ASHDIN
PUBLISHING

Research Article

A Similarity Retrieval Tool for Functional Magnetic Resonance Imaging Statistical Maps

Rosalia Tungaraza,[1] Jinyan Guan,[4] Linda Shapiro,[1] James Brinkley,[2] Jeffrey Ojemann,[3] and Joshua Franklin[2]

[1]*Computer Science and Engineering, University of Washington, Box 352350, Seattle, WA 98195, USA*
[2]*Biological Structure, University of Washington, Box 357420, Seattle, WA 98195, USA*
[3]*Neurosurgery, University of Washington, Box 359300, Seattle, WA 98195, USA*
[4]*Computer Science, Winona State University, Box 5838, Winona, MN 55987, USA*
Address correspondence to Rosalia Tungaraza, rltungar@uw.edu

Abstract *Objective.* We propose a method for retrieving similar functional magnetic resonance imaging (fMRI) statistical images given a query fMRI statistical image. *Method.* 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 (SMD) of their constituent feature vectors. *Results and conclusion.* Our method is sensitive to similarities in brain activation patterns from members of the same data set. Using a subset of the features such as the centroid location and the average activation value (individually or in combination), maximized the sensitivity of our method. We also identified the similarity structure of the entire data set using those two features and the SMD.

Keywords functional magnetic resonance imaging; content-based retrieval; brain activation patterns; statistical parametric mapping; 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) [13] 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 10 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 [1] 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 data sets 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 [17]. 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 not only 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 [3]. Kherif et al. [6] 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 [10]. 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 toward computing the average activated brain. McNamee et al. [10] 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 was not affected much by outliers and conservative because it downplayed patterns of activation that were not common to all subjects within the data set. 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 data set.

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 intersubject 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 [16], bipartite matching of fMRI-ICA spatial maps [2], encoding the original fMRI image into wavelet coefficients [19] or code-blocks [21], and calculating an RV coefficient of two fMRI images [6]. The similarity between two images in all of these methods is based on the spatial locations and extent of the activated regions.

Surprisingly, in our observation of a neuroscientist scoring fMRI images, we noticed that the features he emphasized when comparing a pair of images were not always the same. For instance, he did not always only consider the spatial locations of the activated regions. 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, making it difficult to translate the scoring process 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. Ultimately, the similarity between two images in our work depended on what combination of these features the user chose to compare those images. This property of our system distinguished it from the above-mentioned methods for retrieving similar fMRI images. Rather than adapting one of the above-mentioned matching techniques (e.g., RV coefficient, percentage overlap), we developed our own similarity metric. This was done to accommodate the fact that our users could use any combination of the available features for retrieving similar images in an fMRI database. The above-mentioned techniques only work if the user choses the location and extent of the activated region as the features for retrieving similar fMRI images. If the user chooses these two features (i.e., location and extent of the activated region) for fMRI image retrieval, then our system performs equivalently to methods such as the percentage overlap.

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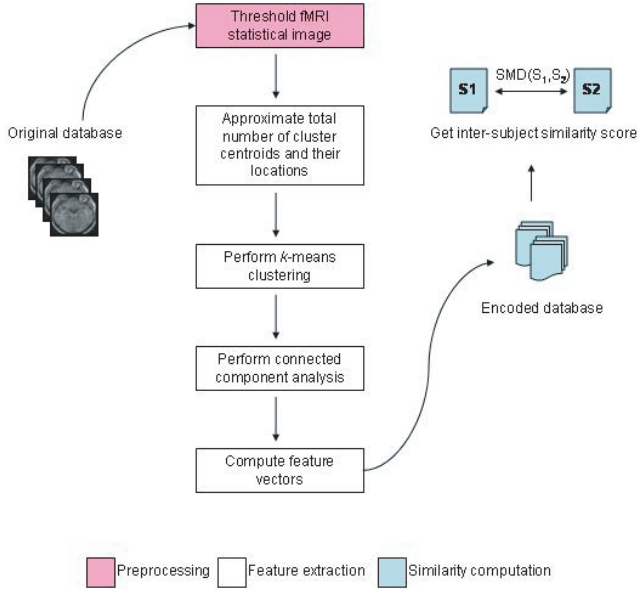
Figure 1: Our methododology from preprocessing, feature extraction to computing the similarity scores.

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.

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 [13]. 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 in 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 [20]. The approximated centroids then serve as initial cluster centers for k-means clustering. We perform both the subtractive and k-means clustering using the voxels that were retained after thresholding the contrast maps as described in Section 2.1. Each voxel is represented by a 4D vector (x, y, z, v), where x, y, z is the spatial location and v is the activation value of that voxel respectively.

It is important for this data set 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 [13] 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. 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 SMD:

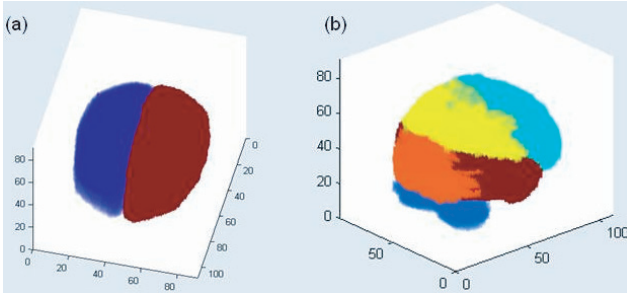$$\text{SMD} = \frac{Q\text{-to-}T \text{ Score} + T\text{-to-}Q \text{ Score}}{2} \qquad (1)$$

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).

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

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

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 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 [8,9].

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 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 data set will be heavily influenced by the coordinates of the centroid if we do not account for these feature property differences. Consequently, we provide an option to replace the Euclidean distance with its normalized version [14] given below:

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

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

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

Given two feature vectors i and j from a database θ 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 θ. Again, each feature unit represents a unique property of the given activation cluster.

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: saggital, axial, or coronal. The figure shows an axial view of the query image. Afterward, the user continues to choose other parameters for the similarity measure as shown in Figure 5. These parameters are the type of
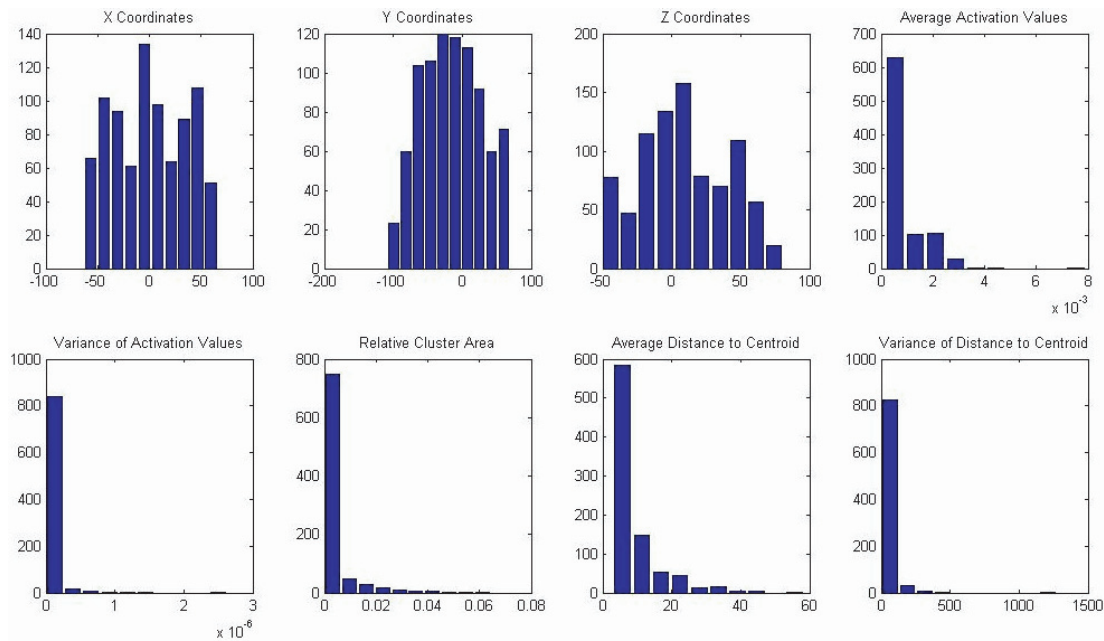
Figure 3: Histograms representing the distribution of the raw values for each region property (or feature vector unit). The corresponding feature property for each histogram is marked at the top of each figure.
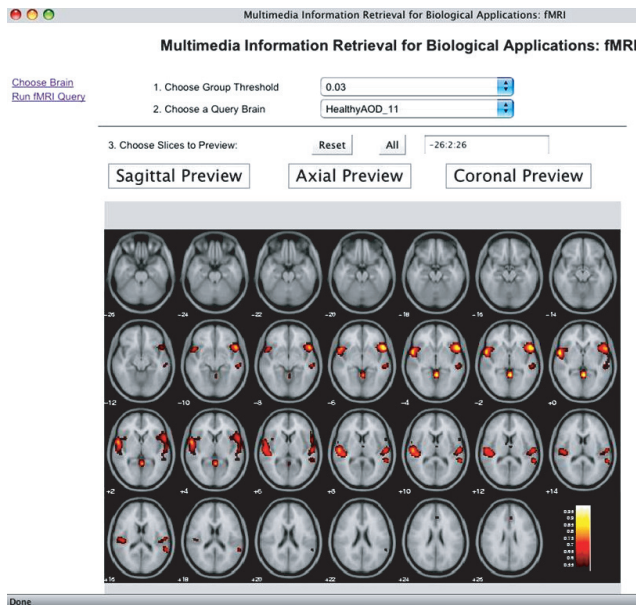

Figure 4: A snapshot of the similarity retrieval tool in which the user can choose the fMRI image for the query and the thresholding level. This page also enables the user to preview the chosen query image in one of three orthogonal views: axial, sagittal, and coronal. The axial preview is shown.
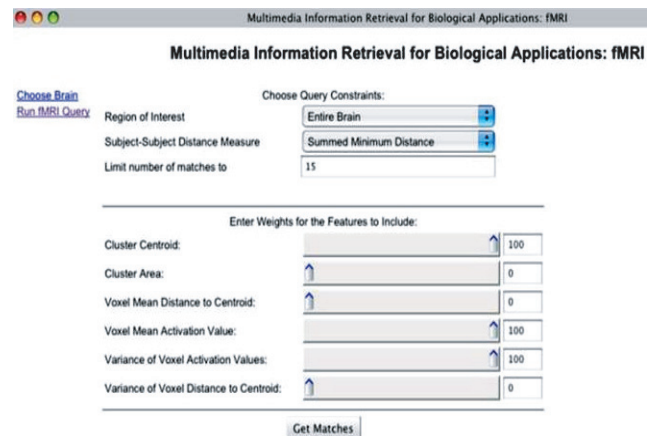

Figure 5: A snapshot of the page where the user chooses the feature properties, type of similarity measure, the brain region to focus on during the similarity computations, and the number of retrievals to return.

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.

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 between the two images are not shown.
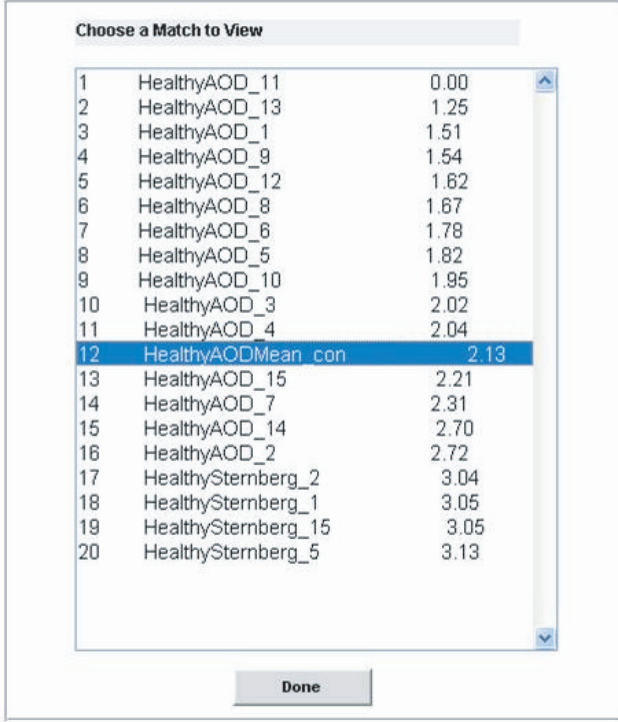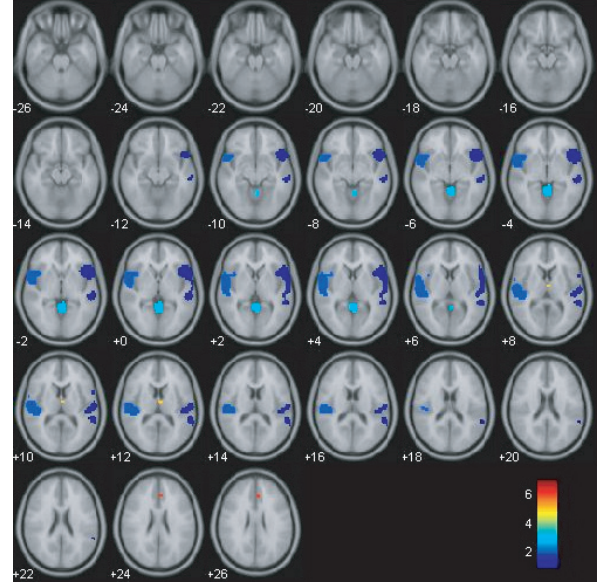
(a) Query image.

(b) Target image.

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.

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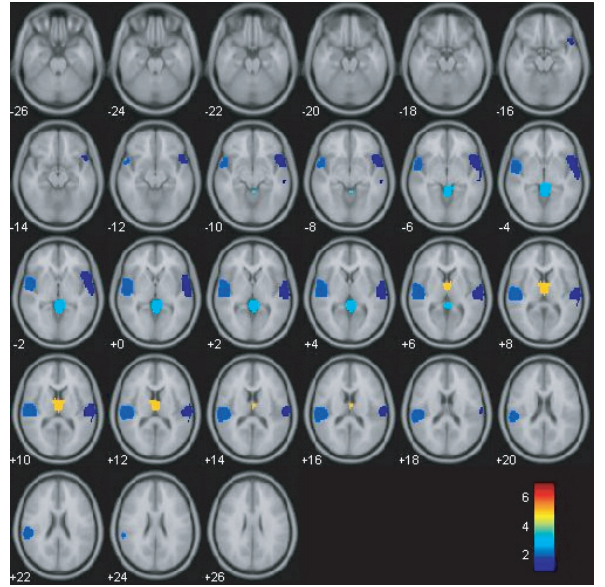
3. Experiments and results

We used a total of 42 t-contrast maps from subjects performing three distinct fMRI experiments: auditory oddball [4], sternberg working memory [4], and face recognition [5]; 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.

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 homogeneous, 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 [12] below to score the retrieval results:

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

Table 1: Description of data sets.

| Data sets	Cognitive process	Task A versus task B
Auditory oddball (AOD) (15 subjects)	Recognize out-of-place sound	Recognize a new tone versus recognize the same repeating tone [4]
Sternberg working memory (SB) (15 subjects)	Recognize memorized alphabets	Recognize memorized alphabets versus recognize non-memorized alphabets [4]
Face recognition (Checkerboard) (12 subjects)	Recognize human faces	Recognize human faces versus recognize a black-and-white checkerboard [5]
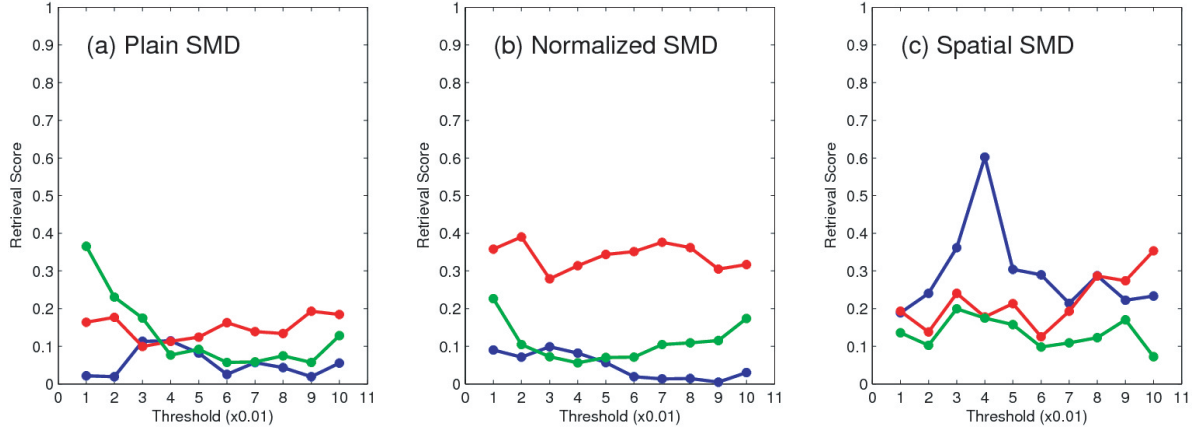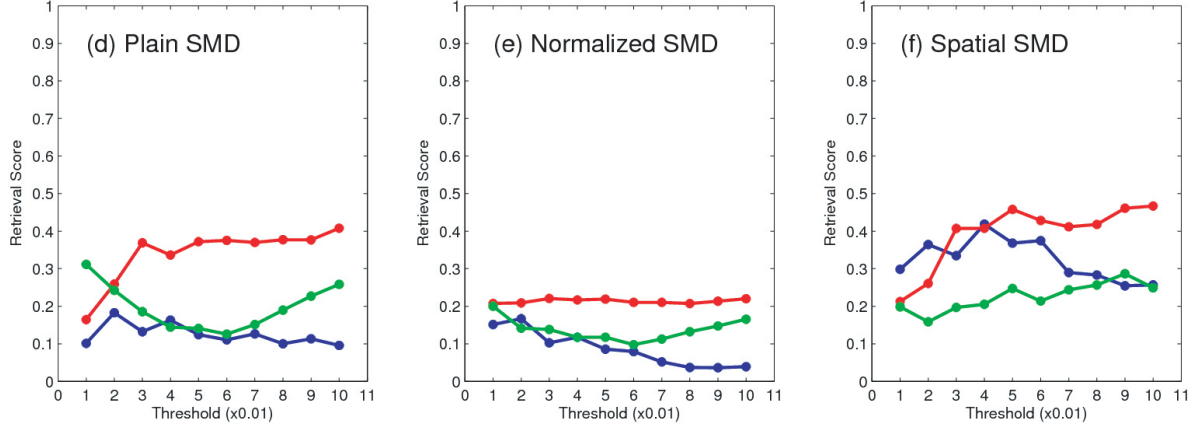
Figure 8: Graphs of the retrieval scores. The top row contains the retrieval scores for each of the RFX models, while the bottom row contains the mean individual retrieval scores for each group. Each group is color coded as follows: Sternberg (green), AOD (red), and Checkerboard (blue). The similarity measures used were (a) and (d) plain SMD, (b) and (e) normalized SMD, (c) and (f) spatial SMD.

N is the total number of brains in the data set, N_{rel} is the total number of brains within the query's group, and R_i is the rank at which the ith relevant brain is retrieved. A perfect retrieval where all the relevant brains are retrieved before any others would receive a score of 0. The larger the retrieval score, the more dissimilar the top ranking brains are to the query brain.

3.2. RFX model retrieval experiments

The retrieval scores using the three RFX models as queries are shown in Figures 8(a), 8(b), and 8(c). We used all features with equal weights for these retrievals. For the similarity measure in Figure 8(a), we used the plain SMD distance (SMD with Euclidean distance). In Figure 8(b), we used the normalized SMD (SMD with normalized Euclidean
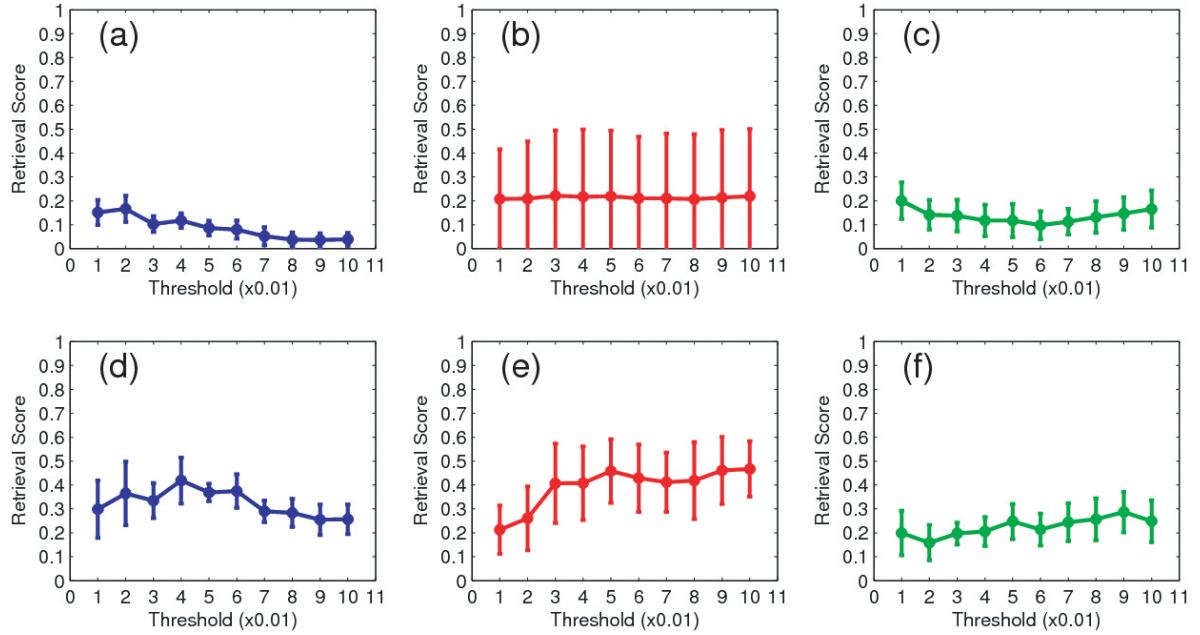
Figure 9: The mean and standard deviations of the retrieval scores across each group. For the top row, we used the normalized SMD distance and for the bottom row we used the spatial SMD. Each group is color coded as follows: Sternberg (green), AOD (red), and Checkerboard (blue).

distance), while in Figure 8(c), 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 8(a)) had the best performance.

3.3. Individual brain query experiments

Ordinarily, queries from real users would be individual brains, not RFX models. Figures 8(d), 8(e), and 8(f) show 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 8(e)) proved superior to both the plain SMD (Figure 8(d)) and the spatial SMD (Figure 8(f)).

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 (Figures 9(a)–9(c)) and the spatial SMD (Figures 9(d)–9(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 (Figures 9(a) and 9(c), respectively) were quite small compared to those for the AOD group (Figure 9(b)). The same pattern emerged when using the spatial SMD (Figures 9(d), 9(e), and 9(f)) despite the fact that the mean retrieval scores for each group in the latter set were higher than with the normalized SMD distance.

3.4. Feature selection experiments

Figures 10 and 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.

3.5. Experiments for testing group homogeneity

In order to visualize the similarity structure of individual images across the entire data set, 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) [15] on each matrix. MDS takes an nxn 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 data set neatly separates into three groups. The locations of the RFX models suggest they are not centrally located in their corresponding groups.
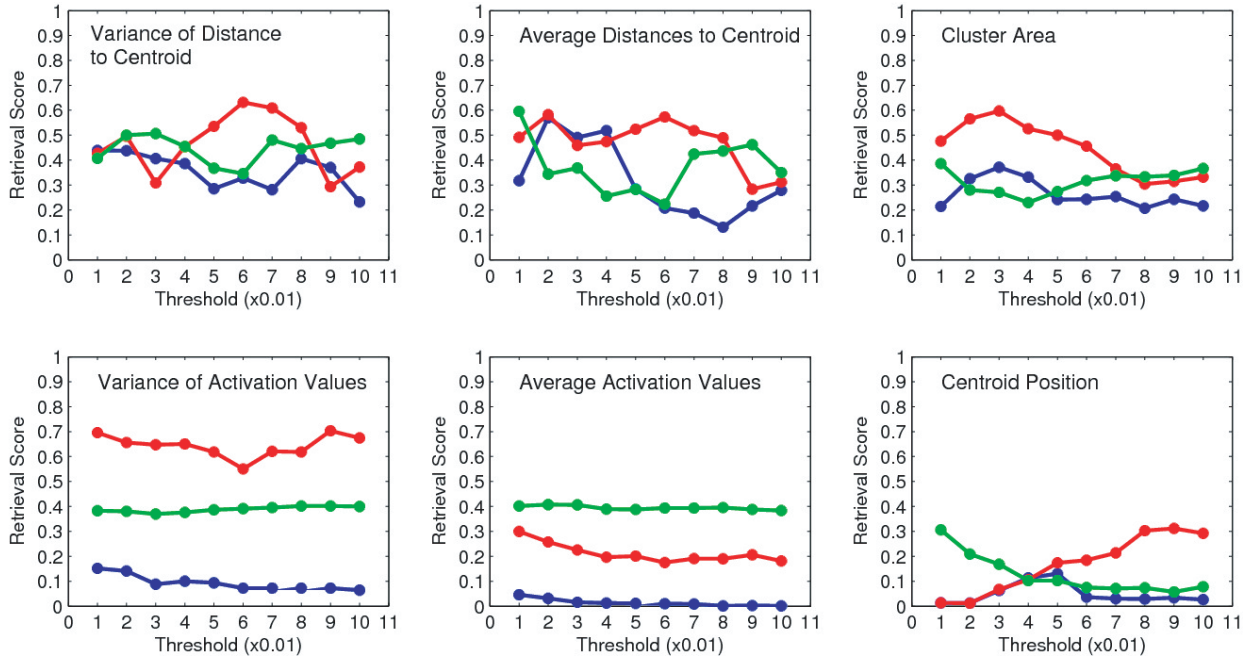
Figure 10: Graphs of the retrieval scores of the three RFX models with only a single feature in similarity measure. Each feature used is indicated at the top of its graph and each group's RFX model is color coded as follows: Sternberg (green), AOD (red), and Checkerboard (blue).

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. Given the cognitive tasks explored in our experimental data, we expected some features to perform better than others and that we would not be able to explore all possible real queries. For instance, we expected the locations of the activated regions in each group to be different. The three cognitive tasks are known to activate distinct areas of the brain [4,5]. 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. Figure 10 shows poor retrieval scores of the RFX models when using each of these feature properties in isolation. 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."

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.

Another observation we glean from Figure 10 is that the retrieval scores for the "average activation values" are nearly constant as one increases the number of voxels used for the retrieval task. This is contrary to the behavior of the graphs when only the "centroid position" was used as the feature. We expected this property because fMRI contrast maps exhibit the strongest signal (i.e., voxel with highest activation value) at the center of an activated region while the surrounding voxels within the same region tend to have lower activation values. Thus, increasing the number of voxels via using different levels of thresholding as depicted in Figure 10 may not affect the average activation value because the activation values of the initial voxels tend to dominate the statistics. On the other hand, the "centroid
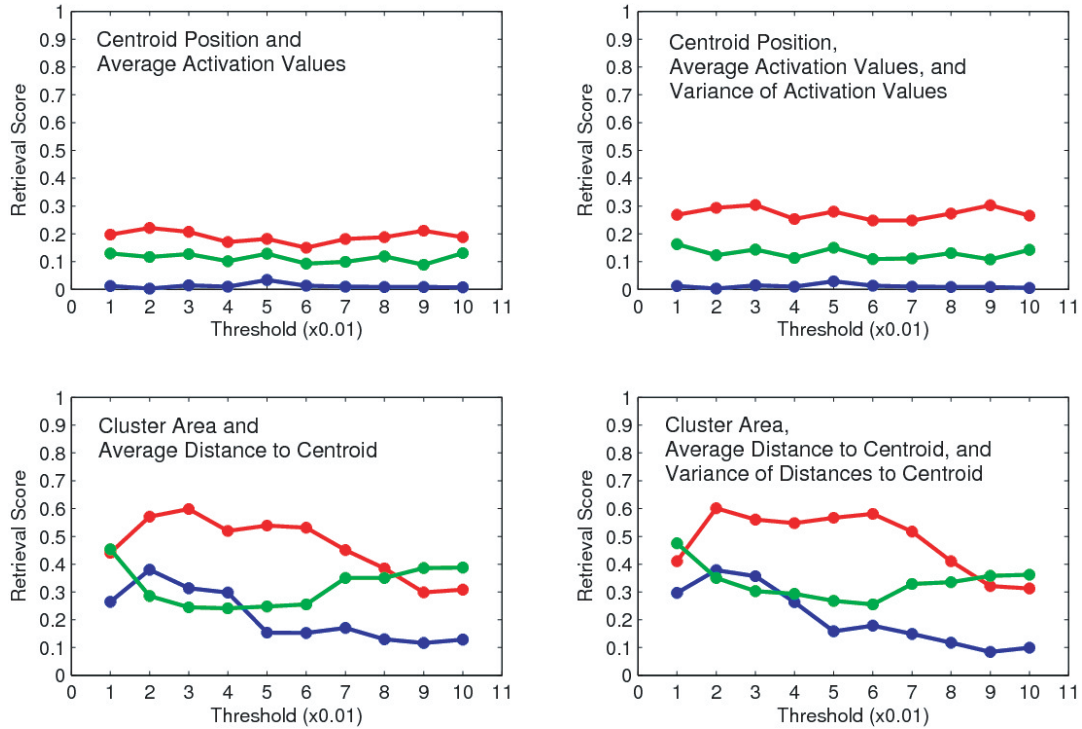
Figure 11: Graphs of the retrieval scores of the three RFX models with combinations of features used for the similarity measure. Each of the feature subsets used is indicated at the top of its graph and each group's RFX model is color coded as follows: Sternberg (green), AOD (red), and Checkerboard (blue).

position" in our work is based on the mean x, y, z location of all the voxels in a given activated region. Increasing the number of voxels for that region changes the size of the region, which can shift the region's "centroid position" resulting in changes of the retrieval score.

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. [13]. The resulting thresholded t-maps for each group would have a similar characteristics to the data set 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 Figures 10 and 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 data set 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 8(f) for instance, where an increase in the threshold value increases 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 [7]. 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
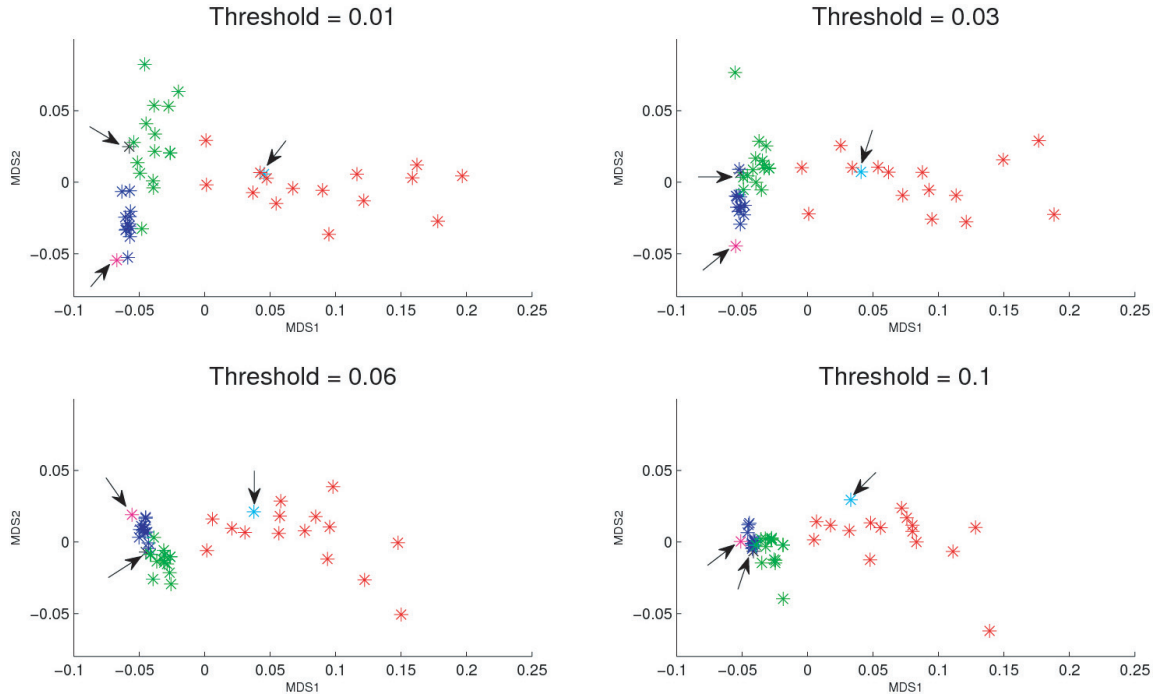
Figure 12: The MDS projections of the all-against-all similarity score matrix of the three groups (AOD: red, Sternberg: green, and Checkerboard: blue) at selected threshold as indicated at the top of each projection. The arrows point to the RFX models for each group. Each RFX model has a unique color: cyan is the AOD RFX, pink is the Checkerboard RFX, and black is the Sternberg RFX.

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 data sets. The data sets we used were just for system testing purposes. They were explicitly chosen for their distinct cognitive tasks. In reality, we do not 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 data sets 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.

References

[1] B. Bai, *Feature extraction and matching in content-based retrieval of functional magnetic resonance images*, PhD thesis, Rutgers University, New Brunswick, NJ, 2007.

[2] B. Bai, P. Kantor, A. Shokoufandeh, and D. Silver, *fMRI brain image retrieval based on ICA components*, in Proc. 8th Mexican International Conference on Current Trends in Computer Science (ENC '07), Michoacan, Mexico, 2007, 10–17.

[3] M. J. Brammer, E. T. Bullmore, A. Simmons, S. C. Williams, P. M. Grasby, R. J. Howard, et al., *Generic brain activation mapping in functional magnetic resonance imaging: a nonparametric approach*, Magn Reson Imaging, 15 (1997), 763–770.

[4] V. D. Calhoun, T. Adali, K. A. Kiehl, R. Astur, J. J. Pekar, and G. D. Pearlson, *A method for multitask fMRI data fusion applied to schizophrenia*, Hum Brain Mapp, 27 (2006), 598–610.

[5] R. N. Henson, T. Shallice, M. L. Gorno-Tempini, and R. J. Dolan, *Face repetition effects in implicit and explicit memory tests as measured by fMRI*, Cereb Cortex, 12 (2002), 178–186.

[6] F. Kherif, J. B. Poline, S. Mériaux, H. Benali, G. Flandin, and M. Brett, *Group analysis in functional neuroimaging: selecting subjects using similarity measures*, Neuroimage, 20 (2003), 2197–2208.

[7] D. W. Loring, K. J. Meador, J. D. Allison, J. J. Pillai, T. Lavin, G. P. Lee, et al., *Now you see it, now you don't: statistical and methodological considerations in fMRI*, Epilepsy Behav, 3 (2002), 539–547.

[8] J. A. Maldjian, P. J. Laurienti, and J. H. Burdette, *Precentral gyrus discrepancy in electronic versions of the Talairach atlas*, Neuroimage, 21 (2004), 450–455.

[9] J. A. Maldjian, P. J. Laurienti, R. A. Kraft, and J. H. Burdette, *An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets*, Neuroimage, 19 (2003), 1233–1239.

[10] R. L. McNamee and N. A. Lazar, *Assessing the sensitivity of fMRI group maps*, Neuroimage, 22 (2004), 920–931.

[11] T. M. Mitchell, R. Hutchinson, R. S. Niculescu, F. Pereira, X. Wang, M. Just, et al., *Learning to decode cognitive states from brain images*, Machine Learning, 57 (2004), 145–175.

[12] H. Müller, S. Marchand-Maillet, and T. Pun, *The truth about corel – evaluation in image retrieval*, in Image and Video Retrieval, M. S. Lew, N. Sebe, and J. P. Eakins, eds., vol. 2383 of Lecture Notes in Computer Science, Springer-Verlag, Berlin, 2002, 38–49.

[13] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, eds., *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Academic Press, London, 2007.

[14] T. R. Reed and J. M. H. Dubuf, *A review of recent texture segmentation and feature extraction techniques*, CVGIP: Image Understanding, 57 (1993), 359–372.

[15] G. A. F. Seber, *Multivariate Observations*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York, 2004.

[16] M. L. Seghier, F. Lazeyras, A. J. Pegna, J. M. Annoni, and A. Khateb, *Group analysis and the subject factor in functional magnetic resonance imaging: analysis of fifty right-handed healthy subjects in a semantic language task*, Hum Brain Mapp, 29 (2008), 461–477.

[17] L. G. Shapiro, I. Atmosukarto, H. Cho, H. J. Lin, S. Ruiz-Correa, and J. Yuen, *Similarity-based retrieval for biomedical applications*, in Case-Based Reasoning on Images and Signals, P. Perner, ed., vol. 73 of Studies in Computational Intelligence, Springer-Verlag, Berlin, 2008, 355–387.

[18] SPM5, http://www.fil.ion.ucl.ac.uk/spm (accessed 12 August 2009), 2005.

[19] Q. Wang, V. Megalooikonomou, and D. Kontos, *A medical image retrieval framework*, in 2005 IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT, 2005, 233–238.

[20] R. Yager and D. Filev, *Generation of fuzzy rules by mountain-clustering*, Journal of Intelligent and Fuzzy Systems, 2 (1994), 209–219.

[21] J. Zhang and V. Megalooikonomou, *An effective and efficient technique for searching for similar brain activation patterns*, in Proc. 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '07), Arlington, VA, 2007, 428–431.