

Graph Representation for Content-based fMRI Activation Map Retrieval

Alba G. Seco de Herrera¹ and L. Rodney Long² and Sameer Antani²

Abstract—The use of functional magnetic resonance imaging (fMRI) to visualize brain activity in a non-invasive way is an emerging technique in neuroscience. It is expected that data sharing and the development of better search tools for the large amount of existing fMRI data may lead to a better understanding of the brain through the use of larger sample sizes or allowing collaboration among experts in various areas of expertise. In fact, there is a trend toward such sharing of fMRI data, but there is a lack of tools to effectively search fMRI data repositories, a factor which limits further research use of these repositories. Content-based (CB) fMRI brain map retrieval tools may alleviate this problem. A CB-fMRI brain map retrieval tool queries a brain activation map collection (containing brain maps showing activation areas after a stimulus is applied to a subject), and retrieves relevant brain activation maps, i.e. maps that are similar to the query brain activation map. In this work, we propose a graph-based representation for brain activation maps with the goal of improving retrieval accuracy as compared to existing methods. In this brain graph, nodes represent different specialized regions of a functional-based brain atlas. We evaluated our approach using human subject data obtained from eight experiments where a variety of stimuli were applied.

I. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) provides methods to non-invasively visualize brain activity by computing statistical evidence of functional brain activations. Task-based fMRI, which we focus on in this paper, is concerned with obtaining this activation data for experiments in which the subject undergoes a specific sensory stimulation, or carries out a specific task [1]. Developing successful fMRI experiments is labor- and time-intensive. It requires careful preparatory work on the experimental design, data acquisition techniques and data analysis methods [2]. In addition to the planning and acquisition planning, there is the cost in maintaining and operating the MR scanner and its operating environment. The combination of these resource demands means that, in any measure, fMRI data is not cheap, and it is therefore of great interest to maximize the use of the collected data. fMRI data sharing may be expected to be a significant factor for maximizing the research value

*This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC).

¹Alba G. Seco de Herrera is with School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK alba.garcia@essex.ac.uk

²L. Rodney Long and Sameer Antani with the Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, MD 20894, USA l.long@nih.gov sameer.antani@nih.gov

of collected fMRI data sets. For example, such sharing will enable experiment replications for validation (or dispute of the results) of the original study. Another significant example is that data sharing will allow analysis work by researchers who do not have physical resources to do the data acquisition themselves. Therefore fMRI data sharing can improve research practices or even reduce the cost of doing science. Despite this observation, fMRI data sharing is not yet a common or easy practice, although some efforts have been made [3], [4]. Databases such as Neurosynth¹, allow further analysis of the experimental output, i.e., of the brain activation maps resulting from the statistical analysis of the experiments. Even though some fMRI data repositories exist, what is still lacking is a retrieval tool for the analyzed data available in repositories. Content-based (CB) fMRI retrieval aims to provide a tool to search collections of fMRI brain activation maps for maps similar to an input query map. This capability supports the analysis of inter-study variations and provides a new tool for investigating study results [5].

Several researchers have published work on CB-fMRI retrieval [6], [7], [8], approaching the problem using various fMRI map descriptors. A common approach is to use the whole-brain, voxel-wise, as a descriptor [9]. Tungaraza et al. [10] proposed a method to threshold the voxels within the brain maps and extract spatially distinct regions from the voxels that remain. Each region is defined by a descriptor containing several geometrical values. Garcia et al. [11] proposed two descriptors to reduce the cost of whole-brain voxel-wise methods: a map layout descriptor (MLD) which captures the spatial distribution of intensity, and a whole-brain, ROI-wise, descriptor. In this work, we present a new a brain activation map descriptor based on treating the activated regions as nodes in a connected graph.

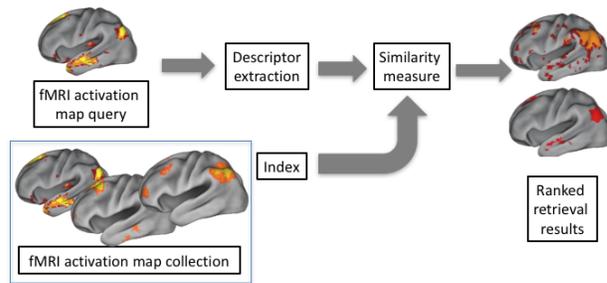


Fig. 1. Overview of the CB-fMRI activation map retrieval approach.

There is an increasing popularity of modeling the brain as

¹<http://neurosynth.org/>

a connected system where nodes represent different regions specialized by function [12], as a **brain graph**. Graph-based approaches provide a way to describe whole-brain patterns [13]. In this representation brain functionality is represented by a number of nodes interconnected by a set of edges [14]. Takerkart et al. [15] address inter-subject pattern analysis by proposing a graph-based approach closely related to our work. They use unsupervised learning to construct attributed graphs that represent fMRI brain activation maps. In this work, we proposed an undirected unweighted atlas-based brain graph to model the brain activation maps. The proposed brain graph is then used as a descriptor for the retrieval step.

II. METHODS

This section describes the dataset and the techniques we employed to carry out the experiments.

a) CB-fMRI activation map retrieval: The goal is to retrieve relevant brain activation maps from a database, for a given query. The query is itself a brain activation map. The retrieval approach follows the typical retrieval schema showed in Figure 1. A descriptor is extracted from the query map and for each of the brain activation maps in the database. Then, a similarity measure is used to compare the descriptor from the query with the descriptors extracted from the database (which are saved in the index). The final result consists of retrieval results ranked in order of similarity to the query.

b) Descriptor extraction: The main features of the two proposed fMRI brain activation map descriptors are described below.

- *ROI-based representation:* A binary descriptor based on whole-brain ROI is defined by Garcia et al. [11]. It combines voxels into functionally distinct ROIs which are provided by the human brain atlas proposed by Craddock et al. [16] (see Figure 2). The human brain atlas provided by Craddock et al. [16] parcellates the brain into spatially coherent, functionally homogeneous regions of interest (ROIs). The Craddock et al. atlas is used in this work because it is functional-based, rather than anatomical-based, which may improve fidelity of functional network modeling [17]. Garcia et al. [11] show that retrieval using this descriptor results in retrieval results similar to those obtained by using a voxel-wise approach, but in a more efficient way. Therefore, we chose the ROI-based representation for comparison with the graph-based representation proposed in this work.
- *Graph-based representation:* Most of proposed human brain graphs are binary brain graphs, where edges between nodes are undirected and unweighted [14]. We follow this approach in this work. The fMRI brain activation maps are represented as undirected unweighted atlas-based brain graphs (see Figure 2). The most common approach of the application of graph theory principles to neuroimaging brain data is to define nodes by the use of anatomical atlases for brain parcellation [18],

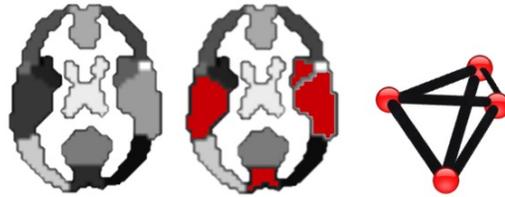


Fig. 2. In the left, a slice of the Craddock et al. atlas with 20 ROIs. In the middle, the same slice with some red ROIs representing the functionally activated ROIs in the brain after a particular stimulus. In the right, a connected graph representation of that activated brain.

[12]. We used the Craddock atlas [16] for consistency with the atlas used for the ROI-based descriptor and because this atlas allows a structural and functional graph representation of the fMRI brain activation maps. For each of the r_i Craddock ROIs there is a corresponding node n_i in the brain graph. Therefore, $N = n_1, \dots, n_N$ is the set of nodes of the brain graph and N is the number of the ROIs in the atlas. To define the edges we define all edges as undirected and unweighted. The set of edges is represented by a binary adjacency matrix A . The adjacency matrix A has binary elements, where each element (a_{ij}) is either 1, if the activation attributes of both connected nodes ($\theta(n_i)$ and $\theta(n_j)$) are 1, or 0 otherwise:

$$A = (a_{ij}) \in \mathbb{B}^{N \times N},$$

$$\text{where } a_{ij} = \begin{cases} 1 & \text{if } \theta(n_i) \neq 0 \text{ and } \theta(n_j) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Since the graph is undirected the adjacency matrix is symmetrical [19]. Therefore, only the upper triangular part of the adjacency matrix above the main diagonal, T , is considered. Similar to Richiardi et al. [20], a binary high-dimensional vector f is then generated by linearizing T . f is the graph-based descriptor which we use for measurement of similarity to other graphs.

c) Similarity measure: Levenshtein distance and histogram intersection (HI) are proposed to measure the similarity between graphs. A description of standard graph metrics which can be applied to brain graphs is given in [21]. For the ROI-based approach, we use HI to compare this approach with the proposed graph-based approach.

- *Levenshtein distance:* Edit distance is a widely accepted, commonly used method for graph matching [22]. Our graph-based descriptor f can be viewed as a string of "0"s and "1"s. This naturally suggests using Levenshtein distance (string edit distance) for similarity comparisons. This distance essentially consists of counting the minimum number of operations (insertions, deletions and substitutions) required to transform one graph-based descriptor f_x into the other f_y .
- *HI:* Due to the high computational complexity of Levenshtein distance [23], we also tested HI. This is possible because our graphs are undirected and unweighted. HI

measures the number of intersecting values in each bin of a histogram [24]; it is similar to Levenshtein distance, in the sense that the dissimilarity of graphs is quantified by the amount of distortion needed to transform one graph into another [23].

III. EVALUATION

The data was obtained from 359 subjects during 8 experiments. Since the distribution of subjects in each experiment is uneven, only four subjects per experiment were considered. The experiments consist of fMRI brain activity comparisons during various tasks such as watching films, using moral dilemmas as probes or doing basic memory exercises. Brain activity of each subject was recorded for each separate experiment (i.e. for each task). This study does not require ethical review because only statistical summaries with no identifying data are used.

In order to get the fMRI brain activation maps, a statistical analysis of fMRI time-series needs to be carried out. The retrieval approach presented in this work operates on brain activation maps regardless of the method of computation of the map; our retrieval is not limited to any particular method, and operates independently of the method used to compute the map. However, in the collection use for the evaluation contains brain activation maps extracted using the probabilistic independent component analysis (PICA) method only. The PICA components contains in each voxel a statistical z -value, indicating the level of activation in a brain voxel. From each subject 10 PICA components are used, and we considered only the highest 10% z -values in each component of the brain activation maps. In our experiments, we used each PICA component in the collection as a query.

For purposes of comparison with previous results, this work follows the same evaluation scenario as in Garcia et al. [11] and Bai et al. [25]. Therefore, a retrieved brain activation map is considered relevant to a given query if they both belong to the same experiment. Each brain activation map in the collection is used as a query. The brain activation maps which belong to the same subject as the query are excluded. Finally, we calculate the area under the ROC curve for each experiment.

IV. EXPERIMENTAL RESULTS

Graph-based algorithms can deal with only with a small number of nodes because of the high computational cost of computing graph distances [13]. In particular, Levenshtein distance grows exponentially with the number of nodes [29]. Therefore, in this work, for the experiments only a small number of nodes was chosen: $n = 10, 20, 30$. Using HI it is possible to use a larger number of nodes. In particular, we used $n = 200$ with HI to compare results with the previous result in Garcia et al. [11] which used 200 ROIs. Results using 200 nodes and the Levenshtein distance, for the graph-based approach approach, are not presented because the computational complexity became unacceptable. Table IV presents the results achieved using the two descriptors and the two similarity measures presented in Section II. The use

TABLE I
AVERAGE OF THE AREA UNDER THE ROC CURVE (%) AND THE STANDARD DEVIATION (SD) ON EIGHT EXPERIMENTS WITH FOUR SUBJECTS EACH ONE (AND TEN PICA COMPONENTS EACH ONE). TWO DESCRIPTORS AND TWO SIMILARITY MEASURES (SM) ARE USED.

Descriptor	n. ROIs	SM	%	SD
ROI	10	HI	49.65	11.70
ROI	20	HI	49.68	11.60
ROI	30	HI	49.66	11.60
ROI	200	HI	49.68	11.59
Graph	10	Levenshtein	50.50	6.79
Graph	20	Levenshtein	53.59	9.35
Graph	30	Levenshtein	52.75	9.49
Graph	10	HI	50.45	11.60
Graph	20	HI	52.04	8.14
Graph	30	HI	52.01	11.04
Graph	200	HI	50.55	13.09

of 200 ROIs does not improve retrieval results in the case of either descriptor. When applying the ROI-based approach and the HI distance the best results were obtained when using 20 ROIs; this yielded the same accuracy as when using 200 ROIs. Applying the graph-based approach, best results were also obtained when using 20 nodes. Overall best results were obtained when using the graph-based representation with the Levenshtein distance; next best was the graph-based representation with HI. In addition, we observed that, when using the graph-based representation, the standard deviation (SD) is smaller. Figure 3 shows the results using 20 ROIs. Graph-based approaches are more stable over the experiments than the ROI-based approach.

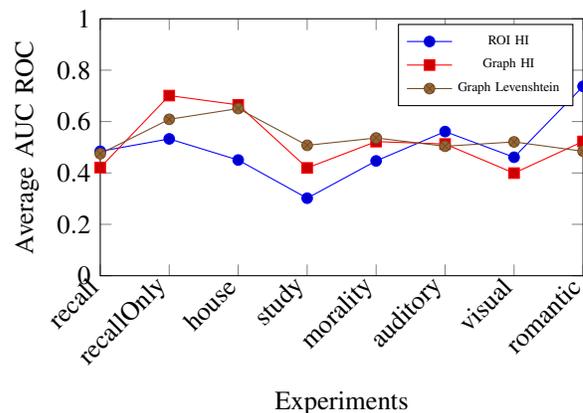


Fig. 3. Average area under the curve ROC per each of the experiments using the best runs. This runs used ROI-based or graph-based descriptors and Histogram Intersection (HI) or Levenshtein distance.

Finally, Table II displays the time needed to search for similarities on the collection for a given query. It shows using the Levenshtein distance, the query takes 18 seconds; with HI, the query takes less than 1 second.

V. DISCUSSION & CONCLUSIONS

We propose a new approach for CB-fMRI retrieval based on fMRI brain graph representation. We define an undirected unweighted atlas-based brain graph to model brain

TABLE II

TIME NEEDED TO SEARCH ON THE COLLECTION WITH 320 SUBJECTS (3,200 BRAIN MAPS SINCE EACH SUBJECT CONTAINS 10 PICA COMPONENTS) FOR ONE BRAIN MAP QUERY. THE COMPARED RUNS USED ROI-BASED OR GRAPH-BASED DESCRIPTORS AND HISTOGRAM INTERSECTION (HI) OR LEVENSHTAIN DISTANCE.

Method	ROI HI	Graph HI	Graph Levenshtein
Seconds	0.1	0.2	18

activation. Each node of the graph represents a specialized region defined by a functional-based brain atlas. We provide experimental retrieval results using two alternative distance measures: Levenshtein distance and HI. Despite the low accuracy achieved for our CB-fMRI retrieval approach, our results show that the proposed brain graph approach achieves better retrieval results on this set of data than the ROI-based approach, (which was shown by Garcia et al. [11] to be more efficient than voxel-based approaches). We feel that the low accuracies are likely due to the inherent difficulty of the problem and also due to the limitations of our evaluation scenario. We show that the proposed graph-based approach provides superior retrieval to the ROI-based approach for this data. Specifically, we show that best results are obtained when using only 20 nodes, and Levenshtein distance. However, this distance has the disadvantage of high computational complexity. HI achieves second best results, with low computational cost.

We expect that increasing sharing of fMRI data among research sites will significantly add to important findings about brain functioning. Tools for managing this data, such as the CB-fMRI retrieval tool we present, are critically needed to support the increasing demands for data sharing; more research and development work is required to meet this need, and the field of tool development for fMRI data search is under-researched at the current time. Our future work will include ROI spatial information in the brain graph, and we expect to expand our collection of brain maps to include a greater variety of experiments.

REFERENCES

- [1] J. A. Detre and J. Wang, "Technical aspects and utility of fmri using BOLD and ASL," *Clinical Neurophysiology*, vol. 113, no. 5, p. 621634, 2002.
- [2] J. M. Chein and W. Schneider, "Designing effective fMRI experiments," *Handbook of neuropsychology*, vol. 9, pp. 299–326, 2003.
- [3] J.-B. Poline, J. L. Breeze, S. S. Ghosh, K. Gorgolewski, Y. O. Halchenko, M. Hanke, K. G. Helmer, D. S. Marcus, R. A. Poldrack, Y. Schwartz, J. Ashburner, and D. N. Kennedy, "Data sharing in neuroimaging research," *Frontiers in Neuroinformatics*, vol. 6, p. 9, 2012.
- [4] R. A. Poldrack and K. J. Gorgolewski, "Making big data open: Data sharing in neuroimaging," *Nature Neurosciences*, vol. 17, no. 11, 2014.
- [5] G. Salimi-Khorshidi, S. M. Smith, J. R. Keltner, T. D. Wager, and T. E. Nichols, "Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies," *Neuroimage*, vol. 45, no. 3, pp. 810–823, 2009.
- [6] B. Bai, P. Kantor, N. Cornea, and D. Silver, "Ir principles for content-based indexing and retrieval of functional brain images," in *Proceedings of the 15th ACM international conference on Information and knowledge management*. ACM, 2006, pp. 828–829.
- [7] B. Bai, P. B. Kantor, N. D. Cornea, and D. Silver, "Toward content-based indexing and retrieval of functional brain images," in *IAIO*, 2007.
- [8] J. Zhang and V. Megalooikonomou, "An effective and efficient technique for searching for similar brain activation patterns," in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*. IEEE, 2007, pp. 428–431.
- [9] B. Bai, "Feature extraction and matching in content-based retrieval of functional magnetic resonance images," Ph.D. dissertation, The State University of New Jersey, 2007.
- [10] R. Tungaraza, J. Guan, L. Shapiro, J. Brinkley, J. Ojemann, and J. Franklin, "A similarity retrieval tool for functional magnetic resonance imaging statistical maps," *International Journal of Biomedical Data Mining*, vol. 2, 2013.
- [11] A. García Seco de Herrera, L. R. Long, and S. Antani, "Content-based fmri brain maps retrieval," in *Brain Informatics and Health: International Conference, BIH 2016*. Springer International Publishing, 2016, pp. 173–180.
- [12] F. D. V. Fallani, J. Richiardi, M. Chavez, and S. Achard, "Graph analysis of functional brain networks: Practical issues in translational neuroscience," *Philosophical Transactions of the Royal Society B*, vol. 369, no. 1653, p. 20130521, 2014.
- [13] J. Richiardi, S. Achard, H. Bunke, and D. Van De Ville, "Machine learning with brain graphs: Predictive modeling approaches for functional imaging in systems neuroscience," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 58–70, 2013.
- [14] E. T. Bullmore and D. S. Bassett, "Brain graphs: Graphical models of the human brain connectome," *Annual Review of Clinical Psychology*, vol. 7, pp. 113–140, 2011.
- [15] S. Takerkart, G. Auzias, B. Thirion, and L. Ralaivola, "Graph-based inter-subject pattern analysis of fMRI data," *PLoS one*, vol. 9, no. 8, p. e104586, 2014.
- [16] R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, "A whole brain fmri atlas generated via spatially constrained spectral clustering," *Human Brain Mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.
- [17] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, "Network modelling methods for fmri," *Neuroimage*, vol. 54, no. 2, pp. 875–891, 2011.
- [18] M. L. Stanley, M. N. Moussa, B. Paolini, R. G. Lyday, J. H. Burdette, and P. J. Laurienti, "Defining nodes in complex bbrain networks," *Frontiers in computational neuroscience*, vol. 7, p. 169, 2013.
- [19] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.
- [20] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville, "Decoding brain states from fMRI connectivity graphs," *Neuroimage*, vol. 56, no. 2, pp. 616–626, 2011.
- [21] H. Bunke, "Structural pattern recognition," Bern, Switzerland, Tech. Rep.
- [22] B. Cao, Y. Li, and J. Yin, "Measuring similarity between graphs based on the levenshtein distance," *Appl. Math*, vol. 7, no. 1L, pp. 169–175, 2013.
- [23] K. Riesen and H. Bunke, "Approximate graph edit distance computation by means of bipartite graph matching," *Image and Vision Computing*, vol. 27, no. 7, pp. 950–959, 2009.
- [24] A. Ulaş, U. Castellani, M. Bicego, V. Murino, M. Bellani, M. Tansella, and P. Brambilla, *Analysis of Brain Magnetic Resonance (MR) Scans for the Diagnosis of Mental Illness*. London: Springer London, 2013, pp. 247–287. [Online]. Available: http://dx.doi.org/10.1007/978-1-4471-5628-4_10
- [25] B. Bai, P. Kantor, A. Shokoufandeh, and D. Silver, "fmri brain image retrieval based on ICA components," in *Conference on Current Trends in Computer Sciencel*. IEEE, 2007, pp. 10–17.
- [26] C. Chatfield and A. J. Collins, *Introduction to multivariate analysis*. Springer, 1980.
- [27] C. F. Beckmann and S. M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 23, no. 2, pp. 137–152, 2004.
- [28] N. F. Ramsey, "fmri paradigm design," Tech. Rep.
- [29] B. Gaüzère, S. Bougleux, K. Riesen, and L. Brun, *Approximate Graph Edit Distance Guided by Bipartite Matching of Bags of Walks*. Springer Berlin Heidelberg, 2014, pp. 73–82.