by NIH through grant 1P41RR13218.

1 Introduction

In previous work [6], we presented a novel information theoretic approach for calculating *f*MRI activation maps. The information-theoretic approach is appealing in that it is a principled methodology requiring few assumptions about the structure of the *f*MRI signal. In that approach, activation was quantified by measuring the mutual information (MI) between the protocol signal and the *f*MRI time-series at a given voxel. This measure is capable of detecting unknown nonlinear and higher-order statistical dependencies. Furthermore, it is relatively straightforward to implement.

In practice, activation decisions at each voxel are independent of neighboring voxels. Spurious responses are then removed by *ad hoc* techniques (e.g. morphological operators). In this paper, we describe an automatic maximum a posteriori (MAP) detection method where the well-known Ising model is used as a spatial prior. The Ising spatial prior does not assume that the time-series of neighboring voxels are independent of each other. Furthermore, removal of spurious responses is an implicit component of the detection formulation. In order to formulate the calculation of the activation map using this technique we first demonstrate that the information-theoretic approach has a natural interpretation in the hypothesis testing framework and that, specifically, our estimate of MI approximates the log-likelihood ratio of that hypothesis test. Consequently, the MAP detection problem using the Ising model can be formulated and solved *exactly* in polynomial time using the Ford and Fulkerson method [4].

We compare the results of our approach with and without spatial priors to an approach based on the general linear model (GLM) popularized by Friston *et al* [3]. We present results from three *f*MRI data sets. The data sets test motor, auditory, and visual cortex activation, respectively.

1.1 Review of the Information Theoretic Approach

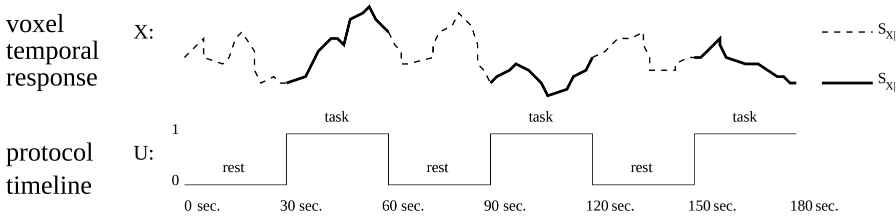


Fig. 1. Illustration of the protocol time-line, $S_{X|U=0}$, and $S_{X|U=1}$.

In [6], each voxel is declared to be active (or not) based solely only the temporal response of that voxel without considering the temporal responses of neighboring voxels. Let $X(\cdot,\cdot,\cdot,\cdot) = \{X(i,j,k,t)|1 \leq t \leq n\}$ denote the observed *f*MRI signal, where i, j, k are spatial coordinates and t is the time coordinate. Each voxel (i, j, k) has an associated discrete-time temporal response, $X_1, \ldots, X_n$, where X_t denotes $X(i, j, k, t)$ for convenience.

Figure 1 illustrates the protocol time-line and an associated temporal response. $S_{X|U=0}$ denotes the set of X_i's where the protocol is 0 while $S_{X|U=1}$ denotes the set of X_i's where the protocol is 1. An implicit assumption in our approach is that $S_{X|U=[0,1]}$ are i.i.d. according to $p_{X|U=[0,1]}(x)$. We treat the protocol U as a discrete random variable taking 0 and 1 with equal probability. In this case the MI between X and U is as follows:

$$\begin{aligned} I(X;U) &= H(U) - H(U|X) = h(X) - h(X|U) \\ &= h(X) - \frac{1}{2}h(X|U=0) - \frac{1}{2}h(X|U=1) \end{aligned}$$

where $H(U)$ is the discrete entropy of U and $h(X)$ is the continuous differential entropy of X. It can be shown that $H(U) \leq 1$ bit and that $0 \leq H(U|X) \leq H(U)$ consequently $0 \leq I(X;U) \leq 1$. Thus, MI as a measure between X and U is normalized.

The differential entropy of X

$$h(X) = -\int_S p_X(x) \log p_X(x) dx$$

is approximated as [6]:

$$\hat{h}(X) \approx -\frac{1}{n}\sum_{i=1}^{n} \log \hat{p}_X(x_i) \tag{1}$$

where $\hat{p}_X(X_i)$ is the Parzen density estimator [5], defined as

$$\hat{p}_X(x_i) = \frac{1}{n\sigma}\sum_j k\left(\frac{x_i - x_j}{\sigma}\right) \tag{2}$$

The kernel $k(x)$ must be a valid pdf (in our case a double exponential kernel).

2 Hypothesis Testing and MI

In order to extend our method to a MAP detection problem using spatial priors it is necessary to formulate a suitable hypothesis test and associated likelihood ratio. Here we show the equivalence of MI to the likelihood ratio of an underlying binary hypothesis testing problem.

2.1 Nonparametric Hypothesis Testing Problem

Consider the following hypothesis test

$$
\begin{aligned}
&H_0 : x_t, u_t \sim p_X(X)p_U(U) \qquad \text{, i.e. } X, U \text{are independent}\\
&H_1 : x_t, u_t \sim p_{X,U}(X, U) \qquad \text{, i.e. } X, U \text{are dependent}
\end{aligned}
$$

where the null hypothesis states that the protocol, U, and the fMRI time-series, X are statistically independent while the alternative states that they are not. The log-likelihood ratio, which is the optimal test statistic by the Neyman-Pearson lemma, is

$$
T_n = \sum_{t=1}^{n} \log \left(\frac{p_{XU}(x_t, u_t)}{p_X(x_t)p_U(u_t)} \right) \tag{3}
$$

assuming i.i.d. samples. It can be shown that [1]

$$
\begin{aligned}
\lim_{n\to\infty} T_n &= nI(X; U) &(4)\\
&= \mathrm{E}\left\{T_n\right\} &(5)
\end{aligned}
$$

consequently using $I(X; U)$ as the activation test statistic is equivalent to the aforementioned hypothesis test. Since the distribution of U is binomial with equal probability it can be shown that $I(X; U)$ simplifies to

$$
\begin{aligned}
I(X; U) &= h(X) - \frac{1}{2}h(X|U = 0) - \frac{1}{2}h(X|U = 1) &(6)\\
&= \frac{1}{2}D(p_{X|U=0}\|p_X) + \frac{1}{2}D(p_{X|U=1}\|p_X) &(7)
\end{aligned}
$$

where $D(p_1\|p_2)$ is the asymmetric Kullback-Leibler divergence.

2.2 Bias in the Estimate of the Likelihood Ratio

When evaluating the likelihood ratio we substitute 1 into 6. The conditional terms are summed over $S_{X|U=0}$ and $S_{X|U=1}$, respectively. The consequence is that we introduce bias into our estimate of the likelihood ratio or equivalently $I(X; U)$. In order to simplify matters consider the likelihood of the fMRI time-series

$$
\begin{aligned}
p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= \prod_{t=1}^{n} p_X(x_t) = \exp(\sum_t \log p_X(x_t))\\
&= \exp(\sum_t \log \hat{p}_X(x_t) + \sum_t \log \frac{p_X(x_t)}{\hat{p}_X(x_t)})\\
&= e^{-n[\hat{h}(X) + \frac{1}{n}\sum_t \log \frac{\hat{p}_X(x_t)}{p_X(x_t)}]} \approx e^{-n[\hat{h}(X) + D(\hat{p}_X\|p_X)]}
\end{aligned}
$$

In similar fashion the likelihood ratio is approximately

$$\frac{p_X(x_1,\ldots,x_n|H_1)}{p_X(x_1,\ldots,x_n|H_0)} \approx e^{n(\hat{I}(X;U)-\gamma)} \tag{8}$$

where $\gamma = \frac{1}{2}[D(\hat{p}_{X|U=0}\|p_{X|U=0}) + D(\hat{p}_{X|U=1}\|p_{X|U=1})] - D(\hat{p}_X\|p_X)$,which is nonnegative due to convexity of Kullback-Leibler divergence. More importantly, the divergence terms asymptotically approach zero. Consequently, the approximation approaches the true likelihood ratio.

3 Modeling Voxel Dependency via the Ising Model

We use the Ising model, a simple Markov random field (MRF), as a spatial prior of the binary activation map. Previously, Descombes *et al* [2] proposed the use an MRF (specifically a Potts model) for *f*MRI signal restoration and analysis. The substantive differences between the proposed method and that of Descombes *et al* include

- Descombes *et al* used simulated annealing to solve the MAP estimation problem with no guarantee of an exact result.
- The MRF prior model was combined with data fidelity terms in a heuristic way.

In contrast, our method combines the likelihood ratio obtained from the data with a spatial prior rigorously within the Bayesian framework leading to an exact solution of a binary MAP hypothesis test.

The Ising model captures the notion that neighboring voxels of an activated voxel are likely to be activated and vice versa. Specifically, let $y(i,j,k)$ be a binary activation map such that $y(i,j,k) = 1$ if voxel (i,j,k) is activated and 0, otherwise. Then this idea can be formulated using Ising model as a prior probability of the activation map $y(\cdot,\cdot,\cdot)$. Let

$$W = \left\{ w : w \in \{0,1\}^{N_1 \times N_2 \times N_3} \right\}$$

be the set of all possible [0-1] configurations and let $w(i,j,k)$ be an element of any one sample configuration.

The Ising prior on $y(\cdot,\cdot,\cdot)$ penalizes every occurrence of neighboring voxels with different activation states as follows:

$$P(y(\cdot,\cdot,\cdot) = w) = \frac{1}{Z}e^{-U(w)} \qquad Z = \sum_{w \in W} e^{-U(w)}$$

$$\begin{aligned} U(w) = \beta \sum_{i,j,k} (&w(i,j,k) \oplus w(i+1,j,k) + w(i,j,k) \oplus w(i,j+1,k) \\ &+ w(i,j,k) \oplus w(i,j,k+1)), \end{aligned}$$

where $\beta > 0$. As in [2] we assume that

$$p(X(\cdot,\cdot,\cdot,\cdot)|Y(\cdot,\cdot,\cdot)) = \prod_{i,j,k} p(X(i,j,k,\cdot)|Y(i,j,k))$$

That is, conditioned on the activation map, voxel time-series are independent. The MAP estimate of the activation is then

$$\begin{aligned}
\hat{y}(\cdot,\cdot,\cdot) &= \arg\max_{y(\cdot,\cdot,\cdot)} p_Y(y(\cdot,\cdot,\cdot)) p_{X|Y}(x(\cdot,\cdot,\cdot,\cdot)|Y(\cdot,\cdot,\cdot) = y(\cdot,\cdot,\cdot)) \\
&= \arg\max_{y(\cdot,\cdot,\cdot)} \log p_Y(y(\cdot,\cdot,\cdot)) + \\
&\quad \sum_{i,j,k} y(i,j,k) \log \frac{p_{X|Y}(x(i,j,k,\cdot)|Y(i,j,k)=1)}{p_{X|Y}(x(i,j,k,\cdot)|Y(i,j,k)=0)} \\
&= \arg\max_{y(\cdot,\cdot,\cdot)} \sum_{i,j,k} \lambda_{i,j,k} y(i,j,k) - \beta \sum_{i,j,k} \left(y(i,j,k) \oplus y(i+1,j,k) \right. \\
&\quad \left. + y(i,j,k) \oplus y(i,j+1,k) + y(i,j,k) \oplus y(i,j,k+1) \right),
\end{aligned}$$

where $\lambda_{i,j,k} = \ln \frac{p_{X|Y}(x(i,j,k,\cdot)|Y(i,j,k)=1)}{p_{X|Y}(x(i,j,k,\cdot)|Y(i,j,k)=0)} = n(\hat{I}_{i,j,k}(X;U) - \gamma)$ is the log-likelihood ratio at voxel (i,j,k) and $\hat{I}_{i,j,k}(X;U)$ is the MI estimated from time-series $X(i,j,k,\cdot)$. The previous use of MI as the activation statistic fits readily into the MAP formulation.

3.1 Exact Solution of the Binary MAP Estimation Problem

There are 2^{N_v} possible configurations of $y(\cdot,\cdot,\cdot)$ (or equivalently elements of the set W) where $N_v = N_1 N_2 N_3$ is the number of voxels. It has been shown by Greig *et al* [4] that this seemingly NP-complete problem can be solved *exactly* in polynomial time (order N_v). Greig *et al* accomplished this by demonstrating that under certain conditions the binary image MAP estimation problem (using MRFs as a prior) can be reduced to the minimum cut problem of network flow. Consequently, the methodology of Ford and Fulkerson for such problems can be applied directly. We are able to employ the same technique as a direct consequence of demonstrating the equivalence of MI to the log-likelihood ratio of a binary hypothesis testing problem. Details of the minimum cut solution are beyond the scope of this paper and we refer the reader to [4] for further details.

4 Experimental Results

We present experimental results on three *f*MRI data sets. The protocols are designed to activate the motor cortex (dominant hand movement protocol), auditory cortex (verb generation protocol), and visual cortex (visual stimulation with alternating checkerboard pattern), respectively. Each data set contains 60 whole brain acquisitions taken three seconds apart. We compare the resulting activation map computed by three methods: GLM, nonparametric MI, nonparametric MI with an Ising prior.

We first apply the GLM method to each data set. The coronal slice exhibiting the highest activation for each data set is shown in the first column of figure 2 with the GLM activation map overlaid in white for each data set. The F-statistic threshold for GLM was set such that the visual inspection of the activation map was consistent with our prior expectation of the number of activated voxels. This corresponded to a p-value of 10^{-10}.

In the next column of the figure the same slices are shown using MI to compute the activation map. In this case, the MI threshold γ was set such that all of the voxels detected by the GLM were detected by MI. Consequently figure 2 (b), (e) and (h) contain additional activations when compared to GLM. Some of these additional activations are spurious and some are not.

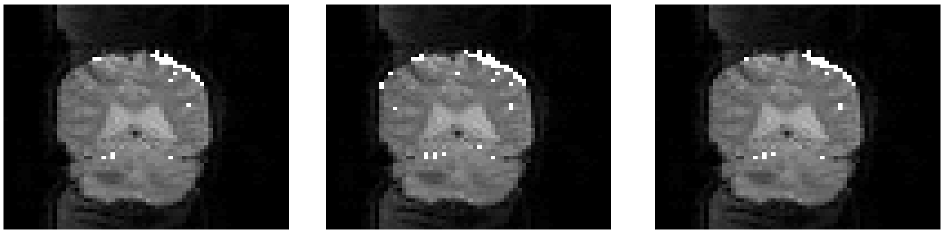
Finally, the Ising prior was applied to the MI activation map with $\beta = 1$. An intuitive understanding of the relationship of γ and β is as follows. If $\beta = 0$, then there is no prior and the method reduces to MI only. For $\beta \neq 0$ the interpretation is not so simple, but we can consider a special case. Suppose the neighbors of a voxel are declared to be active (in our case there are six neighbors for every voxel), then the effective MI activation threshold γ for that voxel has been reduced by $6\beta/n$. Conversely, if all of the neighbors are inactive then the effective threshold is increased by the same amount. For these experiments, $n = 60$ and $\beta = 1$, this equates to a 0.1 nat (equal to 0.14 bits) change in the MI activation threshold for the special cases described.

Comparison of figures 2 (b), (e) and (h) to figures 2 (c), (f), and (i) shows that many of the isolated activations were removed by the Ising prior, but some of the new MI activations remain. Figure 3 shows the temporal responses of the voxels with the lowest GLM score which were detected by MI with prior but not by GLM. Examination of these temporal responses (with protocol signal overlaid) reveals obvious structure related to the protocol signal.

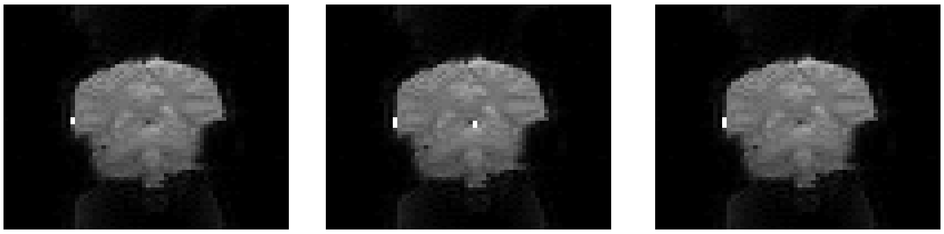
A reasonable question is whether this result is due to an unusually high threshold set for GLM. In order to address this we next lower the GLM threshold such that the voxels of figure 3 are detected by GLM. We then consider regions of the resulting activation map where new activations have appeared in figure 4. The activations of 4a and 4b (motor cortex, auditory cortex), would be considered spurious in light of the region in which they occur. The result for figure 4c is not so clear as these activations are most likely spurious, but might possibly be related to higher-ordered visual processing.

5 Conclusion

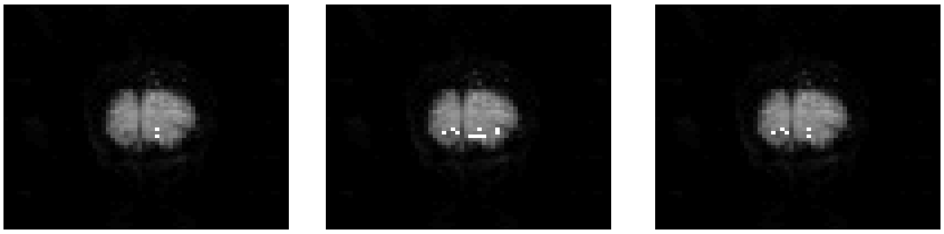
We demonstrated that our previous approach, derived from an information-theoretic perspective, can be formulated in a hypothesis testing framework. Furthermore, the resulting hypothesis test is free of many of the strong assumptions inherent in GLM. As *f*MRI is a relatively new modality for examining cognitive function we think that it is appropriate to examine nonparametric methods (i.e.



(a) GLM: 10th slice (b) MI: 10th slice; $\gamma = 0.50$ bits (c) MI with prior: 10th slice

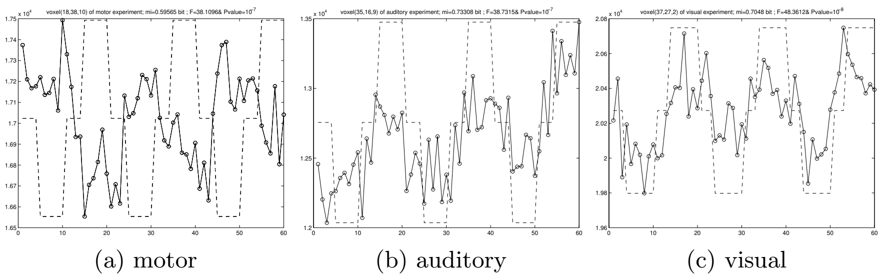


(d) GLM: 9th slice (e) MI: 9th slice; $\gamma = 0.58$ bits (f) MI with prior: 9th slice



(g) GLM: 2nd slice (h) MI: 2nd slice; $\gamma = 0.53$ bits (i) MI with prior: 2nd slice

Fig. 2. Comparison of *f*MRI Analysis results from motor, auditory and visual experiments



(a) motor (b) auditory (c) visual

Fig. 3. Temporal responses of voxels newly detected by the MI using Ising prior

GLM

(a) motor 14th slice (b) auditory 20th slice (c) visual 4th slice

MI with Ising prior