A generalized family of fixed-radius distribution-based distance measures for content-based fMRI image retrieval

John Novatnack [a], Nicu Cornea [b], Ali Shokoufandeh [a,*], Deborah Silver [b],
Sven Dickinson [c], Paul Kantor [d], Bing Bai [e]

[a] Department of Computer Science, Drexel University, 3141 Chestnut Str., Philadelphia, PA, United States
[b] Department of Electrical and Computer Engineering, Rutgers University, 96 Frelinghuysen Road, Piscataway, NJ, United States
[c] Department of Computer Science, University of Toronto, 6 King's College Road, Toronto, ON, Canada
[d] Department of Library and Information Science, Rutgers University, 4 Huntington Str., New Brunswick, NJ, United States
[e] Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ, United States

ARTICLE INFO

ABSTRACT

We present a family of distance measures for comparing activation patterns captured in fMRI images. We model an fMRI image as a spatial object with varying density, and measure the distance between two fMRI images using a novel fixed-radius, distribution-based Earth Mover's Distance that is computable in polynomial time. We also present two simplified formulations for the distance computation whose complexity is better than linear programming. The algorithms are robust in the presence of noise, and by varying the radius of the distance measures, can tolerate different degrees of within-class deformation. Empirical evaluation of the algorithms on a dataset of 430 fMRI images in a content-based image retrieval application demonstrates the power and robustness of the distance measures.

1. Introduction

Functional magnetic resonance imaging (fMRI) is a medical imaging technique that captures brain functionality by measuring the change in oxygen concentration in the blood that flows into different parts of the brain as the subject performs some task. Similar to other emerging imaging modalities, fMRI produces very large datasets (.5 GB) describing a single patient or subject in a time series of high resolution spatial images of the brain. As much as 200 terabytes of data per year are generated in the brain imaging aspects of cognitive studies. These are currently indexed only by sparse meta-data, such as the experimental conditions, and not by the actual content or brain activation patterns.

Indexing by these activation patterns enables the development of content-based image retrieval systems. These systems will be invaluable to scientists, who will be able to query the wealth of fMRI data by content and not rely on the human specified meta-data. Moreover, they could have profound impact on cognitive science, as previously unknown patterns in brain functionality are revealed. However, in order for these systems to be realized, accurate measures of brain similarity must first be developed. In this paper, we present a family of novel distance metrics for brain images which may be leveraged to develop a comprehensive framework for content-based brain retrieval.

Content-based image retrieval (CBIR) encompasses a broad array of general purpose techniques, the most comprehensive survey of which can be found in (Veltkamp and Tanase, 2000). Most CBIR methods compute weak image abstractions, such as color or texture histograms or appearance-based statistics (Niblack et al., 1993), while some assume that query objects have been correctly segmented and seek to encode their shape (Pentland et al., 1994). The ultimate goal of content-based image retrieval systems is invariance to within-class deformation or appearance changes, and while weak image abstractions tend to be color- or texture-specific, they do tolerate within-class shape variation and changes in viewpoint. Moreover, many such features are computed globally, without requiring figure-ground segmentation. But while such systems are designed for less structured scenes, they are, in fact, inappropriate for our domain, for their representations are too abstract to capture the subtle changes in brain activation between two conditions. For example, two brain images with the same number of activated pixels may yield similar abstract representations, e.g., histograms, despite the fact that the locations of the activation in the two images may be substantially different. Clearly, the problem of fMRI image retrieval requires a method more tailored to the problem.

Indeed, there have been efforts supporting content-based retrieval of specific fMRI data collections. In addition to

* Corresponding author. Tel.: +1 215 895 2671; fax: +1 215 895 0545.
E-mail address: ashokouf@cs.drexel.edu (A. Shokoufandeh).

content-based retrieval, there have been efforts to start collecting fMRI experimentations (Van Horn and Gazzaniga, 2002; BRAin Image Database, 2005). Currently, these only support keyword retrieval, and not content-based retrieval. Preliminary studies use a collection of Alzheimer's patients. Notable other efforts include Megalooikonomou et al. (1999), Megalooikonomou et al. (2000), Megalooikonomou et al. (2003), Lazarevic et al. (2001), Ford et al. (2001, 2002, 2003), Wang et al. (2004), Kontos et al. (2004), Saykin et al. (2002). In data mining efforts, databases of fMRI images are mined to find relationships between brain lesions and functional deficits (e.g., vision or speech deficits) (Megalooikonomou et al., 1999), or to classify patients based on fMRI activation patterns (Megalooikonomou et al., 2003; Ford et al., 2001, 2002, 2003). Also related to the efforts presented here are 3D-based content retrieval systems (such as the work of Osada et al., 2002). These are more directed towards CAD-type applications, and do not retrieve the type of 3D objects of variable density as we have here.

A robust measure of similarity between brain images plays a fundamental role in a content-based image retrieval system. Fig. 1 shows an example of such a system. First, fMRI images are processed using standard techniques to extract an activation map. Next, we apply a distance measure appropriate for matching objects of variable density, to quantify their similarity. The result is a distance measure between any two brain images, shown as a hit map in Fig. 1, that can be directly exploited in a content-based image retrieval system. The quality of the distance metric plays a pivotal role in this system, as it should support accurate discrimination between activation patterns.

The problem of matching brain images has a number of unique characteristics that affect the design of our distance metrics. First, brain images are normalized in post-processing, so that the representation of brain activation in each of the subjects occurs in roughly the same spatial location. Therefore, a metric for determining the similarity between two brains must be designed to match activation patterns which are spatially consistent. Second, fMRI images are subject to noise introduced by inaccuracies in the physical processes, the sensors or the post-processing. A metric must be robust in the presence of this noise. In order to account for both of these requirements, we have developed a novel family of brain image distance metrics. These metrics allow for the spatial localization that is critical for accurate retrieval of brain images, and they are robust in the presence of noise. Another key issue is the large amount of processing that is normally applied to fMRI

images. The processing converts the 3D + time images to one 3D statistically significant dataset, representing the areas of the brain which are relevant in the scan. If the preprocessing tools and procedures used to create a query image are not identical to those used to create a database image (the two may have been created in different laboratories, for example), then additional discrepancies between two conditions may result. It is essential that the content-based image retrieval system be able to tolerate such forms of within-class image deformation.

The format for this paper is as follows: first, a general formulation of brain matching is presented that leads to a number of efficient and robust algorithms for computing brain similarity. These algorithms can be seen as more constrained versions of algorithms previously studied in the context of graph matching and network flow. Next, in order to test the effectiveness of these algorithms, we gathered a varied set of fMRI images from cooperating laboratories. Through a comprehensive set of computational experiments we show increased performance of content-based image retrieval using our distance metrics as compared to the metrics that have been previously proposed.

2. Representation model

As stated in the introduction, fMRI images generally undergo an extensive amount of processing before analysis begins. In order to obtain a robust map of active voxels, the raw fMRI images are first preprocessed. (Note: a raw fMRI image consists of a sequence of 3D brain scans typically conducted at 2-s intervals during a cognitive experiment. Experiments can sometimes last up to 30 min or more. This results in a 4D dataset (3D + time).) The preprocessing step eliminates structurally irrelevant components, such as the skull, from the images, while correcting some of the image acquisition noise. We start our processing by applying motion correction to align all the 3D images in a time series to the same position, correcting the effects caused by small movements of the subject during the experimental run. This step is followed by the removal of voxels representing the skull and applying a temporal high pass filter to remove low frequency trends over time due, for example, to the increasing temperature of the device. The preprocessing is sometimes concluded with an optional step of applying spatial smoothing filters to control spatial noise, which we do not do. Due to differences in the shapes and sizes of individual brain images, the preprocessing is followed by a mapping of the 3D
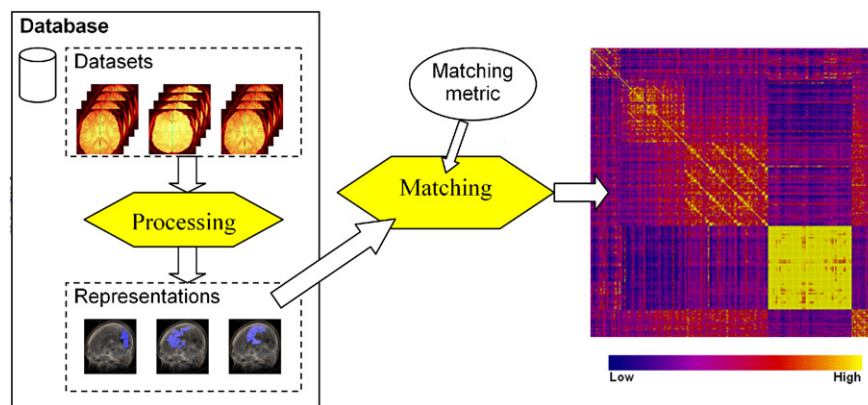

Fig. 1. A content-based image retrieval system that exploits our family of distance metrics. First, raw fMRI images are processed and thresholded to extract activation maps. Secondly, the similarity between all brain images is computed using our novel family of distance metrics. This similarity measure is visualized as a color-coded hit matrix representing how each dataset maps to the other datasets in the repository. Each row (and column) of this matrix corresponds to the similarity obtained of a fMRI image against the rest of the database. The intensity of each cell denotes the magnitude of similarity among corresponding fMRI images, i.e., yellow is used to represent high similarities, and blue represents low similarities (color map shown below the similarity matrix). The hit matrix can be leveraged to determine fMRI images with similar activation patterns. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)
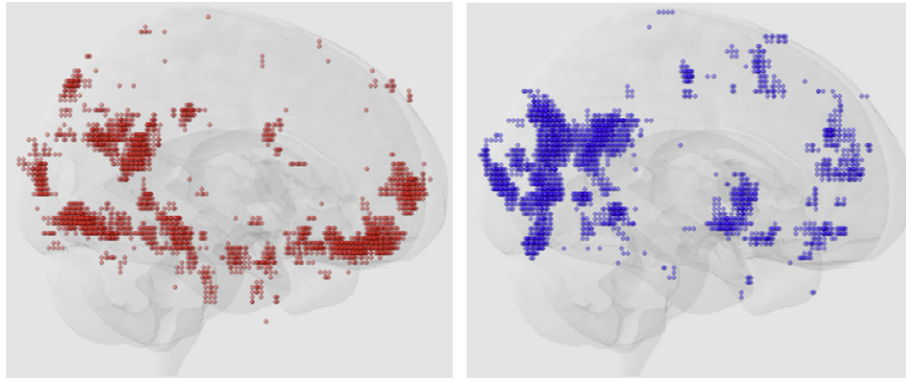
Fig. 2. A 2D view of two fMRI brain images. The raw fMRI data has been preprocessed and thresholded so that the resulting images contain only the most significant voxels. The significant voxels are shown as red (left) or blue (right) spheres. Both of these datasets are part of the Study–recall experiment. The dataset on the left represents the Study-face (SFace) condition of subject 4 (SUBJ4), and the dataset on the right represents the same condition (SFace) for subject 8 (SUBJ8). (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

images to a standard brain template. This step allows for a comparison between individuals that is robust to differences in the sizes of the subjects' brains.

In order to estimate the activation level, or significance, of each voxel in the activation map, we use the general linear model (GLM). In the standard GLM approach, each stimulating condition, such pictures of one's beloved, is presented on a controlled time schedule. The blood flow to each voxel (actually, smoothed with a 5 mm averaging function) is then compared, by regression analysis with the stimulus, which is first convolved with a standard 'hemodynamic response function' representing a typical delay of a few seconds in refresh blood. The standardized coefficient of the response is then converted to a statistical measures using a t-statistic or an F-statistic. Finally, the voxels are ranked in decreasing order of the value of this statistic. Since approximately the same number of voxels are selected from each image, the selection will not be representable as a single uniform cutoff in the value of the statistic. This unusual approach appears to compensate for variations in human brain response, at least for the populations studied here. An explanatory variable originates in a hypothesized response to a certain type of stimulus, such as the picture of one's beloved. Due to a lag in the arrival of new oxygen to active regions of the brain, each explanatory variable is generated by convolving the stimulus time series with a Hemodynamic Response Function (HRF). In our experiments, the HRF is modeled as a mixture of two gamma distributions, which is the default HRF model provided by the FSL software environment (Smith et al., 2004).

The mean and variance of the weight for each explanatory variable is calculated by linear regression. By comparing each weight with its estimated variance, the (Fisher) t-statistic can be estimated, which will be referred to as the t-value of the weight. The t-value is a measure of the degree to which the level of activation is correlated with the corresponding stimulus. The entire regression analysis is performed separately for each voxel v with coordinates (x_v, y_v, z_v), and a brain t-map is formed, showing the corresponding t-value τ_v at each voxel v. The activation or t-map is a scalar 3D dataset. Therefore, after GLM processing, the 4D raw images from an experiment result in one 3D dataset. The activation map is then thresholded to include only the voxels with the most significant t-values, as voxels with low t-values are more likely to be noise than to be causally related to the stimulus.[1] Of course, time and space costs increase as more voxels are included. Fig. 2 shows an example 2D view of two fMRI brain images after pre-

processing and thresholding to retain the most significant voxels. There are other types of processing that can be performed on fMRI datasets that do not include hypothesis testing. In this paper, we focus our efforts on GLM processed data, although our methodology may be extended to other types of processing as well.

We have collected data from various cognitive experiments for our preliminary content-based brain image retrieval tests. Each experiment involves a condition (which is the cognitive process under investigation), and a number of different subjects who underwent the experiment. The experiments in this collection include: oddball (recognition of an out-of-place image or sound), event perception (in either cartoon movie or a real film) (Zaimi et al., 2004), morality problems (in which subjects make decisions about problem situations having or lacking combinations of moral and emotional content) (Greene et al., 2001), and study–recall (study and recall, or recognition of faces, objects or locations) (Polyn et al., 2004). Some experimental conditions had many runs and many subjects, and we have selected randomly from those large sets to produce a more balanced collection. This results in a total of 430 processed experiment/subject combinations. Each was processed using FSL (a GLM-based package Worsley and Friston, 1995). The activation was then thresholded to yield the final t-maps. Table 1 gives a summary of the datasets. The data used here were collected at three different imaging sites, using different magnets. However, all the sites were studying healthy college-age subjects, which is a limitation of the present research. Extension to populations of the very young and the very old is planned.

3. Distance metrics

After processing and thresholding, the 4D fMRI datasets are converted into spatially distributed regions. In this section, we present a family of distance functions which measure the distance

---

[1] Specifically, we choose the threshold which 'activates' 1% of the voxels, accounting for the fixed scale of the brain and the difference in absolute activation levels across subjects.

Table 1
Summary of 430 t-maps used in this paper

Experiment	Number of conditions	Number of t-maps	Number of subjects	Citation
Event perception	2	53	28	Zaimi et al. (2004)
Oddball	2	8	4	N/A
Morality	3	150	22	Greene et al. (2001)
Recall	9	189	9	Polyn et al. (2004)
Romantic	2	30	15	Aron et al. (2005)

The t-maps were constructed in varying types of experiments, each with a number of experimental conditions.

between two fMRI images represented as spatially distributed regions (referred to as thresholded *t*-maps or *t*-maps for short), $\mathcal{T}_1 = \{(v_1, \tau_{v_1}), \ldots, (v_N, \tau_{v_N})\}$ and $\mathcal{T}_2 = \{(u_1, \tau_{u_1}), \ldots, (u_M, \tau_{u_M})\}$. The proposed measure defines the distance between two fMRI images as a function of the actual intensities in the voxels in or very near the region of overlap between the sets of active voxels in their respective *t*-maps $\mathcal{T}_1$ and $\mathcal{T}_2$. The distribution-based distance, also known as the Earth Mover's Distance (EMD), is designed to evaluate the similarity of multi-dimensional distributions. In our application of distribution-based distance, we will assume that a distance measure (v, u) between every pair of voxels $v \in \mathcal{T}_1$ and $u \in \mathcal{T}_2$, called the *ground distance*, is given. The distribution-based distance then 'lifts' this distance from individual voxels to a full distribution distance $\Delta(\mathcal{T}_1, \mathcal{T}_2)$ between the *t*-maps $\mathcal{T}_1$ and $\mathcal{T}_2$.

The above formulation could allow the activation in one voxel in $\mathcal{T}_1$ to move to *any* voxel in $\mathcal{T}_2$. Although such a flow may be part of a global optimum, it may violate the physical constraints of activation flow in the brain over a fixed time interval. Imposing such a physically-based constraint on flow yields a modified distance measure that restricts flow to a fixed-radius around the voxel. The formulation of the resulting fixed-radius distribution-based distance as an optimization problem is based on the well-known transportation problem, whose solution determines the minimum amount of work required to transform one distribution into the other, subject to a given upper bound on the ground distance d_{vu} between individual *t*-values τ_v and τ_u in $\mathcal{T}_1$ and $\mathcal{T}_2$.

Specifically, for *t*-maps $\mathcal{T}_1 = \{(v_1, \tau_{v_1}), \ldots, (v_N, \tau_{v_N})\}$ and $\mathcal{T}_2 = \{(u_1, \tau_{u_1}), \ldots, (u_M, \tau_{u_M})\}$, we define $D = [d_{ij}]_{N \times M}$ to be the pairwise ground distance matrix, where d_{ij} is the ground distance between voxels $v_i \in \mathcal{T}_1$ and $u_j \in \mathcal{T}_2$. Our objective is to find the optimal flow assignment matrix $F = [f_{ij}]$, with f_{ij} representing the flow of *t*-values in τ_{v_i} to τ_{u_j}, that minimizes the overall cost:

$$\Delta(\mathcal{T}_1, \mathcal{T}_2) = \sum_{i=1}^{N} \sum_{j=1}^{M} f_{ij} d_{ij}.$$

We avoid the trivial solution $F = [0]_{N \times M}$ by ensuring that maximum possible flow will leave or enter each voxel, specifically by minimizing the sum of the slacks between mass and flow for each voxel:

$$\widetilde{\Delta}(\mathcal{T}_1, \mathcal{T}_2) = \sum_{i=1}^{N} \left(\tau_{v_i} - \sum_{j=1}^{M} f_{ij} \right) + \sum_{j=1}^{M} \left(\tau_{u_j} - \sum_{i=1}^{N} f_{ij} \right)$$

subject to the following constraints:

$$f_{ij} \geqslant 0, \quad 1 \leqslant i \leqslant N, \; 1 \leqslant j \leqslant M, \tag{1}$$

$$(\rho - d_{ij}) f_{ij} \geqslant 0, \quad 1 \leqslant i \leqslant N, \; 1 \leqslant j \leqslant M, \tag{2}$$

$$\sum_{j=1}^{M} f_{ij} \leqslant \tau_{v_i}, \quad 1 \leqslant i \leqslant N, \tag{3}$$

$$\sum_{i=1}^{N} f_{ij} \leqslant \tau_{u_j}, \quad 1 \leqslant j \leqslant M. \tag{4}$$

The constraints in Eq. (1) will guarantee the non-negativity of the flow matrix. The constraints in Eq. (2) will impose the desired upper bound ρ on the ground distance the flows can move. Specifically, if the ground distance between voxel-pair (v_i, u_j) satisfies $d_{ij} > \rho$, the corresponding constraints in Eq. (2) and non-negativity of flow will force $f_{ij} = 0$. Finally, Eqs. (3) and (4) relate the flow values out of each voxel to the *t*-values. The optimal value of the objective function $\Delta(\mathcal{T}_1, \mathcal{T}_2)$ defines the distribution-based distance between two *t*-maps $\mathcal{T}_1$ and $\mathcal{T}_2$. Solving the above optimization problem for a full brain can be complex and time consuming. We have explored two simplifications: (1) removing the effect of distance terms d_{ij} (ρ-radius network flow), and (2) ignoring the activation value of voxels f_{ij} (ρ-radius bipartite method).

The ρ-radius distribution-based distance between *t*-maps $\mathcal{T}_1$ and $\mathcal{T}_2$ can be effectively computed in polynomial time using linear programming by creating a convex combination of $\Delta(\mathcal{T}_1, \mathcal{T}_2)$ and $\widetilde{\Delta}(\mathcal{T}_1, \mathcal{T}_2)$, and reporting $\Delta(\mathcal{T}_1, \mathcal{T}_2)$. There are simplified formulations for ρ-radius distribution-based distance with better complexity than linear programming. The individual flow values f_{ij} will not be scaled by the ground distance between corresponding voxels v_i and v_j, provided that $d_{ij} \leqslant \rho$. It is not hard to see that this problem can, in turn, be formulated in terms of computing the maximum flow in a flow network $G_{\mathcal{T}_1, \mathcal{T}_2} = (V, E)$. The vertex set V of this graph consists of voxels $\{v_1, \ldots, v_N, u_1, \ldots, u_M\}$, a source node s, and a sink t. For every pair of voxels $v \in \mathcal{T}_1$ and $u \in \mathcal{T}_2$, if $d(u, v) \leqslant \rho$, a directed edge $\langle v, u \rangle$ of capacity $\min(\tau_v, \tau_u)$ is in E. We also add to E directed edges $\langle s, v \rangle$ with capacity τ_v, for each $v \in \mathcal{T}_1$, and $\langle u, t \rangle$ with capacity τ_u, for each $u \in \mathcal{T}_2$. The maximum flow computed in flow network $G_{\mathcal{T}_1, \mathcal{T}_2}$ will, in turn, be used to calculate the modified distance $\Delta(\mathcal{T}_1, \mathcal{T}_2)$. It should be noted that the solution to the maximum flow problem in the network $G_{\mathcal{T}_1, \mathcal{T}_2}$ can be computed in time $O(n^{7/6} m^{2/3})$, where $n = |V|$ and $m = |E|$ (Karger and Levine, 1997). We refer to this distance measure as the ρ-radius network flow distance. Note that this measure of similarity bounds the deformation, but does not penalize larger allowed deformations.

Assigning a unit *t*-value to active voxels in thresholded *t*-maps yields a binary representation of active/inactive voxels, simplifying the computation of ρ-radius distribution-based distance. Specifically, we build a bipartite graph $G_{\mathcal{T}_1, \mathcal{T}_2}(V_1, V_2, E)$ with color classes V_1 and V_2 representing active voxels in *t*-maps $\mathcal{T}_1$ and $\mathcal{T}_2$. We include an edge (u, v) with $u \in \mathcal{T}_1$, $v \in \mathcal{T}_2$ of weight $d(u, v)$ in E if $d(u, v) \leqslant \rho$. As a result, the ρ-radius distribution-based distance between $\mathcal{T}_1$ and $\mathcal{T}_2$ can be formulated based on the weight of the maximum-cardinality minimum-weight matching in the bipartite graph $G_{\mathcal{T}_1, \mathcal{T}_2}$. We call this the ρ-radius bipartite method. It should be noted that such a bipartite matching can be computed in time $O(m\sqrt{n})$ (Cormen et al., 1990), where $n = |V|$ and $m = |E|$. Note that setting $\rho = 0$ effectively computes the *intersection* between two thresholded *t*-maps.

4. Experiments

In this section, we examine the performance of the ρ-radius network flow and bipartite distance functions for the task of content-based image retrieval. Using a database of 430 *t*-maps gathered under a number of different experimental conditions, we first explore the impact of the radius parameter, ρ, on the performance of the network flow and bipartite distance functions. We then compare the performance of these with a number of other distribution-based distance functions.

In order to test the retrieval quality obtained with each distance function, we perform a retrieval using each *t*-map in the database as a query. A retrieval is performed for a specific query by computing a ranked list of the distances between the query and the other *t*-maps in the database. For each query, the ranked list of *t*-maps is examined to determine which are *relevant* to the query. A *t*-map is relevant to the query only if it shares the same experimental conditions. As we move a pointer down the ranked list of retrieved items, we hope to see that the relevant *t*-maps predominate early in the list. To assess this we track the number of cumulated relevant and irrelevant *t*-maps as the pointer moves.

We summarize the performance of a method using the receiver operating curve (ROC). The ROC plots the probability of true positives and false positives along the normalized *Y*- and *X*-axis, respectively. We use the area under this curve as a measure of the retrieval performance, where an area of 1.0 is associated with a perfect retrieval (100% true positives and 0% false positives)

and 0.5 with random guessing (50% true positives and 50% false positives). Note that the scale of each axis is normalized so that the curve is unaffected by the ratio of relevant to irrelevant t-maps in the dataset. For each query, we measure the area under the ROC curve, and use the mean and standard deviation of the ROC curves for all the queries to evaluate a particular distance function.

We implemented the ρ-radius network flow and ρ-radius bipartite distance functions in C++ in a Linux environment. On a machine with dual 1.50 GHz Xeon processors and 1 gigabyte of RAM, both measures can computed between two brain images, each with 2000 activated voxels, in approximately 1.5 s. Our content-based retrieval system, and therefore this running time, scales linearly with the number of t-maps in the database.

4.1. Effect of the radius parameter

Both the ρ-radius network flow and ρ-radius bipartite distance functions require a specific radius parameter, ρ. The radius determines the spatial constraints on the flow between two t-maps. For example, a radius of 1 causes the flow of any voxel to be constrained to 26 possible voxels. By adjusting the radius parameter, the distance functions can be made more or less sensitive to discrepancies in the spatial distribution of activated voxels in two t-maps.

Fig. 3 shows the mean and standard deviation of the ROC curves of the ρ-radius network flow and ρ-radius bipartite distance functions for increasing neighborhood size. Both distance functions have a maximum ROC area at a radius of 2. A neighborhood size of 1 is apparently too constrained to accurately retrieve similar

brains, whereas neighborhood sizes greater than 2 are underconstrained and result in an increased number of false positives. Also observe from the figure that the bipartite distance function outperforms the network flow distance. This may be due to the fact that the bipartite distance function penalizes flow by the distance it travels, whereas the network flow distance does not. However, with a mean ROC over 0.77 for all radius values, both distance functions perform well at retrieving the t-maps in this dataset. Fig. 4a illustrates the average ROC for the radius-2 bipartite and network flow distance functions. The distance metrics perform similarly, and accurately retrieve relevant experiments from the database. Additionally in Fig. 4b illustrates the histogram of the ROC mean's obtained for the 430 queries done for the radius-2 bipartite distance function. Although the retrieval is not perfect, the curve shows that the distance functions serve as an accurate means of obtaining an initial set of similar experiments prior to examination by a human.

In Fig. 1, one can see a color-coded hit matrix where the matching function is plotted for each dataset. All of the datasets are represented along the horizontal and vertical axes. The color map is shown under the similarity matrix, with yellow representing high similarity and blue representing low similarity. The yellow line along the diagonal represents datasets matching to each other. Because the datasets are laid out along the axes in groups of experiments, we expect to see blocks of high matching along the diagonal.

Fig. 5 shows a graphical representation of the effect of the radius parameter when matching the two brain images shown in Fig. 2: (run 1 of subject 4, in the study-face condition) and (run 1
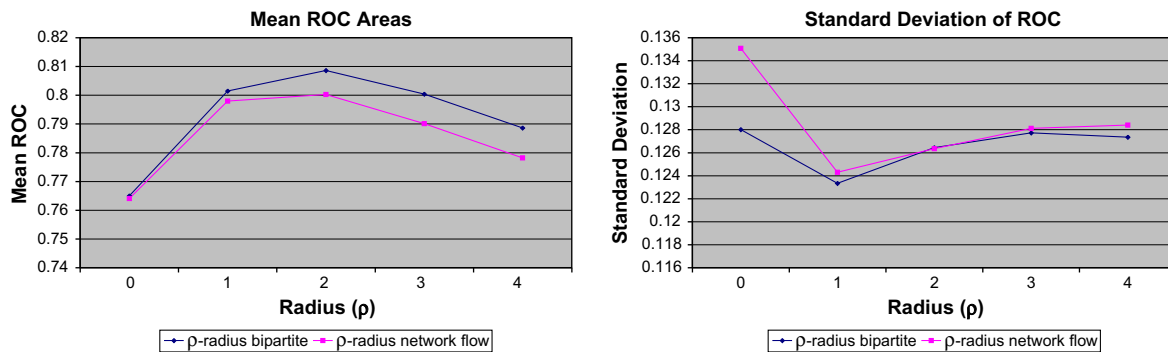
Fig. 3. Mean and standard deviation of the ROC curves produced using the ρ-radius network flow and bipartite matching distance functions for increasing radius values. Both distance measures have a maximum mean ROC with a radius size of 2.
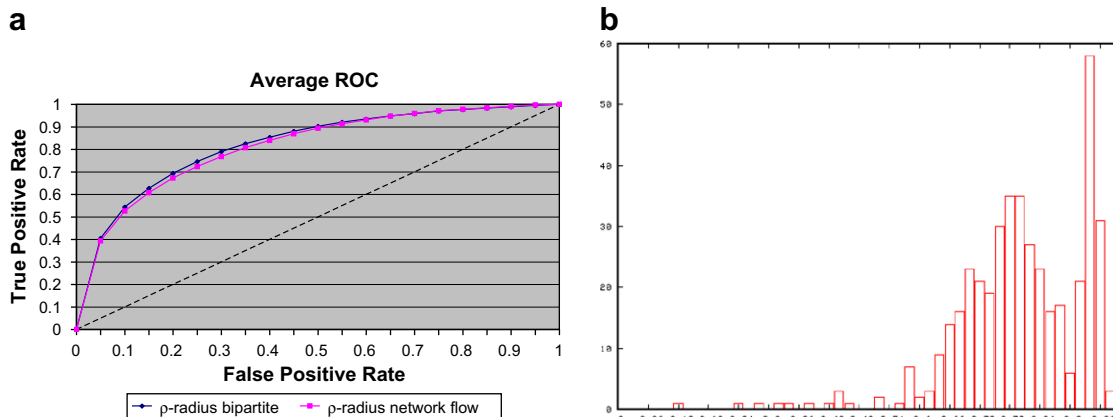
Fig. 4. (a) Average ROC curve for the radius-2 bipartite and network flow distance functions. (b) Distribution of the ROC means obtained for the 430 queries for radius-2 bipartite distance function.
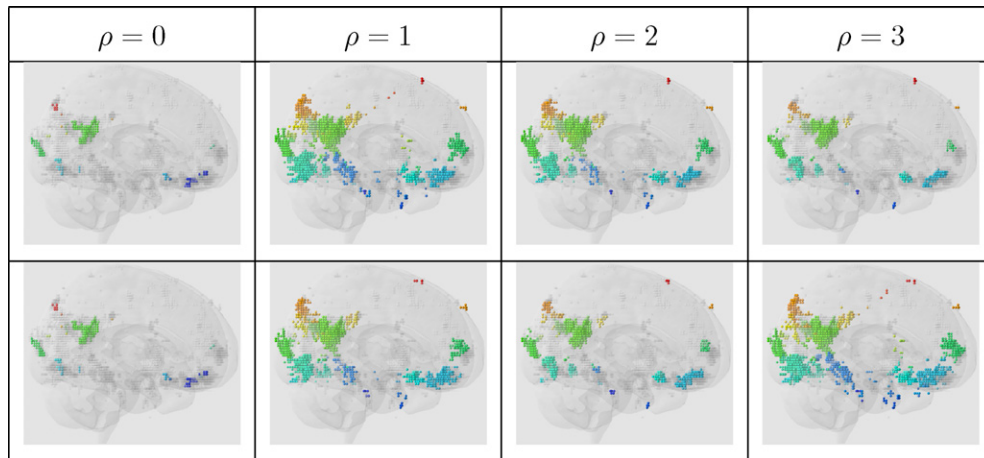
Fig. 5. The effect of increasing the radius parameter. The two rows of images show the corresponding voxels for the two brain images shown in Fig. 2, for the (ρ)-radius bipartite (first row) and network flow (second row) distance functions, respectively. The voxels in the two datasets that are found to be in correspondence to a voxel from the other dataset (matched voxels) are shown as colored spheres. Each connected cluster of matched voxels is shown in a different color. As the (ρ) parameter increases, a greater number of activated voxels are found to be in correspondence. By setting the (ρ) parameter appropriately, the distance functions are robust in the presence of noise, but still enforce spatial constraints on the corresponding voxels. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

of subject 8, study-face condition). The first row shows the effect of the radius parameter on the (ρ)-radius bipartite distance function, and the second row shows the effect in the network flow distance function, for various values of the radius parameter (ρ). The colored spheres represent voxels from each of the two datasets which are found to be in correspondence with a voxel from the other dataset. Each connected cluster of voxels found to be in correspondence is shown in a different color. For example, for $\rho = 0$ (overlap), voxels which overlap are shown in green. As the radius parameter increases, a larger number of voxels are found to be in correspondence (colored voxels). Through this parameter, the distance function is robust in the presence of noise while enforcing spatial constraints of the correspondences.

4.2. Effectiveness of the distance functions for content-based image retrieval

In this section, we compare the performance of the ρ-radius network flow and ρ-radius bipartite matching distance functions to a number of other histogram-based distance measures, including the Kullback–Leibler divergence, Brodmann cosine and simple overlap. The Kullback–Leibler divergence is a popular and effective measure of the similarity that considers the relative entropy between those two probability distributions. In the Brodmann cosine measure, the whole brain space is first partitioned into 96 left/right differentiated Brodmann regions. The *t*-map is then represented as a 96-dimensional vector, whose elements correspond to the ratios of activated voxels in that region over the total size of that Brodmann region. The similarity between two *t*-maps is computed by taking the cosine of the angle between the two corresponding vectors. In addition, we also experimented with using principle component analysis (PCA), as a means of data reduction, after which the similarity between two brains is measured as the angle vector representing the projection of the *t*-map on the top 10 significant principal components.

Table 2 shows the mean and standard deviation of the ROC curves for each distance measure. The ρ-radius bipartite and network distance functions each have a radius value of 2. Observe that the ρ-radius bipartite and network flow distance functions outperform the other techniques. By including the radius parameter ρ in the measures, we are able to more accurately retrieve similar brains, with varying amounts of noise or discrepancies in the brain

Table 2
Brain retrieval performance using a number of distance functions

Distance measure	Mean ROC	Standard deviation of ROC
ρ-Radius bipartite	0.808	0.126
ρ-Radius network flow	0.800	0.126
Simple overlap (0-radius bipartite)	0.765	0.128
Kullback Leibler divergence	0.758	0.135
PCA	0.747	0.146
Brodmann cosine	0.723	0.143

Due to the relaxation of the spatial constraints in the voxel matching, the ρ-radius bipartite and network flow distance functions outperform the other distance metrics in a content-based image retrieval task.

normalization. This is evident by comparison with the fully constrained overlap measure ($\rho = 0$). Although the Kullback–Leibler divergence is a powerful measure of the similarity between two distributions, our experiments suggest that it is too spatially constrained for the task of comparing brain fMRI images. The Brodmann cosine measure performance is significantly worse, most likely due to an overly coarse discretization of the brain into Brodmann regions. Overlap is intuitive and fast to compute; however, the method will be inaccurate if there is noise, or discrepancies in the brain normalization. Clearly, the results presented here are dependent upon the preprocessing done on the fMRI scans and on the thresholding performed on the *t*-maps.

The average ROC results obtained by our algorithm are comparable to those obtained by Ford et al. (2003), a state of the art content-based retrieval system. This is impressive, as unlike (Ford et al., 2003) who used machine learning techniques, we achieved these results simply with our improved distance function. This shows that our distance measures serve as a powerful means of discriminating fMRI brain images that can be leveraged as the underlying distance measure in a content-based retrieval of fMRI images that employs an explicit training stage.

5. Limitations

The above experiments make some important assumptions, reflecting the current limitations of our approach. Our concept of distance penalizes object translation, rotation, and scale in the image, i.e., the distance between two regions of activation (in two

images) depends on their relative position, orientation, and size in a brain-aligned coordinate system. While appropriate for the task of computing the similarity of two brain activation patterns, this lack of invariance makes the framework less appropriate, for example, for the task of detecting lesions in an image, if what constitutes a lesion is independent of position, orientation, or scale. This limitation can be overcome if the EMD measure is computed as an inner loop, with an outer loop which seeks to superimpose, align and scale the images to each other. A second limitation concerns the uniformity with which activation is compared. As domain specialists become able to provide the needed information, our metric could be easily augmented to include measures of 'relative significance' of the various parts of an image, to provide a better task-specific metric. Finally, our database is rather small, and methods for indexing effectively into larger databases, in order to limit the number of candidates that need to be verified using our distance measure, need to be developed. We are currently exploring abstract image representations that will support fast indexing.

6. Conclusions

In this paper, we have presented a family of distance functions that may be leveraged to develop a content-based image retrieval system. Using two different assumptions about the underlying brain images, we have developed two powerful image similarity measures, the ρ-radius network flow and bipartite distance functions, for general and binary images, respectively. Increasing the value of ρ effectively increases the two distance measures' invariance to increasing amounts of noise and within-class deformation. Through an experimental evaluation in the domain of content-based brain image retrieval, we have shown that by using a graphical network flow formulation to match images in a way that conserves the amount of activation, and by accounting for the inherent noise in the data, our distance functions outperform other commonly used measures. The two methods, in fact, add to the realism of the approach in two complementary ways. One penalizes long moves, while the other conserves a measure of the correlation. This suggests that if we can find an efficient way to conserve correlation and also penalize distance, we will achieve somewhat better performance, while increasing the realism of the model.

Acknowledgements

References

Aron, A., Fisher, H., Mashek, D., Strong, G., Li, H., Brown, L., 2005. Reward,motivation, and emotion systems associated with early-stage intense romantic love. Neurophysiol 94, 327–337.

BRAin Image Database. 2005. <http://braid.uphs.upenn.edu/websbia/braid/>.

Cormen, Thomas H., Leiserson, E., Charles, Rivest, Ronald L., 1990. Introduction to Algorithms. MIT Press. COR th 01:1 1.Ex.

Ford, J., Makedon, F., Megalooikonomou, V., Saykin, A., Shen, L., Steinberg, T., 2001. Spatial comparison of fMRI activation maps for data mining: A methodology of hierarchical characterization and classification. In: Presented at the 7th Annual Meeting of the Organization for Human Brain Mapping (OHBM01), Brighton, UK.

Ford, J., Shen, L., Makedon, F., Flashman, L.A., Saykin, A.J., 2002. A combined structural–functional classification of schizophrenia using hippocampal volume plus fMRI activation. In: Second Joint Meeting of the IEEE Engineering in Medicine and Biology Society and the Biomedical Engineering Society.

Ford, J., Farid, H., Flashman, L., Saykin, A.J., 2003. Patient classification of fMRI activation maps. In: Medical Image Computing and Computer-Assisted Intervention, Montreal, Quebec, Canada.

Greene, J., Sommerville, R., Nystrom, L., Darlye, J., Cohen, J., 2001. An fMRI investigation of emotional engagement in moral judgment. Science 293 (5537), 2105–2108.

Karger, D.R., Levine, M., 1997. Finding maximum flows in undirected graphs seems easier than bipartite matching. In: Proc. 30th Annual ACM Symposium on Theory of Computing.

Kontos, D., Megalooikonomou, V., Makedon, F., 2004. Computationally intelligent methods for mining 3D medical images. In: 3rd Hellenic Conf. on Artificial Intelligence, Samos Island, Greece. pp. 72–81.

Lazarevic, A., Pokrajac, D., Megalooikonomou, V., Obradovic, Z. 2001. Distinguishing among 3-D distributions for brain image data classification. In: 4th Internat. Conf. on Neural Networks and Expert Systems in Medicine and Healthcare, Milos Island, Greece. pp. 389–396.

Megalooikonomou, V., Davatzikos, C., Herskovits, E.H., 1999. Mining lesion-deficit associations in a brain image database. In: ACM SIGKDD Internat. Conf. on Knowledge Discovery and Data Mining, San Diego, CA. pp. 347–351.

Megalooikonomou, V., Ford, J., Shen, L., Makedon, F., 2000. Data mining in brain imaging. Statist. Methods Med. Res. 9, 359–394.

Megalooikonomou, V., Kontos, D., Pokrajac, D., Lazarevic, A., Obradovic O. Boyko, Z., Saykin, A., Ford, J., Makedon, F., 2003. Classification and mining of brain image data using adaptive recursive partitioning methods: Application to alzheimer disease and brain activation patterns. In: Human Brain Mapping Conf., OHBM'03.

Niblack, Wayne, Barber, Ron, Equitz, William, Flickner, Myron, Glasman, Eduardo H., Petkovic, Dragutin, Yanker, Peter, Faloutsos, Christos, Taubin, Gabriel, 1993. The qbic project: Querying images by content, using color, texture, and shape. In: Storage and Retrieval for Image and Video Databases (SPIE). pp. 173–187.

Osada, Robert, Funkhouser, Thomas, Chazelle, Bernard, Dobkin, David, 2002. Shape distributions. ACM Trans. Graph 2, 807–832.

Pentland, Alex, Picard, Rosalind W., Sclaroff, Stan, 1994. Photobook: Tools for content-based manipulation of image databases. In: Storage and Retrieval for Image and Video Databases (SPIE). pp. 34–47.

Polyn, S., Cohen, J., Norman, K., 2004. Detecting distributed patterns in an fMRI study of free recall. In: Society for Neuroscience Conf., San Diego, CA.

Saykin, A.J., Wishart, H.A., Flashman, L.A., McAllister, T.W., McHugh, T., Ford, J.C., Shen, L., Steinberg, T., Makedon, F., 2002. Structure/function relationships in brain disorders: Strategies for mining volume, shape, lesion and bold fMRI activation data. In: Society for Biological Psychiatry meeting, Philadelphia, PA.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 (S1), 208–219.

Van Horn, J., Gazzaniga, M., 2002. Databasing fMRI studies: Towards a discovery science of brain function. Nature 3, 314–318.

Veltkamp, Remco C., Tanase, Mirela, 2000. Content-based image retrieval systems: A survey. Technical Report UU-CS-2000-34, Universiteit Utrecht.

Wang, Q., Kontos, D., Li, G., Megalooikonomou, V., 2004. Application of time series techniques to data mining and analysis of spatial patterns in 3D images. In: IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Montreal, Quebec, Canada.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited – again. NeuroImage 2, 173–181.

Zaimi, A., Hanson, C., Hanson, S., 2004. Event perception of schema-rich and schema-poor video sequences during fMRI scanning: Top down versus bottom up processing. In: Proc. Annual Meeting of the Cognitive Neuroscience Society.