Analysis of Event-Related fMRI Data Using Diffusion Maps

Xilin Shen and François G. Meyer

University of Colorado at Boulder, Boulder CO 80309, USA

Abstract. The blood oxygen level-dependent (BOLD) signal in response to brief periods of stimulus can be detected using event-related functional magnetic resonance imaging (ER-fMRI). In this paper, we propose a new approach for the analysis of ER-fMRI data. We regard the time series as vectors in a high dimensional space (the dimension is the number of time samples). We believe that all activated times series share a common structure and all belong to a low dimensional manifold. On the other hand, we expect the background time series (after detrending) to form a cloud around the origin. We construct an embedding that reveals the organization of the data into an activated manifold and a cluster of non-activated time series. We use a graph partitioning technique–the normalized cut to find the separation between the activated manifold and the background time series. We have conducted several experiments with synthetic and in-vivo data that demonstrate the performance of our approach.

1 Introduction

The goal of functional neuroimaging is to map the activity of the brain in space and time. Functional magnetic resonance imaging (fMRI) has become one of the main tools for noninvasive assessment of human brain functions since its invention in the early 1990s. Event-related fMRI makes it possible to study the transient changes triggered by cognitive and sensory stimulation. Unlike block paradigm, event-related fMRI allows mixing of different task conditions on a trial-by-trial basis and provides a means of examining the dynamics and time-course of neural activity under various conditions.

However, the increase in the signal during an event-related fMRI experiment only lasts for a short period of time. The analysis is further complicated by the variation in the shape and amplitude of the hemodynamic response across different cortical regions [1]. Therefore, methods of analysis that rely on a specific model of the hemodynamic response will not be optimal. Several methods [2], [3] address this problem by considering a family of hemodynamic responses constructed from a set of basis functions. Others [4], [5] sought solutions via non-parametric data-driven methods.

Here, we regard the fMRI data as a very large set of time series $x_i(t)$ in $\mathbb{R}^T$, indexed by their position i. After removing the low frequency components

G.E. Christensen and M. Sonka (Eds.): IPMI 2005, LNCS 3565, pp. 652–663, 2005.

from the time series in the preprocessing, we assume any significant changes in the fMRI signal are related to the experimental paradigm. Although the experiment could recruit several cortical regions with different temporal responses, we restrict our attention to scenarios with no more than one type of temporal response. We now consider the set of all activated time series taken from the same cortical region. We assume that this set constitutes a manifold in $\mathbb{R}^T$. Clearly, if we were to use a parametric model for the hemodynamic response, the set of all hemodynamic responses generated from all the possible values of the parameters would form a manifold. The fact that the activated time series belong to a manifold in $\mathbb{R}^T$ has two implications. First, the activated time series reside only in a very small part of $\mathbb{R}^T$. Second, the activated time series are similar to each other, and one can go smoothly from one to the other one. In practice, the activated times series are corrupted by noise, and the associated manifold may exhibit some roughness.

A meaningful geometric description of the data in $\mathbb{R}^T$ would exhibit the presence of the activated manifold, and thus would have the power to discriminate between the activated and the non-activated time series. The diffusion maps method [6] is known to be capable of generating efficient representations of complex geometric structures. In particular, it can be applied to describe the geometry of a low dimensional manifold in high dimensions. In this work, we apply this technique to event-related fMRI data. The diffusion maps provide us with an embedding of the dataset. We then use a graph partitioning technique called the normalized cut [7] to separate the activated time series from the background time series. Because the normalized cut is closely related to the diffusion maps, it provides a natural method to perform the clustering of the time series. The paper is organized as follows. In the next section, we give a brief review to diffusion maps and the way we modify the graph construction. In section 3, we describe the normalized cut criterion and its relation to the diffusion maps. Results of experiments conducted on synthetic and in-vivo ER-fMRI data are presented in section 4.

2 Diffusion Maps and Graph Construction

The problem of finding meaningful structures and geometric descriptions of a data set has been the central interest in many areas like information retrieval, artificial intelligence and statistical data analysis. Different techniques have been developed to construct a representation for data lying on a low-dimensional manifold embedded in a high-dimensional space, among which are the classical linear methods including Principle Component Analysis (PCA) and Multidimensional Scaling (MDS), and the kernel methods like Local Linear Embedding (LLE) [8], Laplacian eigenmaps [9], and Hessian eigenmaps [10]. Most recently, Coifman and Lafon [6] have shown that all kernel methods are special cases of a general framework based on a diffusion process. By defining a random walk on the data set, they associate a Markov matrix to this data set. The spectral analysis of the Markov matrix provides a family of mappings which they termed "diffusion maps".

2.1 Diffusion Maps, Diffusion Distances

To construct the diffusion maps, we consider two time series $x(t)$ and $y(t)$ as being the nodes of a graph G. For two distinct nodes of the graph G we define a weight function $K(x, y)$ that measures the similarity between the time series according to

$$K(x,y) = \exp(-\frac{\|x-y\|^2}{\sigma}),$$

where $\|x-y\|^2 = \sum_t (x(t)-y(t))^2$. K is a symmetric, positive semi-definite matrix. We define the diffusion matrix: $A = D^{-1}K$ where D is a diagonal matrix with $d_{xx} = \sum_y k(x,y)$. By definition, this matrix is row-stochastic and can be viewed as the transition matrix of a random walk on the data set. Moreover, we can show that this random walk is reversible, which is the same to say that A is conjugate to a symmetric matrix:

$$\tilde{A} = D^{\frac{1}{2}} A D^{-\frac{1}{2}} \tag{1}$$

Then we define the diffusion distance between x and y at time m by:

$$D_m^2(x,y) = \tilde{a}^m(x,x) + \tilde{a}^m(y,y) - 2\tilde{a}^m(x,y) \tag{2}$$

$$= \sum_{j \geq 0} \lambda_j^{2m} (\phi_j(x) - \phi_j(y))^2 \tag{3}$$

where $\{\lambda_j\}$ and $\{\phi_j\}$ are the eigenvalues and eigenvectors of $\tilde{A}$, and $1 = \lambda_0 \geq \lambda_1 \geq \lambda_2 \geq ... \geq 0$. The quantity $\tilde{a}^m(x,y)$ denotes the entry of $\tilde{A}$ for row x and column y and it represents the probability of transition from x to y in m steps. The diffusion distance $D_m^2(x,y)$ is then a sum over all paths of length less than or equal to m between x and y. The value of $D_m^2(x,y)$ decreases when the number of paths between x and y gets larger. So it is a measure of connectivity of the points in the graph and it is robust to noise.
Last we introduce the family of diffusion maps $\{\Phi_m\}$ by

$$\Phi_m(x) = \begin{pmatrix} \lambda_0^m \phi_0(x) \\ \lambda_1^m \phi_1(x) \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \tag{4}$$

The diffusion maps convert diffusion distances into Euclidean distances:

$$\|\Phi_m(x) - \Phi_m(y)\|^2 = \sum_{j \geq 0} \lambda_j^{2m} (\phi_j(x) - \phi_j(y))^2 = D_m^2(x,y) \tag{5}$$

This embedding is related to the way a random walk propagates over the data set. In particular, it is sensitive to all sorts of bottleneck effects, therefore it is able to discover different clusters in the data set.

2.2 Modified Way of Constructing the Graph

Before applying the diffusion maps described in the previous section to fMRI data, we need to construct the adjacency graph. We consider each time series to be a node in the graph, and we put an edge between a pair of time series x and y if they are similar. Similarity can be defined by the Euclidean distance between the two time series, $\|x-y\|^2 = \sum_t (x(t)-y(t))^2$. We connect the two nodes x and y of the graph if x is among the n nearest neighbors of y, or y is among the n nearest neighbors of x. This approach is chosen for its simplicity and is guaranteed to give well connected graphs.

However, this definition of connectivity is not adapted to data set with low SNR, which is the case of fMRI data. Indeed, if the data are noisy, this approach will create many unfaithful connections between activated time series and non-activated time series, thereby severely obscuring the geometric structure of the data set. The idea can be better illustrated by the following computation. Suppose we have activated voxels i, j and non-activated voxel k. The time series from these voxels are given by:

$$\begin{aligned} x_i(t) &= s_i(t) + n_i(t) \\ x_j(t) &= s_j(t) + n_j(t) \\ x_k(t) &= n_k(t) \\ t &= 1, 2, ..., T \end{aligned}$$

where $s(t)$ is the signal time course generated by the hemodynamic response of an activated voxel, and $n(t)$ is the noise. The T dimensional distance between the time series are:

$$\begin{aligned} d_{ij} &= x_i(t) - x_j(t) = (s_i(t) - s_j(t)) + (n_i(t) - n_j(t)) \\ d_{ik} &= x_i(t) - x_k(t) = s_i(t) + (n_i(t) - n_k(t)) \end{aligned}$$

If the variability of the signal across different activated voxels is small, then $(s_i(t) - s_j(t))$ is close to a zero. So d_{ij} can be viewed as a noise vector, while d_{ik} contains an additional term: the signal information $s_i(t)$. But the Euclidean norm maps the T dimensional vector to a single number. Under cases of high SNR, we have $\|d_{ij}\|^2 < \|d_{ik}\|^2$ with high probability. However as the SNR decreases, this inequality becomes much weaker, making it more difficult to differentiate between two activated time series and an activated time series and a background time series. Graphs constructed via the n nearest neighbors (with the Euclidean norm) are flawed in the sense that the connection between nodes does not always reflect the intrinsic organization of the original data.

We keep the Euclidean norm to compare time series, but we enforce the existing the spacial correlation present in the fMRI data during the construction of the graph. Indeed, truly activated voxels tend to be spatially clustered. Therefore, we add additional edges to the adjacency graph obtained by the nearest neighbor search. If two voxels are neighbors spacially, then the corresponding time series will be connected in the graph, irrespective of the Euclidean between

these time series. In summary, the criteria for putting an edge between a pair of nodes in the graph is summarized as follows:

Put an edge between the nodes x and y if

- $\|x-y\|^2$ *is among the* n *smallest values of* $\|x-z\|^2$, *for all* $z \in G$ *or,*
- $\|x-y\|^2$ *is among the* n *smallest values of* $\|y-z\|^2$, *for all* $z \in G$ *or,*
- *if* $\|p(x)-p(y)\|^2 \leq r$, *where* $p(x)$ *is the spatial position of the voxel from which the time series* x *originates.*

By imposing this spacial neighborhood criterion, we strengthen the connections within the activated nodes as well as the connections within the non-activated nodes in the graph. This feature is very important for the diffusion process performed over the graph afterward: a random walk starting from an activated node will have much higher probability landing on another activated node rather than a non-activated node in a given number of steps. Fig. 2 shows the improvement in revealing the structure of the data set by the modified approach of the graph construction.

3 Normalized Cut

3.1 Segmentation by the Normalized Cut

Now that we have a description of the geometry of the dataset provided by the diffusion maps, separating the activated voxels from the non-activated voxels becomes a clustering problem. A closely related algorithm for clustering under a general graph-theoretic framework has been recently proposed [7]. Given a weighted graph $G = (V, E)$, we seek to partition the vertices into two disjoint subsets $A, B, A \cap B = \emptyset, A \cup B = V$, so that the similarity within each subset A and B is high and across A and B is low. Shi *et al* defined the following disassociation measure called the normalized cut ($Ncut$),

$$Ncut(A,B) = cut(A,B)(\frac{1}{vol(A)} + \frac{1}{vol(B)})$$

$$cut(A,B) = \sum_{u \in A, v \in B} w(u,v)$$

$$vol(A) = \sum_{u \in A, v \in V} w(u,v)$$

The optimal partitioning of the graph can be obtained by minimizing $Ncut$ over all possible subsets A and B. The combinatorial problem turns out to be NP-complete. However, we can relax the optimization problem by allowing the indicator function to take real values. The problem reduces then to minimizing the Laplacian of the graph, which can be computed efficiently. It is shown in [7] that

$$minNcut(A,B) = min_{\mathbf{f}} \frac{\mathbf{f}^T \mathbf{L} \mathbf{f}}{\mathbf{f}^T \mathbf{D} \mathbf{f}} \tag{6}$$

with the condition $\mathbf{f^T D1} = \mathbf{0}$, where $\mathbf{L}$ is the Laplacian of the graph. We can minimize (6) by solving the following generalized eigensystem,

$$\mathbf{Lf} = \lambda \mathbf{Df}. \tag{7}$$

The second smallest eigenvector $\mathbf{f}_2$ of the above eigensystem is the real valued solution to the normalized cut problem. However, since the eigenvectors take on continuous values, they are no longer indicator functions. We can still split the graph into two components by clustering the nodes according to the sign of $\mathbf{f}_2(i)$.

3.2 Relationship to Diffusion Maps

The normalized cut approach is one of the clustering techniques developed in the field of spectral graph theory [11]. Interestingly, the solution obtained in the previous section has very close ties to diffusion maps. Let $\mathbf{g} = \mathbf{D}^{\frac{1}{2}}\mathbf{f}$, assuming $\mathbf{D}$ is invertible,

$$\frac{\mathbf{f}^T\mathbf{Lf}}{\mathbf{f}^T\mathbf{Df}} = \frac{\mathbf{g}^T\mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}\mathbf{g}}{\mathbf{g}^T\mathbf{g}} = \frac{\mathbf{g}^T\tilde{\mathbf{L}}\mathbf{g}}{\mathbf{g}^T\mathbf{g}} \tag{8}$$

Recall that in section 2, we have defined a symmetric version of the diffusion matrix $\tilde{\mathbf{A}}$ in (1). It is easy to verify that $\tilde{\mathbf{L}} = \mathbf{I} - \tilde{\mathbf{A}}$. So the eigenvectors for $\tilde{\mathbf{L}}$ are the exactly the same ones for $\tilde{\mathbf{A}}$. Partitioning the graph subject to the normalized cut criterion using the embedding given by the diffusion maps is in a certain sense equivalent to itinerating the diffusion process over the graph, but with updated weight function dependent on the diffusion distances between pairs of points.

4 Experiment and Results

4.1 Details About the Implementation of the Algorithm

Given an fMRI dataset, we set up a weighted graph $G = (V, E)$ by taking each time series as a node and connecting pairs of nodes according to the criterion defined in 2.2. In the experiment we connect each node to 6 of its nearest neighbors in terms of the Euclidean distance and 4 of its spacial neighbors taking $r = 1$. The similarity kernel $k(x, y) = exp(-\frac{\|x-y\|^2}{\sigma})$ is chosen for its relative simplicity and isotropic property. The value of σ is typically set to 10 to 20 percent of the total range of the Euclidean distance between pairs of time series. Then the diffusion maps are constructed by computing the eigenvectors of the diffusion matrix $\tilde{A}$ in (1). We eliminate the first eigenvector, and keep the following four eigenvectors to characterize the structure of the data. The selection is justified by the fact that we expect to find only one activated manifold in the dataset. Furthermore, four degrees of freedom should provide enough richness to characterize this manifold. Now we use the new coordinates to rebuild the weighted graph. By splitting the second smallest eigenvectors of the Laplacian matrix, we obtain the partition of the graph. In the final step, we label the activated

voxels and examine the corresponding time series. Notice the fact that usually the cluster of the activated voxels has much smaller size than the cluster of the non-activated voxels. We can use this cue to quickly tell the two clusters apart. Further examination of the time series from the activated voxels is necessary to ensure that the activation detected is indeed related to the experiment paradigm.

4.2 Artificial Event-Related Data

We first apply our approach to an artificial data set of event-related fMRI. The fMRI signal time series are generated by convolving a stimulus time course with the hemodynamic filter $h(t)$ of SPM [12].

$$h(t) = (\frac{t}{d})^a exp(-\frac{t-d}{b}) - c(\frac{t}{d'})^{a'} exp(-\frac{t-d'}{b'}) \tag{9}$$

where $d = ab$ is the time to peak, $d' = a'b'$ is the time to undershoot, with $a = 6, a' = 12$ and $b = b' = 0.9s; c = 0.35$. This model has become the most frequent one since it models both activation, undershoot and that both modes are not symmetrical. The signal time series $s(t)$ is shown in the upper plot of Fig. 1(c). It remains deterministic for all activated voxels.

The synthetic data set contains $N = 30 \times 30 = 900$ brain voxels with one small activated focus of 13 voxels, see Fig. 1(a) for the spacial localization of the activation region. We add white Gaussian noise to the signal $s(t)$ to create the activated time series, and use white Gaussian noise for non-activated time series. Four data sets with SNR= $0.6, 0.7, 0.8, 1$ were generated. We apply the diffusion maps followed by the normalized cut to each data set.

Fig. 2 compares the result of diffusion process on graphs constructed with and without spacial information. At SNR= 0.8, the diffusion process on a graph constructed without the spacial information fails to reveal any interesting structure of the data set. As one can observe in the upper row in Fig. 2, the red stars that represent the activated time series are buried by the black circles which

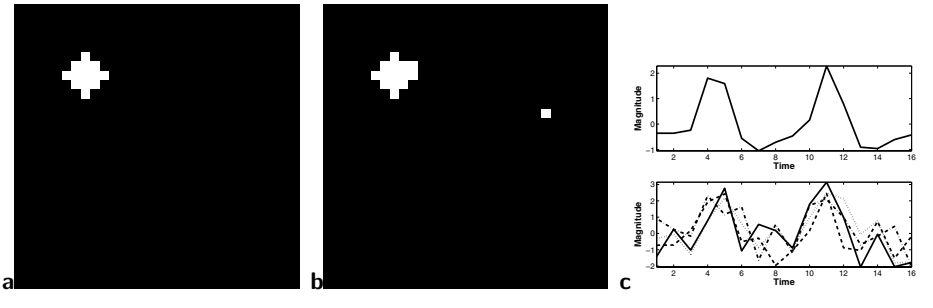

Fig. 1. Left: the true activation map with activated voxels in white and non-activated voxels in black; Middle: activation map for a data set with SNR= 0.8, obtained using our method; Right: (upper) the event-related signal time course $s(t)$, (lower) some of the activated time courses detected

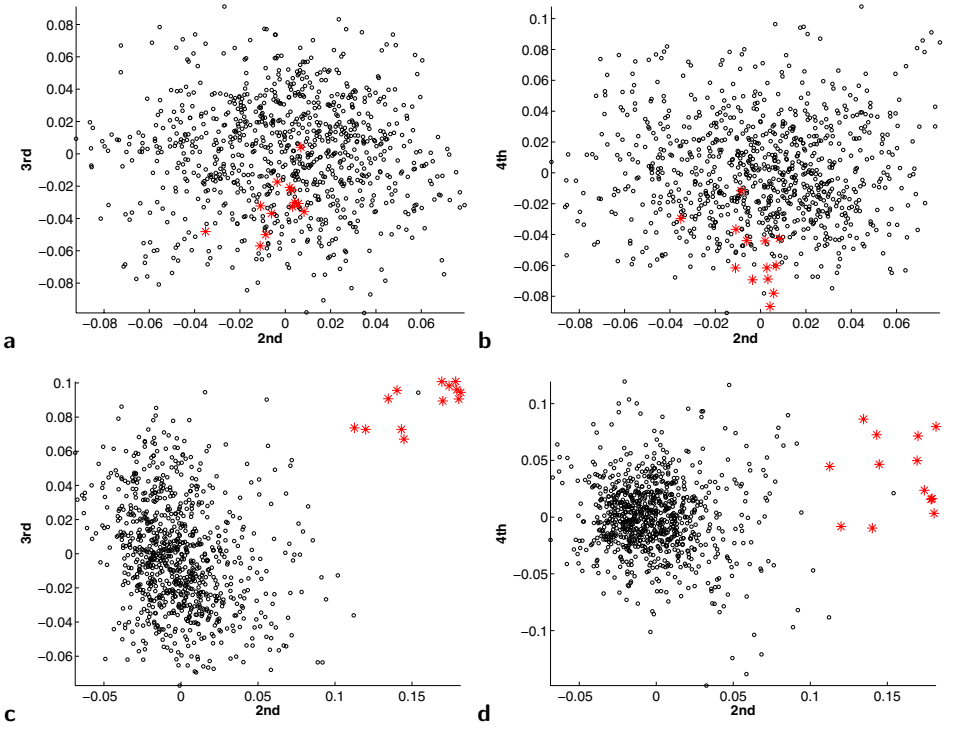

Fig. 2. Upper: diffusion on graph constructed without spacial information; Lower: diffusion on graph constructed with spacial information. The majority are non activated nodes represented by black circles. The activated nodes are represented by red stars

represent the non-activated time series. While for exactly the same data set, by imposing the connections between spacial neighbors, there is a considerable improvement in separating the activated nodes from the non-activated nodes, see the lower row in Fig. 2. Fig. 1(b) shows the activation map obtained for this data set, and the lower plot in Fig. 1(c) displays some of the activated time series detected.

We now compare the performance of our approach with the method of regression. The comparison is based on the number of true and false positives for each value of SNR. The true activation rate is the ratio between the number of true positives detected by the algorithm and the total number of true positives. The false activation rate is the ratio between the number of false positives detected by the algorithm and the total number of true negatives. For each SNR, we generate 10 independent data sets and the averaged statistics are shown in Fig. 3.

The Regression Method. Since we have the absolute knowledge of the signal time series $s(t)$ of the synthetic data set, we can regress each time series onto the signal time series, and apply a Student t-test to the regression coefficient to test

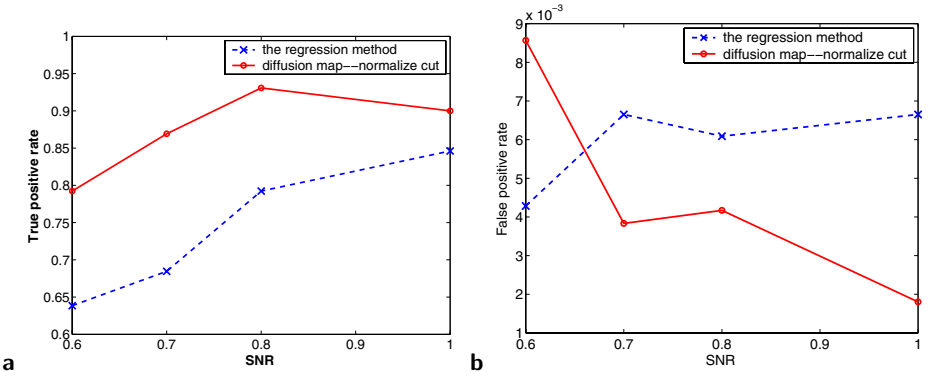

Fig. 3. True activation rate (a) and false activation rate (b) obtained with our diffusion-normalized cut method and the regression method

its significance. The null hypothesis is that the regression coefficient is not significant, which means the time series is not from an activated voxel. The activation map based on the regression method is obtained by thresholding the p value at $p = 0.005$. Knowing exactly the signal time series is a very strong condition which could never be realized in real fMRI experiment. We should expect the method of regression to give the best performance one can expect statistically. Fig. 3(a) shows that our method does very well in terms of detecting the activation. But it suffers from a high false positive rate when the SNR drops below 0.7. One thing worth mentioning here is that once the SNR is greater than 0.7, the false positives detected by our method are always spatially isolated, as is shown in Fig 1(b). This makes it possible to further reduce the number of false positives.

4.3 In-Vivo ER-fMRI Data

We present here the results of an experiment conducted with event-related fMRI data, provided by Dr. Gregory McCarthy (Brain Imaging and Analysis Center, Duke University), demonstrate prefrontal cortex activation in the presence of infrequent events. Visual stimuli were presented to the subjects: most of the images were squares. Infrequent events (targets) consisted in the appearance of circles at random times. Occasionally, images of everyday objects (novels) were also presented. A picture was displayed every 1.5 seconds. The subject was asked to mentally count the number of occurrences of the circles and report that number at the end of each run for total of 10 runs. The experiment was designed to study whether the processes that elicit P300, an event-related potential caused by infrequent target events whose amplitude is dependent on the preceding sequence of stimuli, could also be measured by fMRI. The data was acquired with a gradient echoplanar EPI sequence (TR=1500ms, TE=45ms, NEX=1, FOV=40×20cm, slice thickness =7mm, and imaging matrix 128×64). More details about the experiment are available in [13].

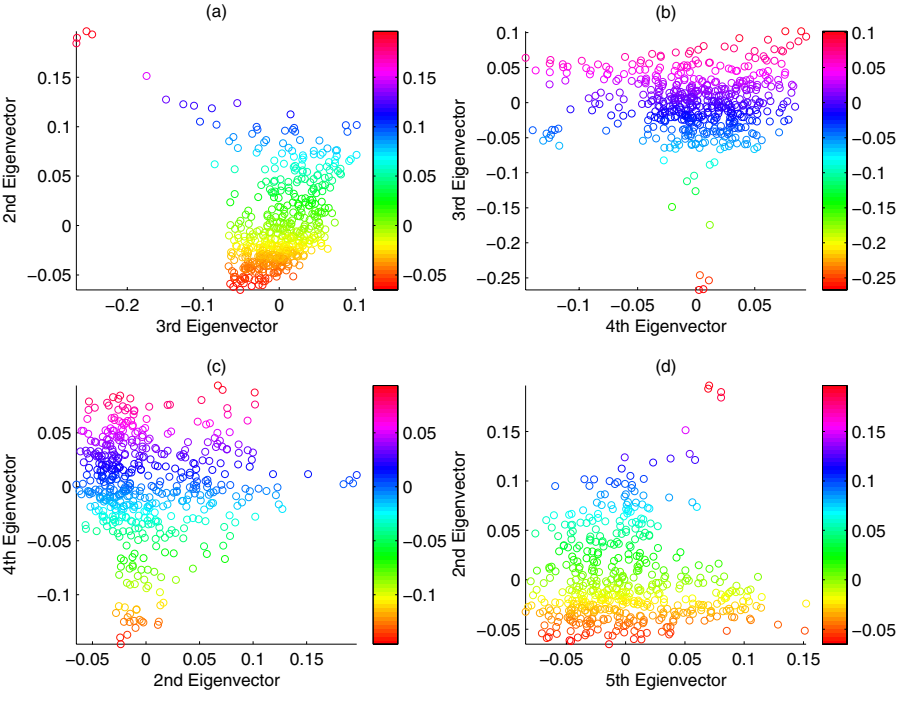

Fig. 4. The first few coordinates given by the diffusion maps

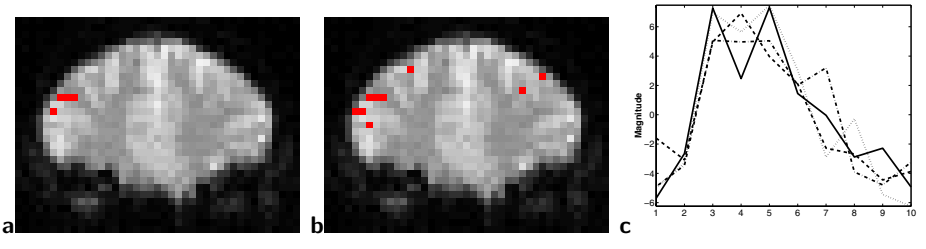

Fig. 5. Left: activation map generated using our method; Middle: activation map generated using the correlation analysis. Right:Time series from the activated voxels

In order to demonstrate the application of our approach to the study of activations by infrequent events, we extract 10-image segments consisting of the 10 consecutive images starting at the target onset. We have in each run about 5 to 6 targets for a total of 49 targets (the data for the 10th run were lost). These segments of images were averaged in order to increase the SNR. The mean value of the time series was removed voxel-wise before we applied our approach. Fig. 4 shows the data structure represented by the first few coordinates provided by the diffusion maps.

A small cluster of four voxels well detached from the mass of the point cloud can be seen in Fig. 4 especially in sub-plot(a). It indicates the presence of activated time series. The activation map generated using our approach is shown in Fig. 5(a). It is compared with the activation map generated by a correlation analysis, shown in Fig. 5(b). We compute the correlation between the time series and the hemodynamic response model defined in (9). The correlation threshold is 0.6. Meanwhile the time series from the four activated voxels are shown in Fig. 5(c).

5 Conclusion

We have presented in this paper a new approach that detects activation in an fMRI dataset. We view all the time series as vectors in a high dimensional space (the dimension is the number of time samples). We assume that all activated times series share a common structure and all belong to a low dimensional manifold. We constructed an embedding that reveals the organization of the data into an activated manifold and a cluster of non-activated time series. We use a graph partitioning technique–the normalized cut to find the separation between the activated manifold and the background time series.

Unlike most fMRI data analysis methods our approach does not require any model of the hemodynamic response or any *a priori* information. It could also be applied to block designed fMRI experiment. In fact, because the signal in block designed experiment has a stronger statistical power, we should expect our method to perform well. In our current implementation, we assume that there is only one type of activation. This assumption makes the presentation of the method simpler, and leads to easier implementation. The approach extends naturally to multiple activations associated with different hemodynamic responses. Although we focused our attention on event-related fMRI, our approach could be applied to the analysis of other biomedical datasets.

Acknowledgments. The authors are extremely grateful to Dr. Mauro Maggioni and Dr. Stéphane Lafon for discussions of the theory and the implementation of diffusion maps. This work was supported by the program on Multiscale Geometry and Analysis in High Dimensions at the Institute of Pure and Applied Mathematics (IPAM) at UCLA.

References

1. Aguirre, G.K., Zarahn, E., D'esposito, M., "The variability of human BOLD hemodynamic responses" Neuroimage **8** (1998) 360–369
2. Josephs O, Turner R, Friston K, "Event-related fMRI" Human Brain Mapping **5** (1997) 243–248
3. Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., "Event-related fMRI: characterizing differential responses" Neuroimage **7** (1998) 30–40

4. Clare, S., Humberstone, M., Hykin, J., Blumhardt, L., Bowtell, R., Morris, P., "Detecting activations in event-related fMRI using analysis of variance" Magne. Reson. Med. **42** (1999) 1117–1122
5. Meyer, F.G., Chinrungrueng, J., "Analysis of Event-related fMRI data using best clustering bases" Medical Imaging **22** (2003) 933–939
6. Coifman, R.R., Lafon, S., "Diffusion Maps" submitted to Applied and Computational Harmonic Analysis (2004)
7. Shi, J., Malik, J., "Normalized Cuts and Image Segmentation" Pattern Analysis and Machine Intelligence **22** no.8 (2000) 888–905
8. Roweis, S.T., Saul, L.K., "Nonlinear dimensionality reduction by local linear embedding" Science **290** (2000) 2323-2326
9. Belkin, M., Niyogi, P., "Laplacian eigenmaps for dimensionality reduction and data representation" Neural computation **13** (2003) 1373–1396
10. Donoho, D.L., Grimes, C., "Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data" Proceedings of the National Academy of Science **100** (2003) 5591-5596
11. Chung, F.R.K., Spectral Graph Theory Am. Math. Soc. (1997)
12. Friston, K.J., Ashburner, J., SPM 97 course notes (1997)
13. McCarthy, G., Luby, M., Gore, J., Goldman-Rakic, P., "Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI." Journal of Neurophysiology **77** (1997) 1630–1634