

# A Similarity Retrieval Tool for Functional Magnetic Resonance Imaging (fMRI) Statistical Maps

R. F. Tungaraza<sup>a</sup>, J. Guan<sup>d</sup>, L. G. Shapiro<sup>a</sup>, J. F. Brinkley<sup>b</sup>, J. Ojemann<sup>c</sup>, and J. D. Franklin<sup>b</sup>

<sup>a</sup>Computer Science and Engineering, Box 352350, U. of Washington, Seattle, WA 98195;

<sup>b</sup>Biological Structure, Box 357420, U. of Washington, Seattle, WA 98195;

<sup>c</sup>Neurosurgery, Box 359300, U. of Washington, Seattle, WA 98195;

<sup>d</sup> Computer Science, Box 5838, Winona State University, Winona, MN 55987

## Abstract

We propose a method for retrieving similar fMRI statistical images given a query fMRI statistical image. Our 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 that contains the region centroid, the region area, the average activation value for all the voxels within that region, the variance of those activation values, the average distance of each voxel within that region to the region's centroid, and the variance of the voxel's distance to the region's centroid. The similarity between two images is obtained by the summed minimum distance of their constituent feature vectors. Results on a dataset of fMRI statistical images from experiments involving distinct cognitive tasks are shown.

Keywords: fMRI, content-based retrieval, brain activation patterns, SPM, statistical images, biomedical imaging

## 1 Introduction

A fundamental goal in functional neuroimaging is to identify areas of activation in the brain relative to a given task. Functional Magnetic Resonance Imaging (fMRI) is one technique used to identify such changes because changes in neuronal activity along a given region of the brain can be captured by a corresponding change in voxel value intensity on the acquired fMRI image. Statistical Parametric Mapping (SPM) [1] is the current popular technique used to analyze fMRI images. An SPM image contains test statistics determined at each pixel by the ratio between the intensity of the signal and its variance across experimental conditions.

Consider a scenario where a number of research groups have conducted different types of fMRI experiments. Each group has at least ten subjects for each experiment. Following

the experiments, the groups use SPM to identify regions in their subjects' brains that were significantly activated due to the experimental stimuli. At the end of this process each group deposits their fMRI raw data and the accompanying statistical maps into a joint database.

Following this, another researcher wants to find out whether the fMRI activation patterns of a subject not currently in the database are similar to any existing activation patterns in the database. She wants to retrieve other subjects regardless of the experimental condition and/or disorder that could potentially exhibit similar activation patterns. She also wants a numeric representation of the degree of similarity between the query image and all the retrieved images.

One might wonder why this type of information is of any use. Bai [2] proposes several scenarios that could lead a researcher into such an exploratory activity including helping to discover hidden similarities among superficially different studies, identifying similarities between datasets with a not-well-defined stimulus (e.g. subject is watching a movie clip), and discovering similarities in brain activity when the cognitive tasks do not seem to be related based on psychological reasoning alone.

Besides those potential uses of this tool, doctors with patients who respond differently to treatments for a specific disease might be able to use this tool to identify the best group to which a given patient should be assigned and consequently administer the appropriate type of treatment [3]. Suppose there are two distinct categories of brain activity for a given task following a stroke, each with a different treatment plan. Treatment A works best for patients in group A, and treatment B works best for those in group B. Our retrieval tool will enable this doctor to map a new patient with the same disorder into the best group and administer the appropriate treatment. The doctor will not only have a general activation pattern from each group to compare to, but also a score representing how similar/dissimilar the new patient's activation patterns are to every member of the two stroke-patients groups.

One way to accomplish the above task is to calculate the average brain for each group, and then, compare the new brain to the average brains in order to determine its membership. Average brains, however, are not always good representations of the group activation patterns, because they are sensitive to outliers [4]. Kherif et. al. [5] note four sources of outliers within a given group: the presence of a known factor that may cause differences in subject response to a given stimuli (e.g. right and left handed preferences), variations in subjects' brain anatomy, subjects' use of different cognitive strategies to perform the same task, and undetected scanning problems. Besides, some subjects tend to be "high activators" while other are "low activators" given the same experimental conditions [6]. The fMRI volumes of high activators have a lot more active voxels compared to the low activators within the same sample group. The presence of one or more subjects within a group under any of these conditions will result in activation patterns that do not conform to the rest. Consequently, the average brain may not be accurate.

It is these types of inconsistencies within the data itself that create differences in performance

among the methods geared towards computing the average activated brain. McNamee et. al. [6] examined four popular brain averaging methods: random effects analysis, Stouffer method, Fisher’s method, and average t-maps. They found that random effects analysis was the most stable and conservative compared to the other three methods. It was stable because the average brain wasn’t affected much by outliers and conservative because it downplayed patterns of activation that were not common to all subjects within the dataset. The Stouffer method and average t-map had intermediate performance. Fisher’s method was the most liberal and unstable. Each of these methods has strengths and weaknesses. None of them, however, can be used successfully with a non-homogeneous dataset.

There is another subtle problem in providing a similarity score for fMRI images: two neuroscientists (or radiologists) can score the same set of fMRI images differently. This problem becomes worse as the total number of images to be scored increases. An automated tool that can provide scores mimicking the average neuroscientist’s scoring scheme may provide better consistency. By the same token, neuroscientists and radiologists are humans. Therefore they approach the scoring process with a bias toward activation patterns they expect given the experiment. In such situations they may miss novel but significant activation patterns exhibited by a subset of the subjects within a group.

Given these drawbacks from both the averaging methods and the human observer, it is important to create inter-subject similarity measures that can be used not only to provide a similarity score between two fMRI volumes, but also to test the homogeneity of group activation patterns. Though the literature on fMRI inter-subject similarity measures is scarce, available methods utilize techniques such as percentage overlap of the activated voxels within the fMRI images [7], bipartite matching of fMRI-ICA spatial maps [8], encoding the original fMRI image into wavelet coefficients [9] or code-blocks [10], and calculating an RV coefficient of two fMRI images [5].

In our observation of a neuroscientist scoring fMRI images, we noticed that the features he emphasized when comparing a pair of images weren’t always the same. Sometimes the absence or presence of a given activation cluster in one image but not the other played a significant role in determining the final score. Though he considered the location, shape and strength of the activation, the scoring process could not directly be translated into a systematic algorithm. It became clear that the similarity measure needed to permit a certain degree of flexibility in terms of what features to include during the scoring process. We experimented with different ways to represent the shape, location, and strength of activation for each activation cluster. The shape-specific features were the number of voxels found in the activated cluster, the average distance of those voxels to the centroid, and their corresponding variance to the centroid. The location was represented by the centroid of the cluster, and the strength of activation was represented by the average activation values within each cluster and the variance of those activation values. These features were incorporated in our similarity-based retrieval system.

In this paper, we describe both the method for comparing two images and the accompanying retrieval system. It is a similarity-based retrieval system geared towards fMRI images in the form of statistical maps. Given a query statistical map and a database of other such maps from different subjects under different experimental conditions, the system retrieves all images similar to the query image in order of similarity. The rest of the paper is organized as follows: in section 2 we describe our method for computing the inter-subject similarity and give an overview of the retrieval system, section 3 demonstrates our results, section 4 is a discussion of our findings, and in section 5 we conclude our study with some suggestions for future work.

## 2 Methodology

Our methodology has three parts: preprocessing, feature extraction, and similarity calculation. Figure 1 illustrates the main steps in this process. In this section we describe these parts in detail and also discuss the graphical interface through which users query the system.

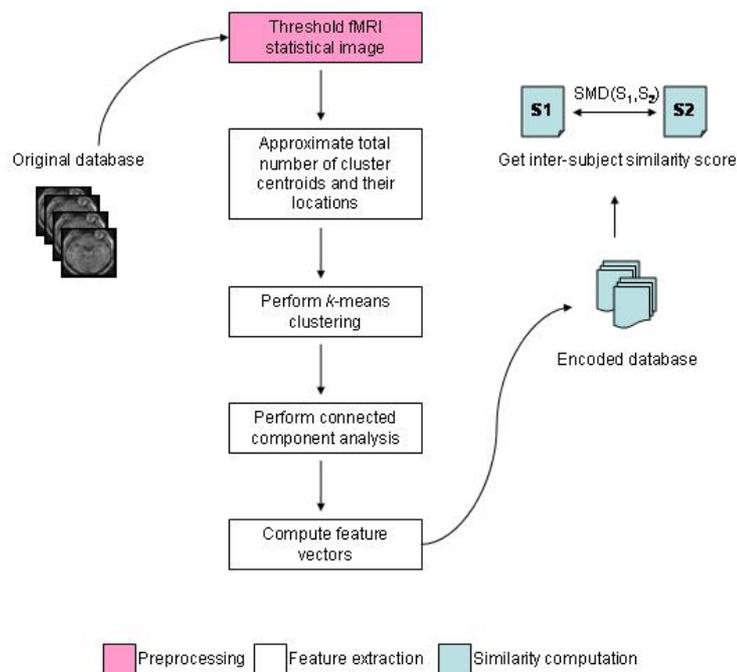


Figure 1: Our methodology from preprocessing, feature extraction to computing the similarity scores.

## 2.1 Preprocessing of the Images

Our goal is to provide a tool capable of retrieving similar fMRI statistical maps given a database of such maps. For demonstration purposes, we used SPM t-contrast maps [1]. Our t-contrast maps are 3D images of the brain with each voxel representing the difference in the mean neuronal activation between two tasks performed by the same subject: task A versus task B. For system testing purposes, we restricted our analysis to those voxels exhibiting more activation for task A than for task B.

A typical fMRI image (including t-contrast maps) has thousands of voxels pertaining to the brain. Among these voxels only a small subset contain task specific information. Mitchell et. al. [11] explored ways to best identify such voxels and found that methods that select the top n most active voxels discriminate better between one task and another given the same subject. A voxel is considered “active” if it has a significantly higher activity level when the subject is performing a given task compared to when the subject is at rest.

Thus, for each contrast map in the database we remove all voxels with activation values less than or equal to zero. Among those voxels that are retained, we further threshold them such that we retain the top X percent of activated voxels. For our system X ranged from 1 to 10 in steps of 1.

## 2.2 Feature Extraction

Next we represent each resulting contrast map with a set of feature vectors. Each such vector defines a spatially distinct region in that 3D image. First, we approximate the total number of cluster centroids given our data using subtractive clustering [12]. The approximated centroids then serve as initial cluster centers for k-means clustering.

It is important for this dataset that the regions we obtain remain spatially distinct. Biologically, neurons activate in clusters in response to a specific task. Voxels within each cluster tend to exhibit similar activation levels. We thus perform connected component analysis on the resulting k-means clusters in order to create spatially distinct regions.

Finally, we define each region using six properties: the region centroid, the region area, the average activation value for all the voxels within that region, the variance of those activation values, the average distance of each voxel within that region to the region’s centroid, and the variance of the voxel’s distance to the region’s centroid. The shape and size of the brains in the database differ from one subject to the other. We thus mapped each brain into a standard stereotaxic space [1] and extracted those feature properties accordingly.

## 2.3 Similarity Measure Computation

We present two methods for determining the similarity between two fMRI 3D images: the Summed Minimum Distance (SMD) and the Spatially Biased Summed Minimum Distance (spatial SMD).

### 2.3.1 Summed Minimum Distance (SMD)

At this point, each brain contains a set of spatially distinct regions (represented by feature vectors) that are defined by the properties listed above. The basic similarity between a query brain and the other brains in the database is calculated using the Summed Minimum Distance (SMD):

$$\text{SMD} = \frac{\text{Q-to-T Score} + \text{T-to-Q Score}}{2} \quad (1)$$

$$\text{Q-to-T Score} = \frac{\sum_{r \in Q} \min_{s \in T} d_E(r, s)}{N_Q} \quad (2)$$

$$\text{T-to-Q Score} = \frac{\sum_{s \in T} \min_{r \in Q} d_E(s, r)}{N_T} \quad (3)$$

between the query brain  $Q$  and the target brain  $T$ . For every feature vector  $s$  in  $Q$  we calculate the Euclidean distance  $d_E(s, r)$  between  $s$  and every feature vector  $r$  in  $T$  and retain the minimum distance.

Then we sum the minimum distances and divide the sum by the total number  $N_Q$  of feature vectors in the query brain to obtain a query-to-target score. We perform the same procedure in the opposite direction to obtain a target-to-query score. The average of the query-to-target score and the target-to-query score is the SMD between the query and the target.

### 2.3.2 Spatial Biased Summed Minimum Distance (Spatial SMD)

As we described above, the locations of the activated cluster in relation to the brain anatomy and the cognitive task play a significant role in determining how a neuroscientist scores a pair of fMRI images. Spatial SMD attempts to mimick this aspect of the human scoring process by incorporating information about the brain anatomy into the similarity measure. The subject’s brain is divided into the following coarse anatomic regions: the frontal lobe, the parietal lobe, the occipital lobe, the temporal lobe, the cerebellum, the left hemisphere and the right hemisphere (see Figure 2). Note that the first four regions are contained within the last two regions. All seven regions were extracted using the WFU PickAtlas tool [13, 14].

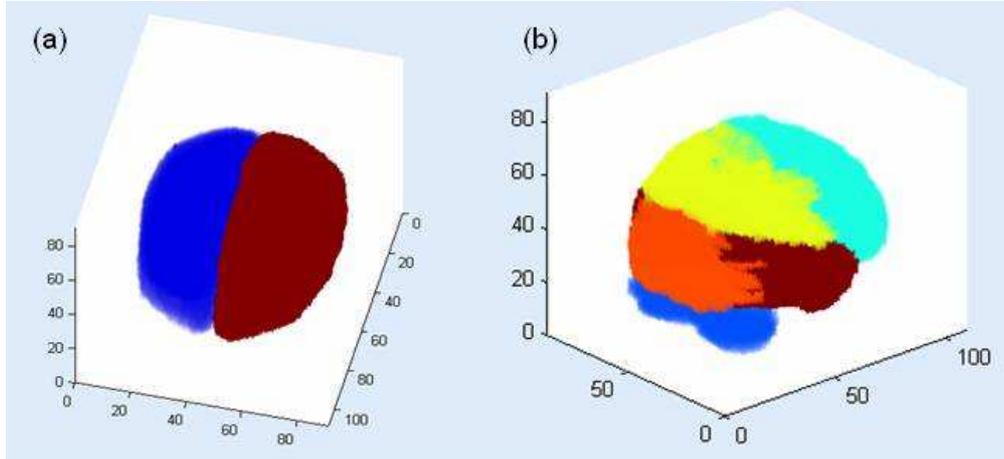


Figure 2: An illustration of the brain regions used when computing the spatial SMD. (a) The two hemispheres: left hemisphere (blue) and right hemisphere (brown) (b) The five inter-hemispheric regions: the frontal lobe (turquoise), the parietal lobe (yellow), the temporal lobe (brown), the occipital lobe (burnt orange) and the cerebellum (blue).

Spatial SMD is an extension of SMD in that the same procedure outlined for SMD is also used for the spatial SMD with one major difference: a feature vector from brain  $Q$  can only be matched with a feature vector from brain  $T$  if both are in the same anatomic region. For example, given a feature vector  $r$  in region  $P$  of brain  $Q$ , feature vector  $s$  in brain  $T$  is the best match for  $r$  only if  $s$  is in region  $P$  of brain  $T$  and has the minimum distance to  $r$  of all vectors in that region. If no match is found, then the matching score is incremented by  $\text{maxDist}$ , a constant.

### 2.3.3 Normalized Euclidean Distance for Both SMD and Spatial SMD

As explained in section 2.3.1, we initially computed the inter-feature distances using the Euclidean distance measure. Figure 3 shows that the individual units of our feature vectors (i.e. the properties for each cluster) are not isotropic. Some properties such as the values for the average activation for each cluster tend to concentrate near zero, while others such as the coordinates of the centroid do not show a similar pattern. The resulting similarity score between pairs of fMRI images in this dataset will be heavily influenced by the coordinates of the centroid if we don't account for these feature property differences. Consequently, we provide an option to replace the Euclidean distance with its normalized version [15] given below:

$$d^2(\theta^{(i)}, \theta^{(j)}) = \sum_{k=1}^K \frac{|\theta_k^{(j)} - \theta_k^{(i)}|^2}{var\theta_k} \quad (4)$$

$$var\theta_k = \sum_{m=1}^M |\theta_k^{(i)} - m_k|^2 \quad (5)$$

$$m_k = \frac{1}{M} \sum_{m=1}^M \theta_k^{(i)} \quad (6)$$

Given two feature vectors  $i$  and  $j$  from a database  $\theta$  of feature vectors, the normalized Euclidean distance between  $i$  and  $j$  is the absolute difference between the values of their corresponding feature units  $k$  weighed by the reciprocal of the variance of  $k$  across  $\theta$ . Again, each feature unit represents a unique property of the given activation cluster.

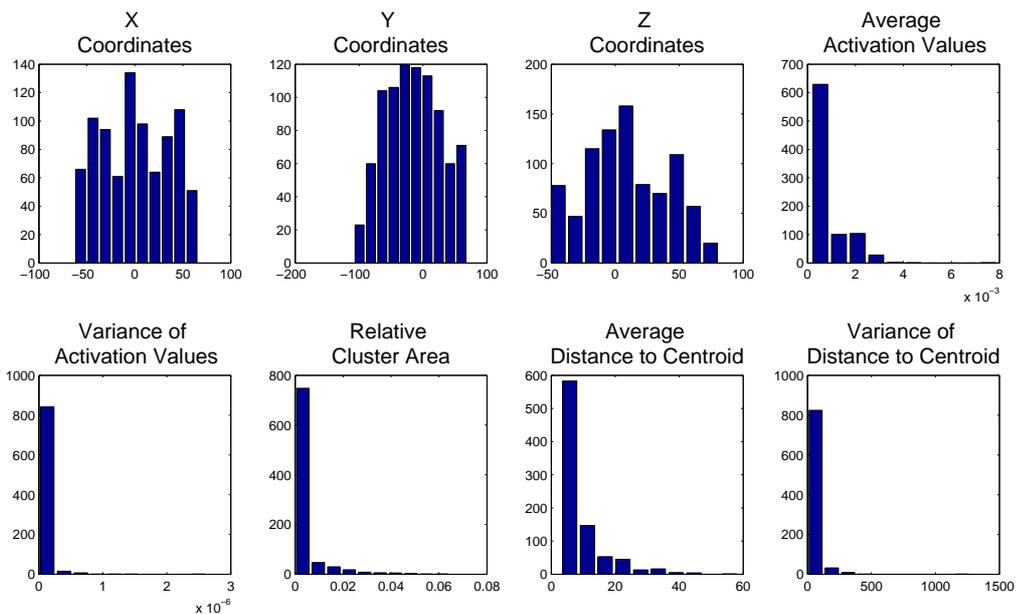


Figure 3: Histograms representing the distribution of the raw values for each region property (or feature vector unit). The corresponding feature property for each histogram is marked at the top of each figure.

## 2.4 User Interface

We have developed a user interface for our inter-subject similarity method. Briefly, given a query fMRI image and a set of user-selected parameters, the interface calls on the similarity-

based retrieval system to compute the similarity score between the query image and all other images in the database. It then returns a list of the best N matches to the query, where the value of N is selected by the user. Figure 4 is the entry page. Here the user decides what image to use as a query and the level of the threshold. S/he also can preview the query image in one of the standard neurological/radiological views: sagittal, axial, or coronal. The figure shows an axial view of the query image. Afterwards, the user continues to choose other parameters for the similarity measure as shown in Figure 5. These parameters are the type of similarity measure to use, what part of the brain to focus on, how many targets to return, and how much to weigh each feature property.

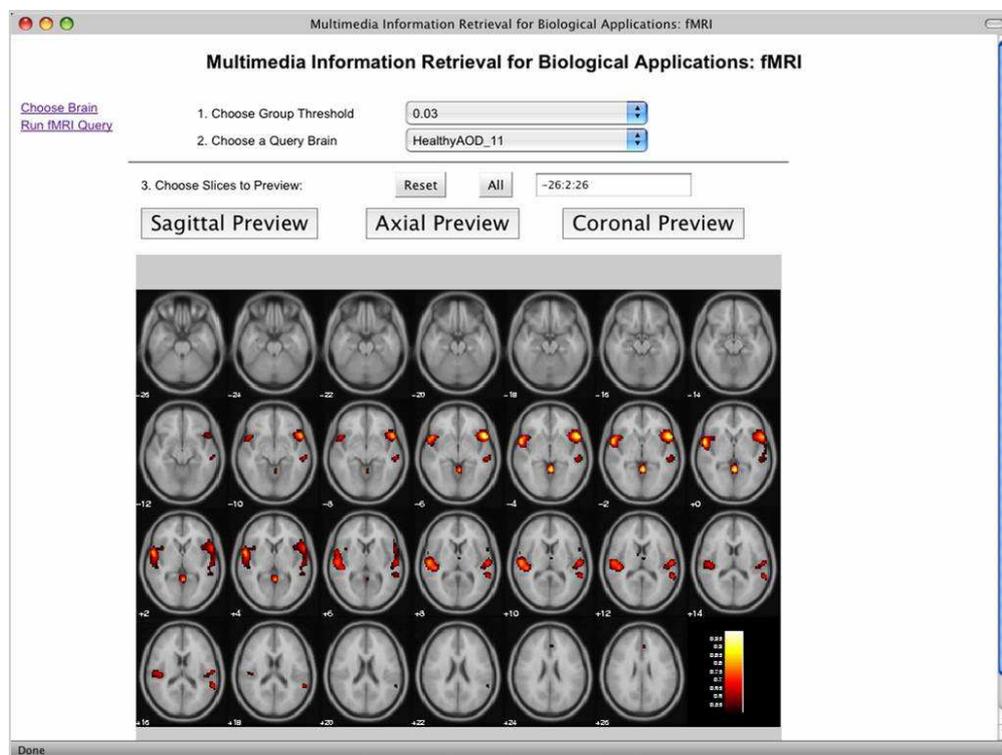


Figure 4: A snapshot of the similarity retrieval tool in which the user can choose the fMRI image for the query and the thresholding level. This page also enables the user to preview the chosen query image in one of three orthogonal views: axial, sagittal, and coronal. The axial preview is shown.

After computing the similarity, the system returns a list of target images ranked according to their similarity scores as shown in Figure 6. The closer the score is to zero, the more similar that target is to the query. The user can then choose one image from that list in order to view its activated regions in relation to the corresponding regions in the query brain. The target and query image are displayed as shown in Figure 7. Regions within the query and target brain that were matched have the same color. Regions that found no correspondences

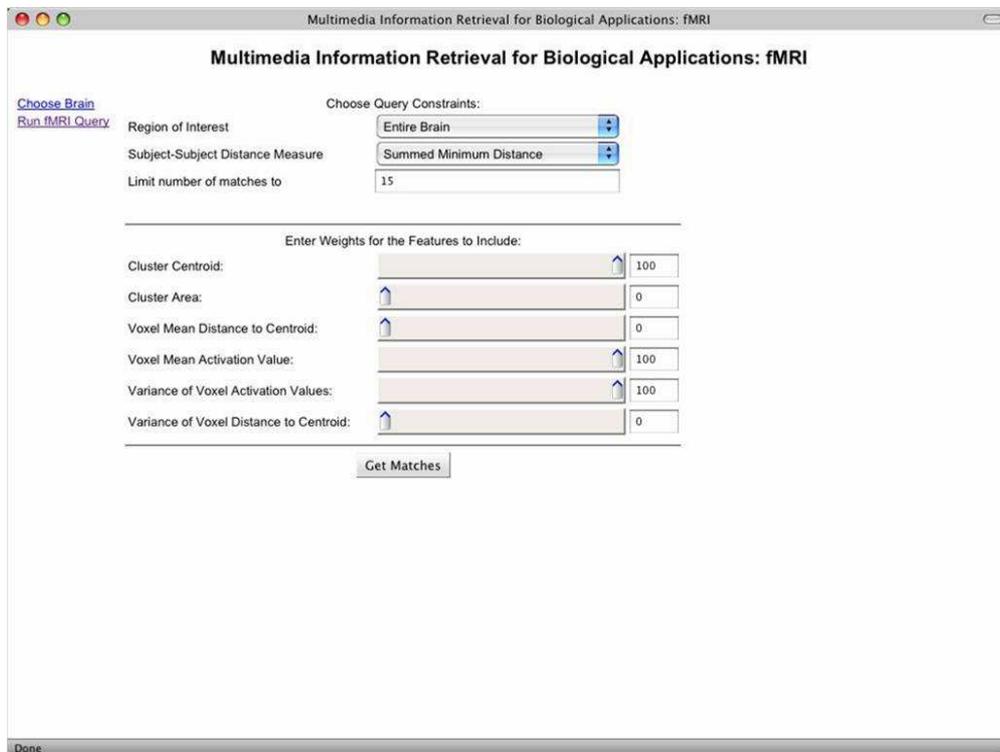
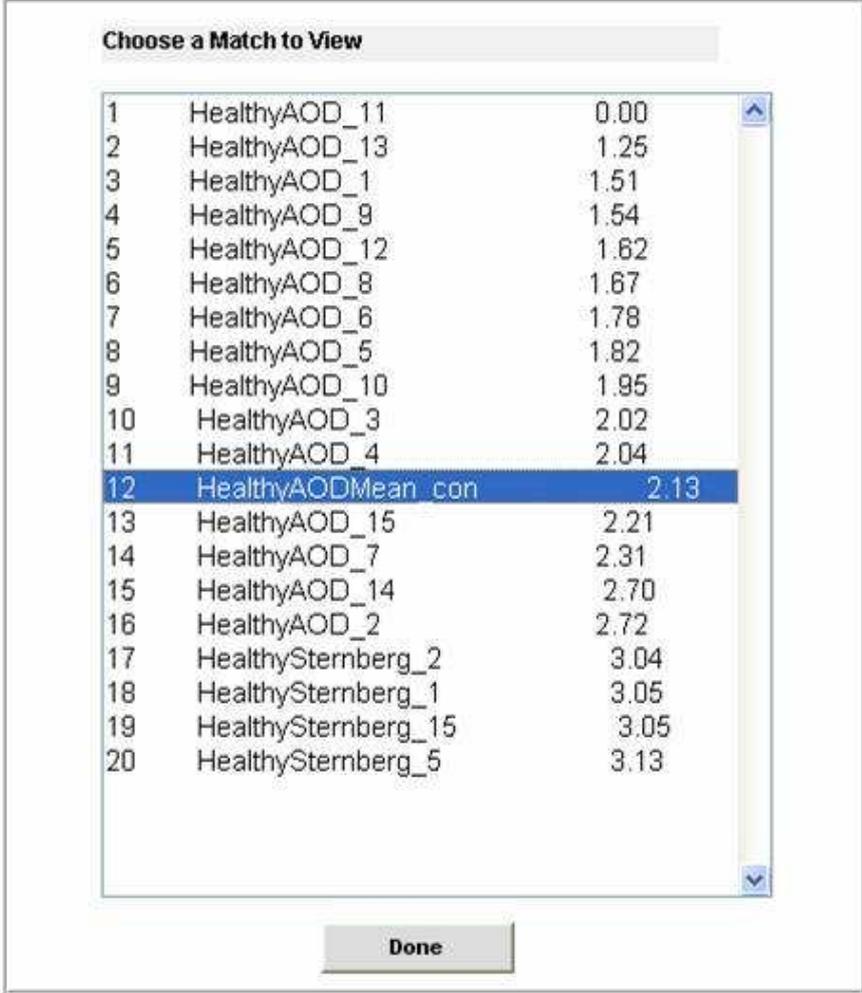


Figure 5: A snapshot of the page where the user chooses the feature properties, type of similarity measure, the brain region to focus on during the similarity computations, and the number of retrievals to return.

between the two images are not shown.



Index	Item Name	Distance
1	HealthyAOD_11	0.00
2	HealthyAOD_13	1.25
3	HealthyAOD_1	1.51
4	HealthyAOD_9	1.54
5	HealthyAOD_12	1.62
6	HealthyAOD_8	1.67
7	HealthyAOD_6	1.78
8	HealthyAOD_5	1.82
9	HealthyAOD_10	1.95
10	HealthyAOD_3	2.02
11	HealthyAOD_4	2.04
12	HealthyAODMean_con	2.13
13	HealthyAOD_15	2.21
14	HealthyAOD_7	2.31
15	HealthyAOD_14	2.70
16	HealthyAOD_2	2.72
17	HealthySternberg_2	3.04
18	HealthySternberg_1	3.05
19	HealthySternberg_15	3.05
20	HealthySternberg_5	3.13

Figure 6: Results of a query. The query image was HealthyAOD\_11, which scored a distance of 0.0 from itself. Most of the first 20 results are from the same AOD group.

### 3 Experiments and Results

We used a total of 42 t-contrast maps from subjects performing three distinct fMRI experiments: Auditory Oddball [16], Sternberg Working Memory [16], and Face Recognition [17]; details are given in Table 1. We expected the activation patterns from these three tasks to be spatially distinct, because the brain regions responsible for these tasks are spatially different.

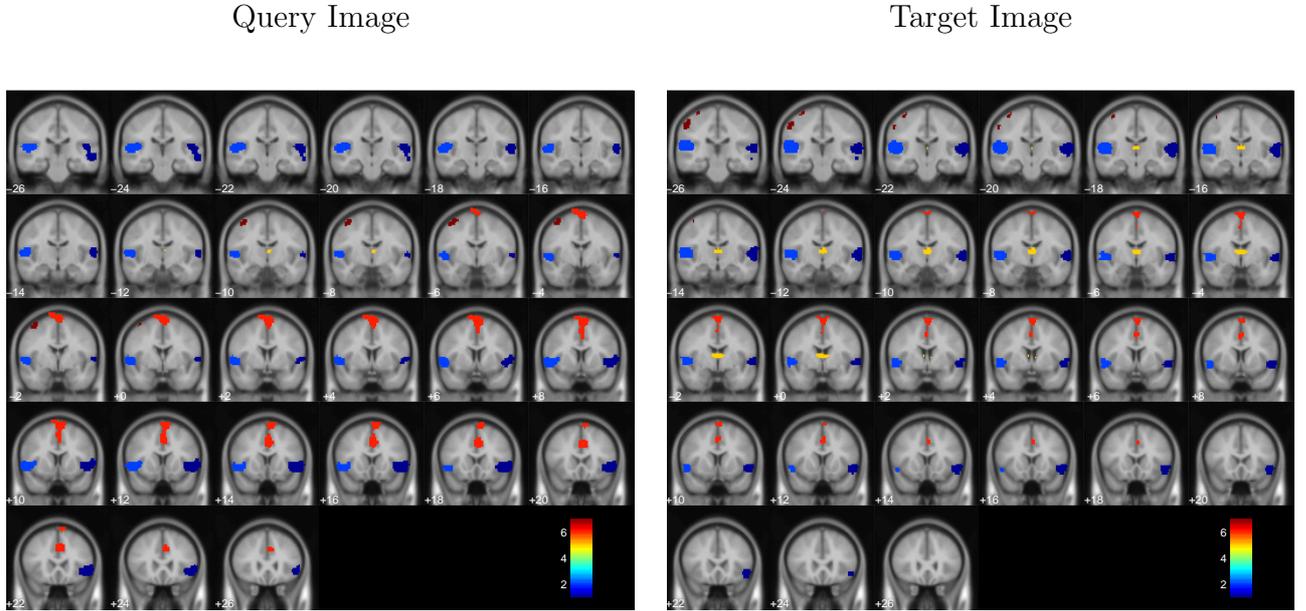


Figure 7: After the user has selected a target image to preview, the system displays the activation patterns of both that image (right) and the query image (left). Regions that were matched have the same color.

### 3.1 Evaluation

We evaluated the retrieval performance of our method in three separate sets of experiments. In the first set of experiments, we used a random effects model (RFX) as the query for each group. An RFX model of a group of fMRI images represents a very conservative average activation brain, which only incorporates those activated voxels that are present in all members of the group. We computed the RFX models for the AOD, Sternberg and Checkerboard groups by performing a second level t-test on their t-contrast maps using the SPM5 software package [18]. Our logic was as follows: if the query is an RFX contrast map from a given group and the group is relatively homogenous, then the majority of target contrast maps should come from the same group as that of the query. In the second set of experiments, we used each individual 3D image in each group as a query image. In the third set of experiments we used different combinations of the feature properties for the similarity measure computation.

In all three sets of experiments we utilized the retrieval score [19] below to score the retrieval results:

Table 1: Description of Data Sets

Data Sets	Cognitive Process	Task A vs Task B
Auditory Oddball (AOD) ( 15 subjects )	Recognize out of place sound	recognize a new tone vs recognize the same repeating tone [16]
Sternberg Working Memory (SB) ( 15 subjects )	Recognize memorized alphabets	recognize memorized alphabets vs recognize non-memorized alphabets [16]
Face Recognition (Checkerboard) ( 12 subjects )	Recognize human faces	recognize human faces vs recognize a black and white checkerboard [17]

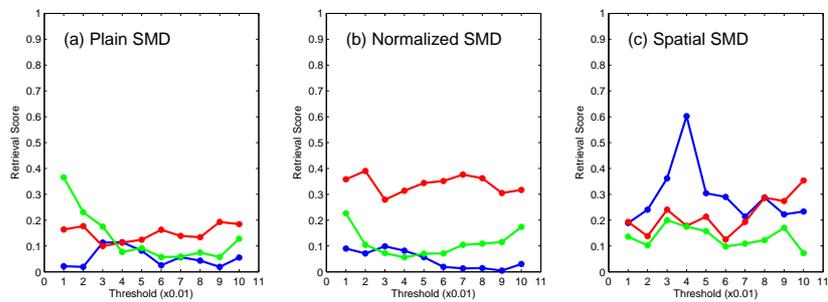
$$\text{Retrieval Score} = \frac{1}{N \times N_{rel}} \left( \sum_{i=1}^{N_{rel}} R_i - \frac{N_{rel}(N_{rel} + 1)}{2} \right) \quad (7)$$

$N$  is the total number of brains in the dataset,  $N_{rel}$  is the total number of brains within the query’s group, and  $R_i$  is the rank at which the  $i^{th}$  relevant brain is retrieved. A perfect retrieval where all the relevant brains are retrieved before any others would receive a score of 0, while the worst retrieval where all the relevant brains are retrieved after all the others would receive a score of 1. In the case of random retrieval, the score is approximately 0.5.

### 3.2 RFX Model Retrieval Experiments

The retrieval scores using the three RFX models as queries are shown in Figure 8 a-c. We used all features with equal weights for these retrievals. For the similarity measure in Figure 8a we used the plain SMD distance (SMD with Euclidean distance). In Figure 8b we used the normalized SMD (SMD with normalized Euclidean distance), while in Figure 8c we used the spatial SMD with normalized Euclidean distance (henceforth spatial SMD). Each data set exhibited its own characteristics when examined with the others. In general, considering all three models, the plain SMD (Figure 8a) had the best performance.

### RFX Model Retrieval Scores



### Mean Individual Retrieval Scores

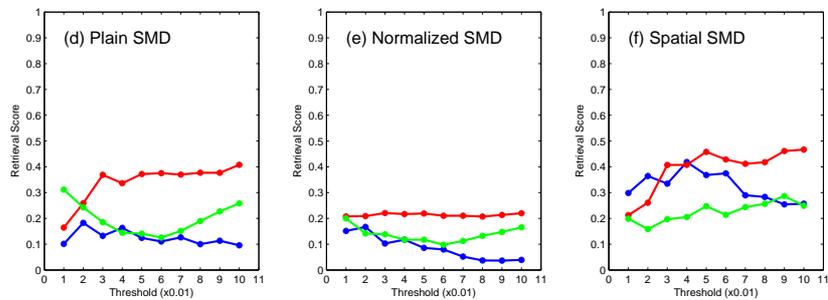


Figure 8: Graphs of the retrieval scores. The top row contains the retrieval scores for each of the RFX models, while the bottom row contains the mean individual retrieval scores for each group. Each group is color coded as follows: Sternberg (green), AOD (red), and Checkerboard (blue). The similarity measures used were (a and d) plain SMD, (b and e) normalized SMD, (c and f) spatial SMD.

### 3.3 Individual Brain Query Experiments

Ordinarily, queries from real users would be individual brains, not RFX models. Figure 8 d-f shows the mean retrieval scores for individual brain queries using all the feature properties in the similarity measure. Unlike the RFX models, the mean retrieval scores across each group exhibited similar characteristics for all three data sets. The normalized SMD (Figure 8e) proved superior to both the plain SMD (Figure 8d) and the spatial SMD (Figure 8f).

We also explored the variances in the retrieval scores across each group. Figure 9 shows the means and standard deviations of the individual retrieval scores for the normalized SMD distance (a-c) and the spatial SMD (d-f), using all the features in the similarity computation. Using normalized SMD, the variance of the group retrieval scores for the Checkerboard and Sternberg images (Figure 9a and Figure 9c, respectively) were quite small compared to those for the AOD group (Figure 9b). The same pattern emerged when using the spatial SMD (Figure 9 d-f) despite the fact that the mean retrieval scores for each group in the latter set were higher than with the normalized SMD distance.

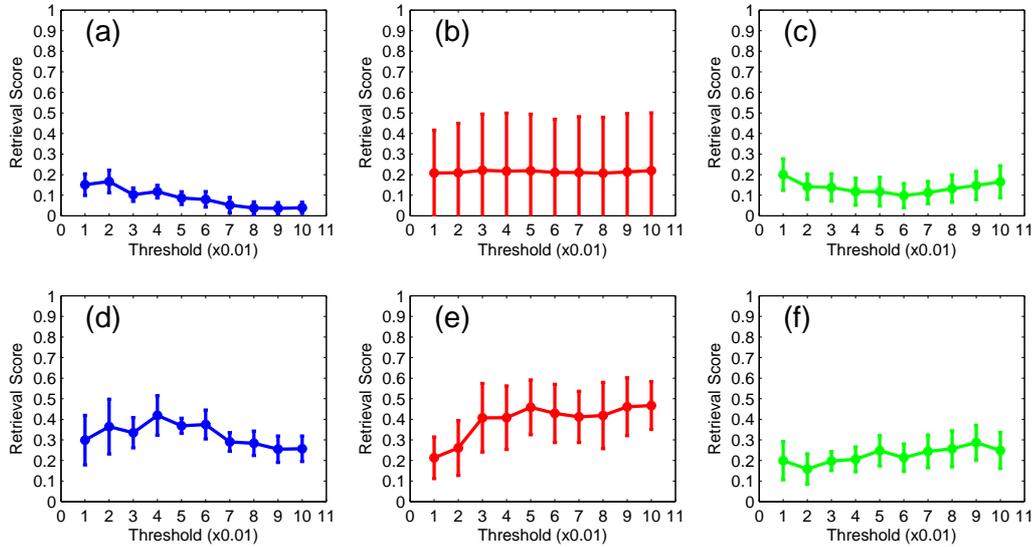


Figure 9: The mean and standard deviations of the retrieval scores across each group. For the top row, we used the normalized SMD distance and for the bottom row we used the spatial SMD.

### 3.4 Feature Selection Experiments

Figure 10 and Figure 11 represent the retrieval scores when each of the RFX models was used as a query, and different sets of feature properties were used. When using one feature property at a time, only the “centroid position” and the “average activation value” had low retrieval scores relative to the other four features as shown in Figure 10. Figure 11 illustrates that combining the “centroid position” with “average activation values” gave the best retrieval scores relative to the other combinations depicted in that figure. That figure also shows that combining the “cluster area”, “distance to centroid”, and “variance of distances to centroid” resulted in higher retrieval scores compared to the other combinations.

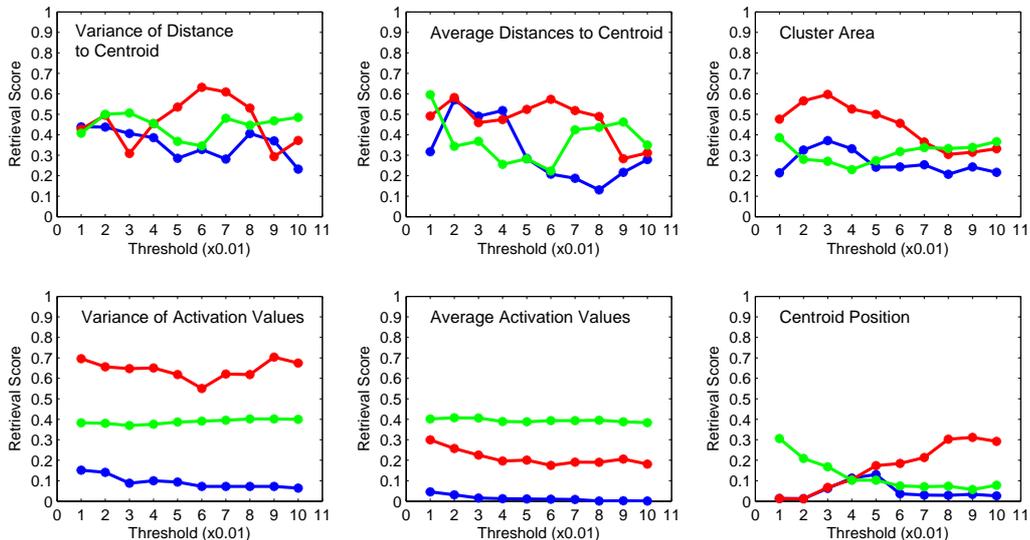


Figure 10: Graphs of the retrieval scores of the three RFX models with only a single feature in similarity measure. Each feature used is indicated at the top of its graph.

### 3.5 Experiments for Testing Group Homogeneity

In order to visualize the similarity structure of individual images across the entire dataset, we generated an all-against-all similarity score matrix for each threshold. We used the entire data set, the normalized SMD, and the features “centroid location” and “average activation value” for computing those matrices. Then we performed multidimensional scaling analysis (MDS) [20] on each matrix. MDS takes an  $n \times n$  similarity matrix and projects it into a lower dimension such that the inter-point distances in that matrix are retained.

Figure 12 illustrates the MDS projections. The dataset neatly separates into three groups.

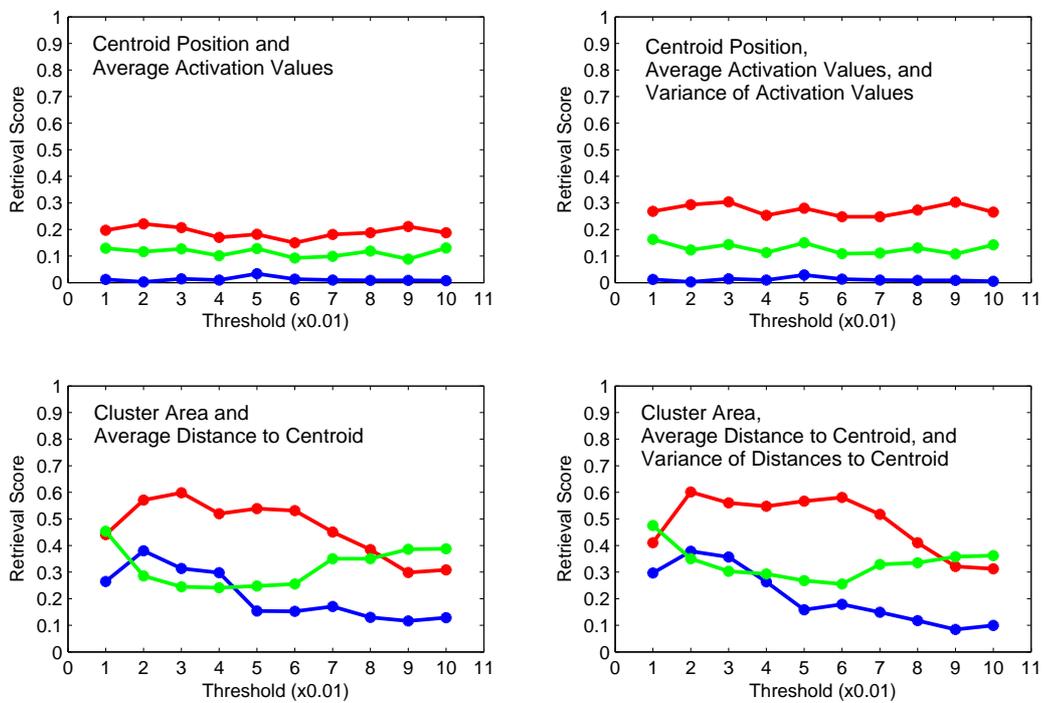


Figure 11: Graphs of the retrieval scores of the three RFX models with combinations of features used for the similarity measure. Each of the feature subsets used is indicated at the top of its graph.

The locations of the RFX models suggest they are not centrally located in their corresponding groups.

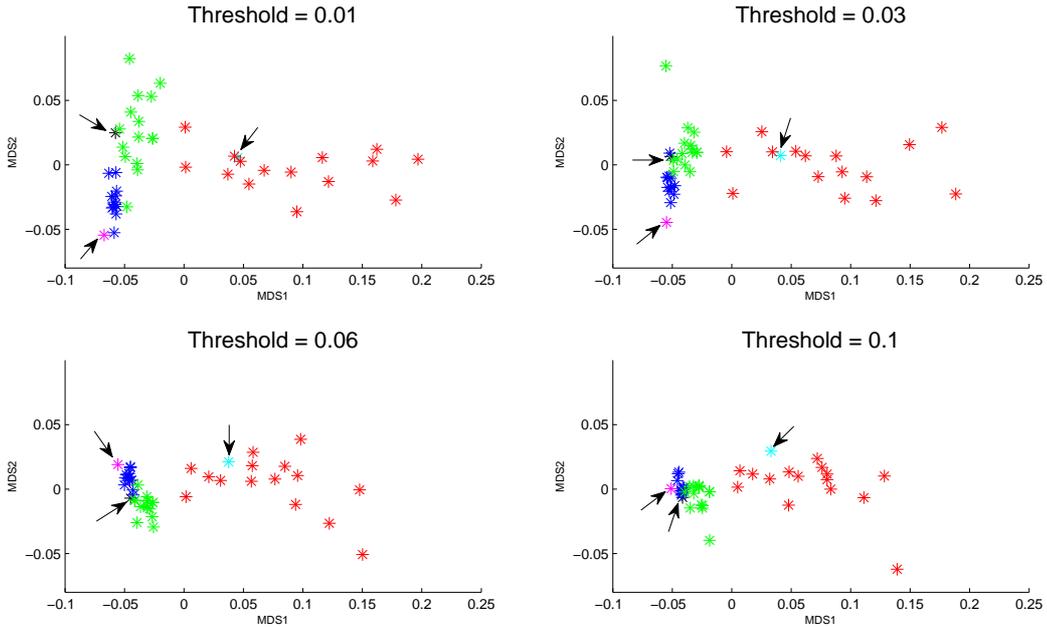


Figure 12: The MDS projections of the all-against-all similarity score matrix of the three groups (AOD: red, Sternberg: green, and Checkerboard: blue) at selected threshold as indicated at the top of each projection. The arrows point to the RFX models for each group. Each RFX model has a unique color: cyan is the AOD RFX, pink is the Checkerboard RFX, and black is the Sternberg RFX

## 4 Discussion

The results show that given a query fMRI image, one can retrieve similar images using an automated computer system. Different features of the activation clusters provide different levels of discriminatory power to the similarity measures. We expected the locations of the activated regions in each group to be different, because the three cognitive tasks are known to activate distinct areas of the brain [17] [16]. Thus, using only the “centroid location” as a feature for the similarity measure should result in low retrieval scores across each group. Figure 10 confirmed our expectation by presenting retrieval scores close to zero for all three RFX models when using only the “centroid location”.

On the other hand, feature properties such as the “variance of distance to centroid”, “average

distance to centroid”, and “cluster area” perform poorly when used alone. In Figure 10 the retrieval scores of the RFX models when using each of these feature properties in isolation were close to random (i.e. 0.5). Our experiments on feature combination, as illustrated in Figure 11, suggest that for our data sets, these three features need to be combined with some of the stronger discriminatory features such as the “centroid location” and the “average activation value”. It is possible however, that for a different group of fMRI images where the shape of the activation area is critical, the “variance of distance to centroid”, the “average distance to centroid”, and the “cluster area” will prove more discriminatory for inter-subject similarity scoring.

Surprisingly, the “average activation value” and the “variance of activation value” created three distinct retrieval score patterns in Figure 10. We examined the distribution of the activation values within each group and discovered that the intensity of voxel activation values was also different across these three cognitive tasks. The AOD group had the most “active” voxels compared to the other two groups, while the Checkerboard group had the least. Differences in the experimental conditions is one explanation of the different voxel activation values. The data sets come from three different experiments performed by two different research groups using different sets of equipments and subjects. It is possible that in a subset of these groups, there is a persistent experimental artifact that caused the voxel intensity values to be so different from the other group(s). A second explanation is that noisier data, or data collected over a shorter time, will have lower activation.

Regardless of what the source of these differences is, these group voxel distributions are a good representation of what does happen in reality when scientists use other statistical maps such as T-maps, Z-scores, and/or F-scores. In this project we only used contrast maps, because we did not have the variance of the test statistics along each voxel. If that information were available, we would have calculated a corresponding t-statistics, generated a t-map for each brain, and thresholded the t-maps to retain only those voxels that were statistically significantly activated using some measure that accounted for multiple comparison. For a detailed mathematical treatment of this process refer to Friston et. al. [1]. The resulting thresholded T-maps for each group would have a similar characteristics to the dataset we used in this project: the number of “active” voxels and the distribution of voxel intensity values would differ from subject to subject and sometimes from group to group.

Both Figure 10 and Figure 11 show a crude way to identify the features with good group discrimination ability. It is crude, because we simply included or excluded certain feature properties. An interesting future study would be to provide weights for the feature properties and automate the process of picking the optimal set of weights for the feature vector given a specific fMRI dataset and a user with specific goals. Our expert fMRI users do want to use the features in this way for their own experiments.

Another interesting observation from this study is how the value of the threshold affects the retrieval score. Consider Figure 8f for instance where an increase in the threshold value in-

creases the retrieval score of the RFX models. In other words, raising the threshold decreases the performance by reducing the discriminatory power of the similarity measure. Studies have shown that the results one gets from fMRI studies heavily depend on the nature of the threshold used [21]. Yet, given the thousands of voxels that an fMRI image has and the fact that most of those voxels represent noise, a threshold of some sort must be applied if one is to obtain meaningful results. It is imperative then that our user interface provide an option for users to pick a threshold for each similarity computation. Our current users select a threshold that is based on what fraction of voxels they want to retain within a given data. However, given the t-statistics of that data, they would be able to select a confidence level, which is more meaningful.

Apart from the choice of threshold level or feature properties, our interface has many other parameters that the user needs to select (see Figures 4 and 5). We opted for this design due to the nature of fMRI images. There is hardly a set of parameters that guarantee best retrieval scores across different subjects and even datasets. The data sets we used were just for system testing purposes. They were explicitly chosen for their distinct cognitive tasks. In reality, we don't expect subjects performing different tasks to be that easily separable. As a matter of fact, even the same group of subjects performing two different tasks may not be easily separable because of the differences in inter-subject cognitive strategies when performing a given task, brain sizes, and brain anatomic locations. Thus, we decided to give the user as much autonomy as possible in selecting the parameters for the similarity measure based on what s/he thinks may improve the resulting retrievals given the available data set.

Besides providing a convenient way for fMRI image analysis to identify similar images, our similarity measure is also useful in identifying whether the members of a given group have homogeneous activation patterns. Figure 12, which is an MDS projection of the similarity scores of each subject in the data set with all other subjects therein, is a good example of this. Using the first two projections (MDS1 and MDS2) we examined what subjects are located closer to each other across the three groups.

We learned that members of the AOD group are more scattered compared to members of the other two groups. The distances between the red stars are bigger compared to the distances between the blue stars or the green stars. We also found that the RFX models are not always good representations of their corresponding group activation patterns. In Figure 12, the pink star, which represents the RFX model of the Checkerboard group is not located at the center of that group. Nonetheless, this figure demonstrates that our similarity measure has the ability to successfully identify fMRI images with similar activation patterns; members of the same group were consistently close to each other with minimal overlap.

## 5 Conclusion and Future Work

In this study we proposed and evaluated a method for retrieving similar fMRI statistical images given a query fMRI statistical image. The method extracts spatially distinct regions from each image after thresholding its constituent voxels. Each region is defined by a feature vector that contains the region centroid, the region area, the average activation value for all the voxels within that region, the variance of those activation values, the average distance of each voxel within that region to the region’s centroid, and the variance of the voxel’s distance to the region’s centroid. The similarity between two images is obtained in two ways: the average summed minimum distance weighted by the inverse of the number of components from the query to the target and from the target to the query and its spatially biased version.

We demonstrated that our method is sensitive to similarities in brain activation patterns from members of the same data set. From our experiments we found that the normalized SMD obtained the best results compared to the other two similarity measures, when used with individual queries. We also learned that using that similarity measure with the features “centroid location” and “average activation value” (individually or in combination), maximizes the performance of the normalized SMD. Lastly, we were able to identify the similarity structure of the entire data set using those two features and the normalized SMD.

In our future work, we plan to improve the spatial SMD measure based on the observation that some activations may lie on the borders of brain regions, instead of lying fully within one region or another. We also want to automate the process of selecting an optimal set of weights for each feature unit in the feature vector. We plan to test this method on datasets that do not have a well-known brain activation pattern in an effort to aid scientific discoveries. In conjunction with collaborating brain researchers, we plan to use the method to study differences between groups affected with particular maladies, such as autism, and control groups. Lastly, we want to extend this method so it can be used to cluster statistical images from a set of subjects under the same experimental conditions into distinct groups based on the similarity of their activation patterns.

## 6 Acknowledgments

This research was supported by the National Science Foundation under Grant Number DBI-0543631. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## References

- [1] Friston, K., Ashburner, J., Kiebel, S., Nichols, T., and Penny, W., eds., [*Statistical Parametric Mapping: The Analysis of Functional Brain Images*], Academic Press (2007).
- [2] Bai, B., *Feature extraction and matching in content-based retrieval of functional Magnetic Resonance Images (fMRI)*, PhD thesis, Rutgers University (2007).
- [3] Shapiro, L. G., Atmosukarto, I., Cho, H., Lin, H. J., Ruiz-Correa, S., and Yuen, J., “Similarity-based retrieval for biomedical applications,” in [*Case-Based Reasoning on Images and Signals*], Perner, P., ed., 355–387, Springer-Verlag (2008).
- [4] Brammer, M. J., Bullmore, E. T., Simmons, A., Williams, S. C., Grasby, P. M., Howard, R. J., Woodruff, P. W., and Rabe-Hesketh, S., “Generic brain activation mapping in functional magnetic resonance imaging: a nonparametric approach,” *Magnetic Resonance Imaging* **15**, 763–770 (1997).
- [5] Kherif, F., Poline, J., Meriaux, S., Benali, H., Flandin, G., and Brett, M., “Group analysis in functional neuroimaging: selecting subjects using similarity measures,” *NeuroImage* **20**, 2197–2208 (2003).
- [6] McNamee, R. L. and Lazar, N. A., “Assessing the sensitivity of fMRI group maps,” *NeuroImage* **22**, 920–931 (2004).
- [7] Seghier, M. L., Lazeyras, F., Pegna, A. J., Annoni, J., and Khateb, A., “Group analysis and the subject factor in functional magnetic resonance imaging: analysis of fifty right-handed health subjects in a semantic language task,” *Human Brain Mapping* **29**, 461–477 (2008).
- [8] Bai, B., Kantor, P., Shoukoufandeh, A., and Silver, D., “fMRI brain image retrieval based on ICA components,” in [*ENC ’07: Proceedings of the Eighth Mexican International Conference on Current Trends in Computer Science*], 10–17, IEEE Computer Society, Washington, DC, USA (2007).
- [9] Wang, Q., Megalooikonomou, V., and Kontos, D., “A medical image retrieval framework,” *2005 IEEE Workshop on Machine Learning for Signal Processing*, 233–238 (2005).
- [10] Zhang, J. and Megalooikonomou, V., “An effective and efficient technique for searching for similar brain activation patterns,” *Proc. 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 428–431 (2007).
- [11] Mitchell, T. M., Hutchinson, R., Pereira, F., and Wang, X., “Learning to decode cognitive states from brain images,” *Machine Learning* **57**, 145–175 (2004).

- [12] Yager, R. and Filev, F., “Generation of fuzzy rules by mountain clustering,” *Intelligent and Fuzzy Systems* **2**, 209–219 (1994).
- [13] Maldjian, J. A., Laurienti, P. J., and Burdette, J. B., “Precentral gyrus discrepancy in electronic versions of the Talairach atlas,” *NeuroImage* **21**, 450–455 (2004).
- [14] Maldjian, J. A., Laurienti, P. J., Burdette, J. B., and Kraft, R. A., “An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets,” *NeuroImage* **19**, 1233–1239 (2003).
- [15] Reed, T. R. and du Buf, J. M. H., “A review of recent texture segmentation and feature extraction techniques,” *CVGIP: Image Underst.* **57**(3), 359–372 (1993).
- [16] Calhoun, V. D., Adali, T., Kiehl, K. A., Astur, R., Pekar, J. J., and Pearlson, G. D., “A method for multitask fMRI data fusion applied to schizophrenia,” *Human Brain Mapping* **27**, 1–13 (2006).
- [17] Henson, R., Shallice, T., Gorno-Tempini, M., and Dolan, R., “Face repetition effects in implicit and explicit memory tests as measured by fMRI,” *Cerebral Cortex* **12**, 178–186 (2002).
- [18] SPM5, “<http://www.fil.ion.ucl.ac.uk/spm/>,” (2005).
- [19] Muller, H., Marchand-Maillet, S., and Pun, T., “The truth about corel-evaluation in image retrieval,” *Proc. International Conference on Image and Video Retrieval* **2383**, 38–49 (2002).
- [20] Seber, G. A. F., [*Multivariate Observations*], Wiley (1984).
- [21] Loring, D. W., Meador, K. J., Allison, J. D., Pillai, J. J., Lavin, T., Lee, G. P., Balan, A., and Dave, V., “Now you see it and now you donot: statistical and methodological considerations in fMRI,” *Epilepsy and Behavior* **3**, 539–547 (2002).