Content-Based fMRI Brain Maps Retrieval

Alba G. Seco de Herrera[(✉)], L. Rodney Long, and Sameer Antani

Lister Hill National Center for Biomedical Communications,
National Library of Medicine, Bethesda, USA
{albagarcia,l.long,sameer.antani}@nih.gov
https://lhncbc.nlm.nih.gov

Abstract. The statistical analysis of functional magnetic resonance imaging (fMRI) is used to extract functional data of cerebral activation during a given experimental task. It allows for assessing changes in cerebral function related to cerebral activities. This methodology has been widely used and a few initiatives aim to develop shared data resources. Searching these data resources for a specific research goal remains a challenging problem. In particular, work is needed to create a global content–based (CB) fMRI retrieval capability.

This work presents a CB fMRI retrieval approach based on the brain activation maps extracted using Probabilistic Independent Component Analysis (PICA). We obtained promising results on data from a variety of experiments which highlight the potential of the system as a tool that provides support for finding hidden similarities between brain activation maps.

Keywords: fMRI retrieval · PICA · Brain activation map

1 Introduction

Functional Magnetic Resonance Imaging (fMRI) is a powerful tool used in the study of brain function. It can non-invasively detect signal changes in areas of the brain where neuronal activity is varying [19]. Following each stimulus the scanner generates a time-series of 3-D volumetric data, where each voxel represents the time course of the Blood Oxygen Level Dependent (BOLD) response at that voxel. Subsequent statistical processing generates additional 3-D brain maps, in which each voxel represents the probability that a statistically significant change in BOLD response occurred between a stimulus and corresponding control [16]. In order to obtain the statistical brain maps, the General Linear Model (GLM) has been extensively used [15]. However, Probabilistic Independent Component Analysis (PICA), a variant of the traditional Independent Component Analysis (ICA) is becoming more popular for fMRI data analysis [1]. In this work PICA components are used as the retrieval unit.

Data sharing is becoming increasingly common, but despite encouragement and facilitation by some research efforts most neuroimaging data acquired today is still not shared due to political, financial, social, and technical barriers [10].

Some efforts have been made to develop fMRI repositories. However, a global content-based (CB) fMRI retrieval capability is still lacking. Such a capability would allow a researcher to retrieve studies relevant to a specified study or interest.

Most of the work carried out addresses the problem of the classification of brain images. Some researchers try to detect activation volumes in the same brain image sequence [13,14], while others try to distinguish experiments with different cognitive tasks [9]. In these studies, machine learning (ML) methods (k-nearest neighbors [14], Bayesian [14], Support Vector Machines (SVM) [13,14] and Fishers Discriminant Analysis [9] are applied to the time series of voxels [13, 14], or results from additional processing, such as t-maps [9] generated by the General Linear Model (GLM). All ML methods characterize the distribution of some set of features for labeled training datasets, and use these characteristics to classify other datasets.

Some work has been done on CB fMRI retrieval [3,5,20]. Laconte [13] assumes a priori knowledge exists in the temporal characteristics of the data. Bai [2] applied several similarity measures to retrieve the brain maps calculated using PICA. Shapiro et al. [16] retrieve fMRI signals based on prior matching of the raw signal data to eight signal templates. In 2013, Tungaraza et al. [18] proposed a method for retrieving similar fMRI statistical images given a query fMRI statistical image. The method thresholds the voxels within those images and extracts spatially distinct regions from the voxels that remain. Each region is defined by a feature vector containing several geometrical values. The similarity between two images is obtained by the summed minimum distance (SMD) of their constituent feature vectors.

A new method for CB fMRI brain maps retrieval is proposed in this work.

The remainder of the paper is organized as follows. Section 2 presents the techniques proposed in this paper as well as the database used to evaluate them. The experiments that we carried out are presented in Sect. 3. Finally, a discussion and the conclusions are given in Sect. 4.

2   Methods

This section describes the techniques developed to create a CB fMRI retrieval system. A description of the dataset and framework used to evaluate the proposed system is also given.

2.1   Dataset

In this paper the data and evaluation scenario provided by Bai et al. [4] are used. The data used is obtained from 359 subjects during 8 experiments. However, the number of subjects per experiment is not uniformly distributed. From each subject 10 PICA components are used, as provided by Bai et al. [4]. PICA

components were calculated using FLS[1] library Each PICA component is a complete brain map (with dimension $91 \times 109 \times 91$) where each voxel has an associated statistical z-value which indicates the level of activation. Table 1 shows the distribution of the number of subjects participating in each experiment. The experiments consist of fMRI brain activity comparisons during various tasks. Brain activity of each subject was recorded for each separate experiment (i.e. for each task). A variety of tasks are covered in the various experiments, such as watching films, using moral dilemmas as probes or doing basic memory exercises.

Table 1. Number of subject per experiment.

Experiment	N. subjects
Oddball-visual	4
Oddball-auditory	4
Event perception–House Active	27
Event perception–Study Active	25
Morality	248
Study-Recall	27
Recall-Only	9
Romantic	15
Total	359

2.2   Techniques

This section describes the basic techniques that we used in in this study. Figure 1 puts together all the basic components in our retrieval system.

We used brain map extracted with PICA as the retrieval unit. Bai et al. [4] used a whole-brain voxel–wise strategy as a brain map descriptor. That method is computationally expensive, sensitive to noise and difficult to interpret [8]. Instead, we present two alternative descriptors: a map layout descriptor (MLD) and a whole–brain ROIs–wise descriptor.

MLD is similar to the Color Layout Descriptor [12] commonly used in 2D CB image retrieval (CBIR). It is designed to capture the spatial distribution of intensity in a volume. The feature extraction process consists of a grid based representative intensity selection.

Whole–Brain Region of Interest (ROI)–Wise – combines voxels into functionality distinct ROIs. The human brain atlas provided by Craddock et al. [8] (see Fig. 2) is used in this paper because it is functionally homogeneous, spatially

---

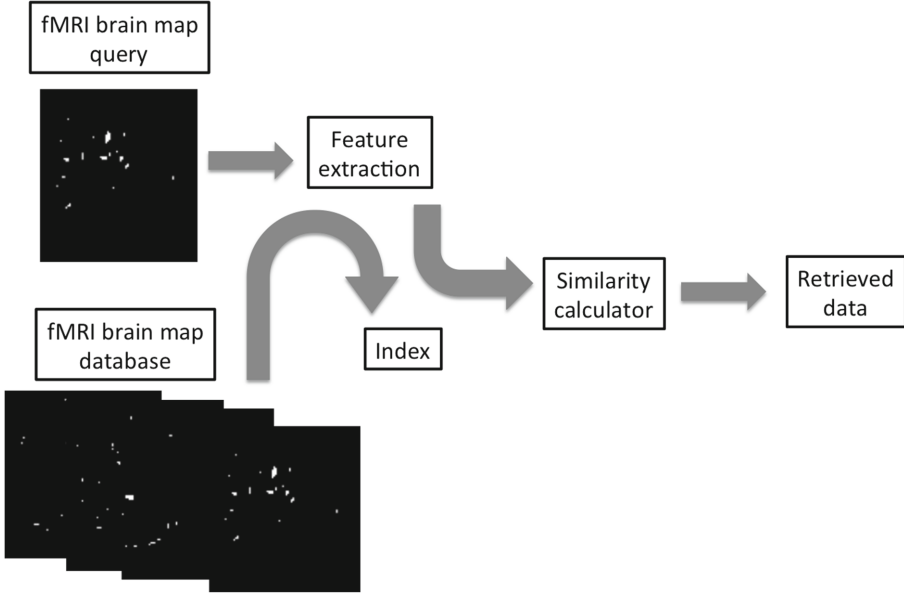[1] http://fsl.fmrib.ox.ac.uk/fsl/f.

Fig. 1. Outline of the basic elements of the retrieval system.

contiguous, and it represent the functional connectivity (FC) patterns of the brain. The feature extraction process comprises the following process. For each brain map $B_i, i = 1, \ldots, n$, there is a set of MNI coordinates[2] $C = \{c_1, \ldots, c_m\}$, where m is the number of voxels in each brain map, (if they are not MNI coordinates they are converted to MNI). Each coordinate contains the z-values from the PICA component. The histogram $\mathbf{h} \in \mathbb{B}^{200}$ is defined as:

$$\mathbf{h_k} = \begin{cases} 1 \text{ if } \exists c_j \subset R_k \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

where R_k is the k-th Craddock ROI, $\forall k, k = 1, \ldots 200$. If the MNI coordinates c_j lie outside the brain but they are within 5 voxels of Craddock ROI R_k, then $\mathbf{h_k = 1}$.

These descriptors allow the extraction of the main features of the fMRI brain activation maps.

In this work, two measures are tested for the similarity comparison for each of the descriptors: histogram intersection (HI) [17] and Euclidean distance. For each subject only the nearest PICA component is considered. CombSum is the fusion strategy applied to combine results of each of the descriptors of the same fMRI brain activation map. See [11] for more details on the chosen fusion strategies.

---

[2] The Montreal Neurological Institute (MNI) defined a standard brain, which is representative of the population, by using a large series of MRI scans on normal controls.
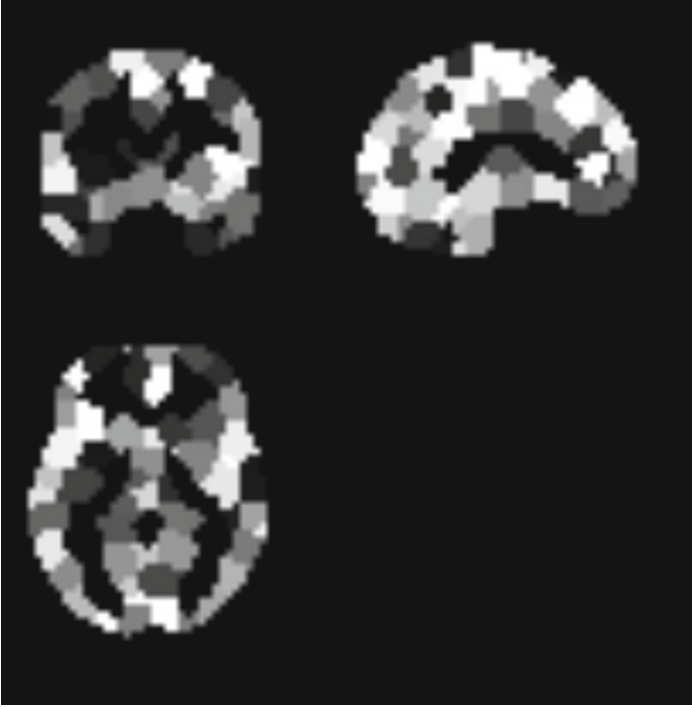
Fig. 2. Atlas provided by Craddock et al. consisting on 200 ROIs.

2.3   Evaluation

The main objective of this work is to evaluate the effectiveness of the techniques discussed in Sect. 2 for fMRI brain activation map retrieval based on PICA components. Therefore, the evaluation scenario used in this work is reused from Bai et al. [4] for comparison. The scenario considers a retrieved brain map relevant to a query if they both belong to the same experiment. This assumption is actually too strict and not totally accurate because there can be hidden similarities among different experiments. Despite the limitations of the evaluation framework, it provides a basic scenario to compare the proposed approaches with the state-of-the-art.

Each brain map in the dataset (see Sect. 2.1) is used as a query. Every brain map belonging to the same subject as the query is excluded. The area under the ROC curve is calculated for every experiment. Results are compared with the best run proposed by Bai et al. [4].

3   Experimental Results

This section details the runs produced for fMRI CB retrieval using the techniques and evaluation framework presented in Sect. 2. The characteristics of each of the 6 runs proposed are presented below:

– *Run1* – the MLD descriptor is extracted and HI is used for similarity comparison;
– *Run2* – the whole–brain ROIs–wise descriptor is extracted and HI is used for similarity comparison;
– *Run3* – the MLD and whole–brain ROIs–wise descriptors are extracted and HI is used for similarity comparison.
– *Run4* – the MLD descriptor is extracted and Euclidean distance is used for similarity comparison;
– *Run5* – the whole–brain ROIs–wise descriptor is extracted and Euclidean distance is used for similarity comparison;
– *Run6* – the MLD and whole–brain ROIs–wise descriptors are extracted and Euclidean distance is used for similarity comparison.

This runs are also compared with the best run proposed by Bai et al. [4].
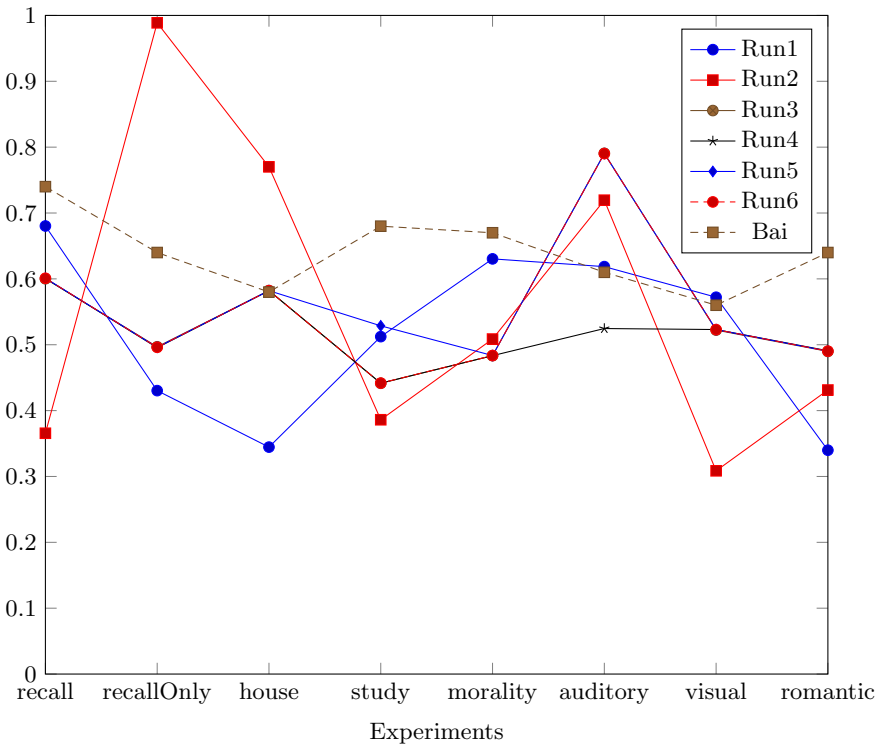

Fig. 3. Average area under the curve ROC per each of the experiments using the various runs and compared with the baseline proposed by Bai et al.

Table 2. Average area under the curve ROC.

	Recall	RecallOnly	House	Study	Morality	Auditory	Visual	Romantic
Run1	0.68	0.43	0.34	0.51	0.63	0.62	0.57	0.34
Run2	0.37	0.99	0.77	0.39	0.51	0.72	0.31	0.43
Run3	0.6	0.5	0.58	0.44	0.48	0.79	0.52	0.49
Run4	0.6	0.5	0.58	0.44	0.48	0.52	0.52	0.49
Run5	0.6	0.5	0.58	0.53	0.48	0.79	0.52	0.49
Run6	0.6	0.5	0.58	0.44	0.48	0.79	0.52	0.49
Bai	0.74	0.64	0.58	0.68	0.67	0.61	0.56	0.64

Figure 3 shows the average area under the ROC curve for each experiment. A more detailed view of the results can be found in Table 2. Results show that there is a big difference between experiments probably due to the uneven distribution of subjects between experiments. However, the use of the proposed descriptors achieved results comparable with the previous work of Bai et al. [4].

4   Conclusions

This article describes the methods and results applied to a novel method for fMRI brain activation map retrieval process. This is an area of research which has not been extensively explored. We propose two new descriptors to represent fMRI brain activation maps and simplify the retrieval of relevant cases.

Experimental data provided by Bai et al. [4] is used for the evaluation of the techniques. We note that it is difficult to assess when a fMRI brain activation map is relevant for a given query and plan to explore additional methods, such as the combination of human judgments with statistical methods [6,7]. Despite this limitation, the framework provides a basic approach to compare alternative feature descriptors and similarity measures with the limited state-of-the-art available in this field.

The results are promising but there is a big difference between experiments. Future work will expand the experimental data set and incorporate better fusion of several descriptors to help the retrieval. Further investigation on applications of CB fMRI brain activation retrieval will also be carried out.

References

1. Ao, J.: An optimized statistical analysis of fMRI data using independent component analysis. Ph.D. thesis, Texas Tech University (2010)

2. Bai, B.: Feature extraction and matching in content-based retrieval of functional magnetic resonance images. Ph.D. thesis, The State University of New Jersey (2007)
3. Bai, B., Kantor, P., Cornea, N., Silver, D.: IR principles for content-based indexing and retrieval of functional brain images. In: Proceedings of 15th ACM International Conference on Information and Knowledge Management, pp. 828–829. ACM (2006)
4. Bai, B., Kantor, P., Shokoufandeh, A., Silver, D.: fMRI brain image retrieval based on ICA components. In: Conference on Current Trends in Computer Sciencel, pp. 10–17. IEEE (2007)
5. Bai, B., Kantor, P.B., Cornea, N.D., Silver, D.: Toward content-based indexing and retrieval of functional brain images. In: RIAO (2007)
6. van den Broek, E.L., Kisters, P.M., Vuurpijl, L.G.: The utilization of human color categorization for conten-based image retrieval. In: Proceedings of SPIE, Human Vision and Electronic Imaging IX, pp. 351–362 (2004)
7. van den Broek, E.L., Kisters, P.M., Vuurpijl, L.G.: Content-based Image retrieval benchmarking: utilizing color categories and color distributions. J. Imaging Sci. Technol. **49**, 293–301 (2005)
8. Craddock, R.C., James, G.A., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S.: A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum. Brain Mapp. **33**(8), 1914–1928 (2012)
9. Ford, J., Farid, H., Makedon, F., Flashman, L.A., McAllister, T.W., Megalooikonomou, V., Saykin, A.J.: Patient classification of fMRI activation maps. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2879, pp. 58–65. Springer, Heidelberg (2003). doi:10.1007/978-3-540-39903-2_8
10. Haselgrove, C., Poline, J.B., Kennedy, D.N.: A simple tool for neuroimaging data sharing. Front. Neuroinformatics **8**, 82 (2014)
11. García Seco de Herrera, A., Schaer, R., Markonis, D., Müller, H.: Comparing fusion techniques for the ImageCLEF 2013 medical case retrieval task. Comput. Med. Imaging Graph. **39**, 46–54 (2015)
12. Kasutani, E., Yamada, A.: The MPEG-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval. In: Proceedings of International Conference on Image Processing, ICIP 2001, pp. 674–677 (2001)
13. LaConte, S., Strother, S., Cherkassky, V., Anderson, J., Hu, X.: Support vector machines for temporal classification of block design fMRI data. NeuroImage **26**(2), 317–329 (2005)
14. Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M., Newman, S.: Learning to decode cognitive states from brain images. Mach. Learn. **57**(1–2), 145–175 (2004)
15. Poline, J.B., Brett, M.: The general linear model and fMRI: does love last forever? Neuroimage **62**(2), 871–880 (2012)
16. Shapiro, L.G., Atmosukarto, I., Cho, H., Lin, H.J., Ruiz-Correa, S., Yuen, J.: Similarity-based retrieval for biomedical applications. In: Perner, P. (ed.) Case-Based Reasoning on Images and Signals. SCI, vol. 73, pp. 355–387. Springer, Berlin (2008)
17. Swain, M.J., Ballard, D.H.: Color indexing. Int. J. Comput. Vis. **7**(1), 11–32 (1991)
18. Tungaraza, R., Guan, J., Shapiro, L., Brinkley, J., Ojemann, J., Franklin, J.: A similarity retrieval tool for functional magnetic resonance imaging statistical maps. Int. J. Biomed. Data Min. **2**, 1–12 (2013)
19. Woolrich, M.W., Beckmann, C.F., Nichols, T.E., Smith, S.M.: Statistical analysis of fMRI data. In: Filippi, M. (ed.) fMRI Techniques and Protocols. Neuromethods, vol. 41, pp. 179–236. Springer, Berlin (2009)
20. Zhang, J., Megalooikonomou, V.: An effective and efficient technique for searching for similar brain activation patterns. In: 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2007, pp. 428–431. IEEE (2007)