A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe ☆

Michael A. Yassa, Craig E.L. Stark *

Center for Neurobiology of Learning and Memory, Department of Neurobiology and Behavior, University of California, Irvine, California, USA

A B S T R A C T

Accurate cross-participant alignment within the medial temporal lobe (MTL) region is critical for fMRI studies of memory. However, traditional alignment approaches have been exceptionally poor at registering structures in this area due to significant inter-individual anatomic variability. In this study, we evaluated the performance of twelve registration approaches. Specifically, we extended several traditional approaches such as SPM's normalization and AFNI's 3dWarpDrive to improve the quality of alignment in the MTL region by using weighting masks or applying the transformations directly to ROI segmentations. In addition, we evaluated the performance of three fully deformable methods, DARTEL, Diffeomorphic Demons, and LDDMM that are effectively unconstrained by number of degrees of freedom. For each, we first assessed the method's ability to achieve optimal overlap between segmentations of subregions of the MTL across participants. Then we evaluated the smoothness of group average structural images aligned using each method to assess the blur that results when voxels of different tissue types are averaged together. In general, we found that when anatomical segmentation is possible, substantial improvement in registration accuracy can be gained in the MTL even with a small number of deformations. When segmentation is not possible, the fully deformable models provide some improvement over more traditional approaches and in a few cases even approach the performance of the ROI-based approaches. The best performance is achieved when both methods are combined. We note that these conclusions are not limited to the MTL and are easily extendable to other areas of the brain.

Published by Elsevier Inc.

Introduction

There is a consistent and growing interest in functional imaging of the medial temporal lobe (MTL), given the abundance of theoretical approaches to understanding this region's role in long-term memory. Accurate cross-participant alignment within the MTL is critical for these fMRI studies. However, cross-participant alignment has traditionally been a difficult challenge. Structures in the MTL exhibit significant variability across individuals (Insausti, et al. 1998). The resulting blur across scans severely limits the signal-to-noise ratio in group analyses, and limits the confidence one has in the precise localization of fMRI signals (Stark and Okado, 2003; Miller et al., 2005; Kirwan et al., 2007). Most studies have made use of an atlas-based approach to register whole brain scans across participants. These traditional approaches look for a global optimum alignment that can be achieved under the limitations imposed by the available degrees of freedom. Unfortunately, achieving a global optimum comes at the cost of regional accuracy. For example, we have demonstrated in the past that a typical whole brain alignment to the Talairach atlas (Talairach and Tournoux, 1988) or to the MNI template results in only a ~50% chance that a voxel from the hippocampus of one participant lies within the hippocampus of another participant after alignment. As a consequence of this poor alignment, functional MRI effects in the MTL can be severely underestimated or in some cases completely missed (Miller et al., 2005).

The purpose of this work is to evaluate the performance of several families of cross-participant registration methods in aligning structures in the MTL. Our goal is to increase sensitivity of fMRI studies to activity in regions with small structures such as the MTL and to increase confidence in anatomical localization of activity by ensuring adequate overlap within regions and across scans. In our previous work, we have proposed a basic registration approach that focuses on regional alignment. The approach, dubbed ROI-AL for Region of Interest based ALignment, operates on two main principles: (1) In the case of limited degrees of freedom, if hypotheses to be tested with fMRI are directed towards specific structures or regions, the available power should be applied locally and not wasted outside of these regions, and (2) When possible, alignment should be based not on

grayscale contrast but on regional anatomical information, where knowledge of the locations, sizes and shapes of structures can be used (Stark and Okado, 2003; Miller et al., 2005). While the MTL is used here, it serves only as a prototypical test case. There is nothing about the MTL per se that would lead us to question the generalizability of these principles to other structures. For example, this approach has been successfully used in aligning several structures in the basal ganglia as well (Nomura et al., 2007; Mattfeld and Stark, submitted for publication).

Conventional registration approaches such as Talairach alignment use a limited number of degrees of freedom to find the best global fit of the source image to the target template. Since the focus is on global and not local optimization, this limitation poses substantial challenges to the methods' ability to model the complex variation across different subjects' medial temporal lobes. In addition, there is often very little in the grayscale images to delineate among adjacent structures. One way to address these issues simultaneously is to draw segmentations of regions of interest on structural images and to align these segmentations directly. We have shown that this approach significantly improves the quality of registration and increases our fMRI detection power in the MTL using alignment algorithms that ranged from a standard 12-parameter affine transform (Stark and Okado, 2003) to separate affine transformations for each ROI (Law et al., 2005) to alignment using large deformation diffeomorphic metric mapping (Miller et al. 2005, Kirwan et al. 2007).

Currently, there are many options for fully deformable nonlinear registration algorithms. Unlike traditional approaches, these methods are not limited by a small number of degrees of freedom and thus are better at estimating local deformations. We considered three such methods here and compared them to several traditional approaches. One such method, DARTEL (Ashburner, 2007), was designed to work on whole brain scans, although it can potentially be modified to work on ROI segmentations as well. Two other methods, Demons (Vercauteren, et al. 2007; Thirion, 1998) and Large Deformation Diffeomorphic Metric Mapping (LDDMM: Beg et al., 2005; Miller et al., 2005), can and have been applied both to whole brain scans and to three-dimensional segmentations of the MTL. This last approach, referred to as Region of Interest LDDMM (ROI-LDDMM), has been our most effective implementation of the ROI-AL approach to date.

We have previously shown that ROI-LDDMM yields error scores much lower than alignment using AFNI's Talairach transformation, SPM's normalization and FSL's registration (Kirwan et al., 2007). Here, we extended this work, by including recently available techniques such as DARTEL and Diffeomorphic Demons, as well as attempts to improve the more traditional techniques. We hypothesized that the large deformation methods will perform much better than the traditional small deformation approaches, which use a limited number of degrees of freedom to estimate the deformations necessary to warp images into the same space. Additionally, we hypothesized that ROI-based approaches (that require the segmentation or anatomical definition of the regions of interest) would outperform the MRI-based approaches (that are applied directly to the grayscale MR images) because they benefit from *a priori* information on individual anatomical variability.

We used two metrics to assess the performance of each approach. First, we used manual segmentations of subregions of the MTL (hippocampus, parahippocampal cortex, perirhinal cortex and entorhinal cortex) to evaluate the degree to which each algorithm could align participants' scans such that the transformed segmentations would overlap across participants. Then, we estimated the smoothness of the average structural scan produced by each method. Smoothness is a function of the blur that results when voxels of different tissue type are averaged together. Thus, higher accuracy results in a decrease in tissue type mismatch, which in turn results in a reduction in the smoothness of the average image.

Materials and methods

MRI data acquisition

Structural MRI data was collected as part of a functional MRI study that was recently published (Bakker et al., 2008). Twenty right handed healthy individuals (10 female) with a mean age of 25 participated in this study. Imaging was conducted on a 3 Tesla Philips scanner (Eindhoven, The Netherlands) equipped with an 8-channel SENSE (sensitivity encoding) coil at the F.M. Kirby Research Center for Functional Brain Imaging at the Kennedy Krieger Institute (Baltimore, MD). A high-resolution T1-weighted 3D magnetization-prepared rapid gradient echo (MPRAGE) scan was acquired for each participant with isotropic 1 mm voxels, a field of view of 256 and an acquisition matrix of 256×256. The scans were aligned parallel to the long axis of the hippocampus covering the entire brain volume in 150 slices.

Registration methods

Talairach alignment

This basic alignment served as the starting point for all other methods, and thus is used as a baseline in our study. The Talairach normalization procedure was conducted in AFNI (Cox, 1996) and aligns MRI scans to the coordinate system defined by Talairach and Tournoux (1988) using a two-step procedure. The initial alignment, often referred to as AC–PC alignment, is a rigid transformation (3D rotation and translation) using two landmarks for the AC (superior edge and posterior edge), one landmark for the PC (inferior edge), and two midsagittal landmarks. The second step starts with a definition of the bounding box for the brain, followed by a 12-parameter piecewise linear transform that registers the AC, PC, anterior, posterior, left, right, superior and inferior boundaries to atlas locations. These landmarks define a set of 12 boxes. The image in each box is independently scaled in three dimensions to fit the Talairach atlas coordinates. The anatomical segmentations were drawn on the Talairach transformed structural images.

Normalization in SPM

We used SPM5 (Wellcome Department of Imaging Neuroscience, University College, London, UK) running under MATLAB 7.5/R2007b (The Mathworks, Sherborn, MA) for all of our SPM-based methods. Normalization in SPM5 generally works by minimizing the sum of squared difference (SSD) in intensity between the source and the target images. For the algorithm to work well, two criteria should be considered: (1) contrast should be similar across the source and target images, and (2) the source image should be roughly aligned to the same space as the target image (e.g. AC–PC). Both criteria aim to improve the starting estimates to increase the likelihood that normalization will converge towards the global optimum.

Estimating normalization parameters in SPM5 is a two-step procedure. The first step determines the optimum 12-parameter affine transformation by matching the whole head first to the template, then proceeding to only include voxels from brain tissue using a canonical brain mask. The procedure uses a Bayesian framework, which maximizes the posterior probability of it being correct (Ashburner et al., 1997). The second step estimates nonlinear deformations that are defined by the linear combination of discrete cosine transform (DCT) functions where each of the deformation fields is described by 1176 parameters (Ashburner and Friston, 1999). The algorithm simultaneously attempts to minimize the membrane energies of the deformation fields and the residual squared difference between the source image and the target. Additional options allow for affine regularization, which penalizes excessive stretching or shrinking, as well as source and target image smoothing, which improves the likelihood of finding a global optimum.

We performed two iterations using this method, one with (SPM − MTL weighted) and one without using a regional weighting mask (SPM − standard). In the approach without weighting, we used default options for estimating normalization parameters (8 mm smoothing for source images, no smoothing for the already smooth target, affine regularization to the ICBM template, nonlinear DCT frequency cutoff set to 25, and 16 nonlinear iterations). In the second approach, we used a binary mask designed to focus the alignment on the MTL. The mask began as our laboratory's model of the MTL in MNI space which was then expanded in all directions using a 20 mm Gaussian blur and then clipped into a binary mask at a grayscale value of 0.5. The final mask was created with dimensions 128×88×72 (RL −63 64, AP −37 50, IS −51 20) and fully encompassed MTL structures. Once again, default options were used for estimating normalization parameters. For both approaches, deformed images were created at 1 mm resolution using trilinear interpolation. In an additional attempt to not waste degrees of freedom, we also ran an iteration with cropped images (using the aforementioned mask), however since SPM cannot deal with zero values or NaN's all that well, the resulting alignment was very poor and was therefore not included here or in subsequent analyses.

Unified segmentation in SPM

Unified segmentation in SPM5 is a single solution that combines segmentation, bias correction and spatial normalization under the same iterative model (Ashburner and Friston, 2005). Estimating the model's parameters is done by alternating between tissue classification, bias correction and registration steps. Registration uses canonical priors from three tissue classes, gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) in addition to a fourth class representing the background of the image. These tissue probability maps are allowed to deform within a set of parameters estimated by a linear combination of a thousand cosine transform bases. Since each voxel could contain more than one tissue type, several Gaussian distributions are used for each tissue class (default is two for GM, two for WM, two for CSF and four for background). Like the previous technique, we performed two iterations, one with (SPM unified seg − MTL weighted) and one without a regional weighting mask (SPM unified seg). We used default options for the entire procedure, except for changing the sampling distance to 1 mm since our scans' resolution was 1 mm. For masking, we used the same mask used for the traditional normalization approach.

AFNI's 3dWarpDrive

We used Analysis for Functional Neuroimages (AFNI v. 2007_03_06_084: Cox, 1996) for several of our registration methods. 3dWarpDrive is a component of AFNI that tries to find a spatial transformation that optimally warps the source to match the target dataset using a defined number of parameters. Currently, it can perform a 12-parameter affine transformation on any dataset (whole brain scans or ROI segmentations). It also has the capacity to perform a 39-parameter fit, although this portion of the program is described as experimental. We started with applying both fits to the ROI maps directly (ROI-AFNI 3dWD12 and ROI-AFNI 3dWD39). The affine version was applied to each of the ROIs individually and then the transformations were combined so that a total of ten affine fits were used, identical to the ROI-AL approach used by Law et al. (2005). In addition, we also applied the techniques directly to cropped versions of the grayscale MRI scans (AFNI 3dWD12 − MTL weighted, AFNI 3dWD39 − MTL weighted). We used cropped versions here instead of the whole brains in order to avoid wasting degrees of freedom outside the MTL. The cropped versions were produced by multiplying the whole brain scans by the mask created in the first method. Although we used the 39-parameter fit with the grayscale images, a limitation of the current version of the software was the inability to apply the resulting transformation to a secondary dataset. Thus, we could only assess the overlap scores from the ROI-based 39-parameter fit and not the smoothness, since we could not apply the deformations to the grayscale images. Likewise, we could only assess the smoothness and not the overlap scores in the grayscale-based 39-parameter fit since the deformations could not be applied to the ROI segmentations.

LDDMM

Large Deformation Diffeomorphic Metric Mapping or LDDMM (Beg et al., 2005; Miller et al., 2005) is a registration algorithm that attempts to assign metric distances on the space of the images to quantify changes in shape and is what we have used extensively in the past for alignment in the MTL region. It computes invertible transformations from displacement vector fields while simultaneously preserving anatomical features such as curve smoothness and surface topology. Diffeomorphic transformations have several desirable properties. They are: (1) one-to-one; disjoint shapes remain disjoint and there is no fusion of points, (2) continuous; connected shapes remain connected, and (3) smooth; smoothness of object boundaries is preserved. The LDDMM approach is effectively unconstrained by degrees of freedom. Deformations are parameterized using time-varying velocity vector fields, on which a smoothness constraint is imposed for regularization. This constraint can be described as the width of the Gaussian shaped filter that is applied to the velocity vector field, which controls the smoothness of the field and the smoothness of the final deformation. This approach has been previously used to map the shape of the hippocampus and how it changes in the course of aging and Alzheimer's disease (Csernansky, et al., 2005, Wang et al. 2003), as well as to increase the statistical power of functional maps in the MTL (Miller et al., 2005; Kirwan et al., 2007). We applied this method directly to the ROI segmentations (ROI-LDDMM) by aligning them to a central tendency model of ROI segmentations derived for a different set of 20 young healthy individuals. Since this method outperformed the use of LDDMM on the grayscale structural images (MRI-LDDMM) in our previous work (Miller et al., 2005), we focused here on the ROI-based implementation.

DARTEL

DARTEL stands for "Diffeomorphic Anatomical Registration Through Exponential Lie Algebra" and was proposed by Ashburner (2007) as a potential alternative to SPM's traditional registration approaches, running as a toolbox in SPM5. Like LDDMM, it utilizes the large deformation framework to preserve topology. As a result, the deformations are also diffeomorphic and invertible. Unlike LDDMM's time-varying velocities approach, DARTEL's deformations are parameterized by a single velocity field that is constant in time. Registration involves simultaneously minimizing the sum of squares difference between the source and target images as well as the linear elastic energy of the deformations used to warp the target image (it can also be adjusted to penalize according to membrane or bending energy or some combination). DARTEL uses a Gauss–Newton optimization scheme to minimize the sum of squares. Template creation is incorporated in the algorithm and is iterative so that a new template based on the entire sample is re-created at the end of each iteration. Since the procedure was designed to operate on whole brain MRI scans, we did not apply it to regional segmentations, although it may be possible to adjust DARTEL's parameters to warp ROIs as well by adjusting the smoothing as well as the form and amount of regularization applied (John Ashburner, personal communication, July 4, 2008).

Demons

The original demons algorithm was based on an analogy with thermodynamic concepts (Thirion, 1998). In 1871, James Maxwell proposed a thought experiment that posed a potential violation of the second law of thermodynamics. He conceived of a closed system in which two compartments existed with a trapdoor in between. The

system is filled with gas molecules at the same temperature. A guard "demon" opens the trapdoor and only allows in fast traveling molecules into one compartment, and slow traveling molecules into the other, eventually resulting in one compartment having a higher temperature than the other, and consequently a decrease in entropy. This appears to violate the second law of thermodynamics, which states that entropy in a closed system must always increase. However, the paradox was eventually resolved some sixty years later when it became clear that for the demon to calculate the speed of molecules it must generate a greater amount of entropy; thus, the total entropy in the system would still increase. The application to image matching is quite simple. One can treat object boundaries in one image (the fixed

image) as semi-permeable membranes and let the other image (the moving image), considered a deformable grid, diffuse through these membranes using effectors within the interfaces. These effectors, much like Maxwell's demons, would act locally to push the grid inside the object boundary if it is labeled "inside" and outside the object boundary if it is labeled "outside" (Thirion, 1998).

When used as an optimization procedure on vector field displacements with simple images like ROI segmentations, the results of demons alignment are typically diffeomorphic (invertible), although the diffeomorphic properties are not strictly enforced. Recently, a fully diffeomorphic version of demons has been described and implemented (Vercauteren al. 2007) and is included along with
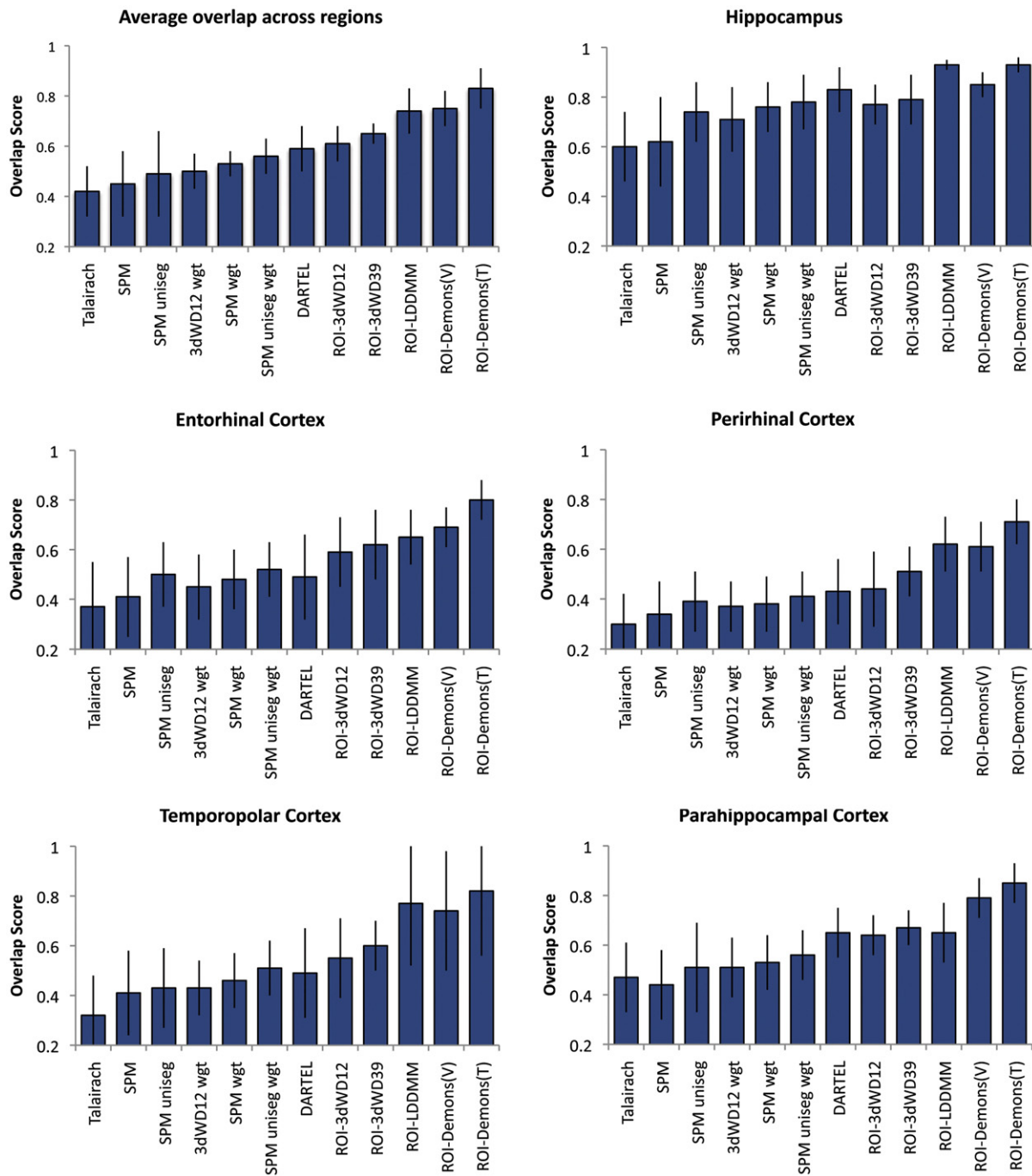

Fig. 1. Comparison of the overlap scores of the ROI segmentations warped by each method. Perfect overlap is 1. Overall, ROI-based approaches perform better than grayscale approaches, with the diffeomorphic methods performing better than the traditional constrained approaches. Within the grayscale methods, diffeomorphic DARTEL performs better than others. In general, Talairach and traditional SPM approaches do not perform as well. ROI-Demons(T) refers to ROI-Demons using Thirion's update rule, while ROI-Demons(V) refers to ROI-Demons using Vercauteren's diffeomorphic update rule. Error bars are ±1 standard deviation.

an implementation of Thirion's original demons algorithm in the freeware package MedINRIA (version 1.7.0, ASCLEPIOS Research Team, France). The diffeomorphic version enforces smoother deformations by operating on a diffeomorphic space of displacement fields. Similar to DARTEL, diffeomorphisms here are parameterized by using constant velocity vector fields, which leads to faster computations of vector field exponentials, although the optimization strategy and the objective function to be minimized are different. We used ROI-Demons(T), which is based on the original Thirion's demons, as well as the newer ROI-Demons(V), which uses Vercauteren's diffeomorphic transformations to align segmentations of our regions of interest. In both cases, the resulting fields were diffeomorphic (nonzero Jacobian determinant), but Thirion's demons gave us better results. Thus, we chose to apply the other metrics to the method using Thirion's rule only. The procedure is composed of two steps: first, correspondence energy is minimized, and then regularization is applied to ensure smoother spatial transformations. The alignment starts with a number of iterations using a coarse subsampling of the image followed by several levels of fine-resolution alignment. We used three levels of alignment with 15 iterations at the coarse level, 10 at the intermediate level, and 5 at the fine level (default options). We applied a smoothness constraint on the vector fields of 1 standard deviation. We did not use histogram matching, which is only important for matching grayscale images with different histograms.

Performance evaluation

Overlap scores

First, the hippocampus, perirhinal cortex, entorhinal cortex, and parahippocampal cortex were segmented in AFNI bilaterally using the methods used in the past in our laboratory (e.g. Yassa and Stark, 2008, Kirwan et al., 2007). The hippocampus was first segmented by starting in the most lateral slice where it was visible in the sagittal plane and proceeding in the medial direction until it completely disappeared. The superior boundary was set by the amygdala rostrally and the choroid fissure and the lateral ventricle caudally. The white matter of the parahippocampal gyrus served as the inferior boundary (Duvernoy, 1998). Temporopolar, entorhinal, perirhinal, and parahippocampal cortices were defined bilaterally in the coronal plane according to the methods described by Insausti et al., (1998). The temporopolar cortex was first outlined as it appeared in anterior coronal slices using an inclusive outline of the gray matter until white matter was visible and the gyrus of Schwalbe and indentation of the inferotemporal sulcus appeared. Then, the outline was reduced to the gray matter that extended from the peak of the gyrus of Schwalbe (the most lateral aspect) and following the gyrus medially until the indentation of the inferotemporal sulcus. The perirhinal and entorhinal cortices were subsequently defined as the portions of the parahippocampal gyrus extending from the most medial aspect of the temporal lobe to include the full length of the collateral sulcus. The boundary separating those two regions depended on the depth of the collateral sulcus, which is subject to inter-individual anatomical variation and is reviewed in detail in Insauti et al. (1998). The parahippocampal cortex was defined bilaterally as the portion of the parahippocampal gyrus caudal to the perirhinal cortex and rostral to the splenium of the corpus callosum. Following hand segmentation, the ROIs were smoothed with a 3×3×3 median filter. The smooth ROIs were used to evaluate each alignment technique. Each method's resulting deformations were applied to the segmented map. Overlap scores were calculated in MATLAB as the ratio of voxels within a region that have the same ROI label in both images versus the total number of labeled voxels in the target image (Eq. (1)).

$$OS_{i,j} = \frac{vol(I_i \cap I_j)}{vol(I_i)}, \quad i = 1 \dots N, j = 1 \dots N. \tag{1}$$

I_i and I_j are two subject images containing a single ROI label. This metric is calculated for each possible pair of subjects using one participant as the source image and another as the target (190 pairings). Scores were averaged over all ROIs and participant pairings. Perfect overlap would be a score of 1, while no overlap at all would be a score of 0. This approach estimates the resulting cross-participant overlap (which is often the most important outcome in registration for fMRI purposes), instead of how well each image is aligned to the target model. This also avoids potential bias towards the ROI-based methods that may result from using the MTL model as the basis for the overlap metric.

FWHM smoothness

For each method, the warped structural scans were averaged together to create a single mean image using AFNI. Misalignment due to tissue type mismatch resulting from registration errors leads to a blurry mean image. This approach has been utilized before to assess the performance of registration algorithms (Wu et al., 2006). We used a quantitative measure of spatial correlation of voxels to assess the smoothness of the mean image. AFNI's program 3dFWHM (Cox, 1996; Xiong et al., 1995) was used to estimate the smoothness of the image. The program starts by assuming that voxels were spatially independent to begin with and then proceeds to estimate the kernel width that would be needed to produce the observed smoothness of an image using the method described by Forman et al. (1995). 3dFWHM estimates smoothness along any given dimension using the following equation (Eq. (2)):

$$FWHM_x = dx \sqrt{\frac{-2\ln 2}{\ln\left(1 - \frac{var(ds)}{2var(s)}\right)}} \tag{2}$$

where dx is the voxel size along the x dimension, $var(ds)$ is the variance of the difference in between neighboring voxels in the x dimension and $var(s)$ is the overall variance at each voxel. In group-averaged structural images, a larger FWHM would be expected with increased mismatch and more alignment errors. Since brain MRI images have voxels that are inherently spatially correlated (MRI images are not composed of random gray, white and black dots), it is important to assess smoothness due to alignment mismatch without the contributions of inherent smoothness in an MRI image. To do this, we simply used a single subject multi-scan average template (colin27) to estimate the inherent smoothness, and compared the mean images' smoothness to that value for each technique.

Hippocampal subfield alignment

To further assess the performance of the diffeomorphic methods on a more challenging task, we tested how well they did at aligning subfields of the hippocampus in a subset of scans ($n = 12$), in which

Table 1
3dFWHM smoothness estimate for each method

Method	FWHMx	FWHMy	FWHMz	Mean (xyz)	% Change
Colin27 (template)	4.33	5.02	3.84	4.39	0.00
Talairach	5.78	6.22	5.17	5.72	30.25
SPM traditional	5.90	7.34	6.20	6.48	47.48
SPM traditional (MTL weighted)	6.83	6.534	5.66	6.34	44.33
SPM unified segmentation	7.31	6.97	5.77	6.68	52.10
SPM unified segmentation (MTL weighted)	7.40	6.92	5.78	6.70	52.50
3dWarpDrive 12 (MTL weighted)	6.40	7.18	5.43	6.34	44.33
3dWarpDrive 39 (MTL weighted)	6.74	7.24	5.94	6.64	51.15
SPM DARTEL	5.70	5.86	4.94	5.50	25.24
ROI-3dWarpDrive 12	5.40	5.62	4.25	5.09	15.90
ROI-Demons (Thirion)	4.97	5.68	4.13	4.93	12.21
ROI-LDDMM	5.40	5.18	4.09	4.89	11.33

subfields had been manually segmented into dentate gyrus (DG)/CA3 subfield, CA1 subfield, and subiculum (Csernansky et al., 2005; Wang et al., 2003; Kirwan et al., 2007) according to the atlas of Duvernoy (1998). The subfields in Duvernoy (1998) are defined on eight coronal slices along the anterior–posterior axis of the hippocampus. Representative slices in each hippocampus that best (closest) resembled the slices described were chosen and segmented according to the atlas description. The segmentation then proceeded from these slices in both directions slice by slice to ensure a smooth transition across slices.

Using Talairach alignment as a reference, we tested DARTEL's performance by applying the grayscale warps to the subfield ROIs. We then tested the performance of ROI-Demons and ROI-LDDMM two ways: (1) by applying the warps from the MTL ROIs directly to the subfields, and (2) by calculating warps based on the alignment of the

subfield ROIs themselves, to see how much can be gained by incorporating this information. We used a central tendency model that was created by taking the mode of all participants' hippocampal subfields and other MTL structures' ROI labels that were transformed using ROI-LDDMM alignment. This model served as the target template for alignment. We then assessed the performance of Talairach, DARTEL, ROI-Demons(T), Subfield ROI-Demons(T), ROI-LDDMM and Subfield ROI-LDDMM using the aforementioned overlap score metric.

Results

Comparison of overlap scores between different methods and regions of interest (Fig. 1) showed that overall, the approaches fully
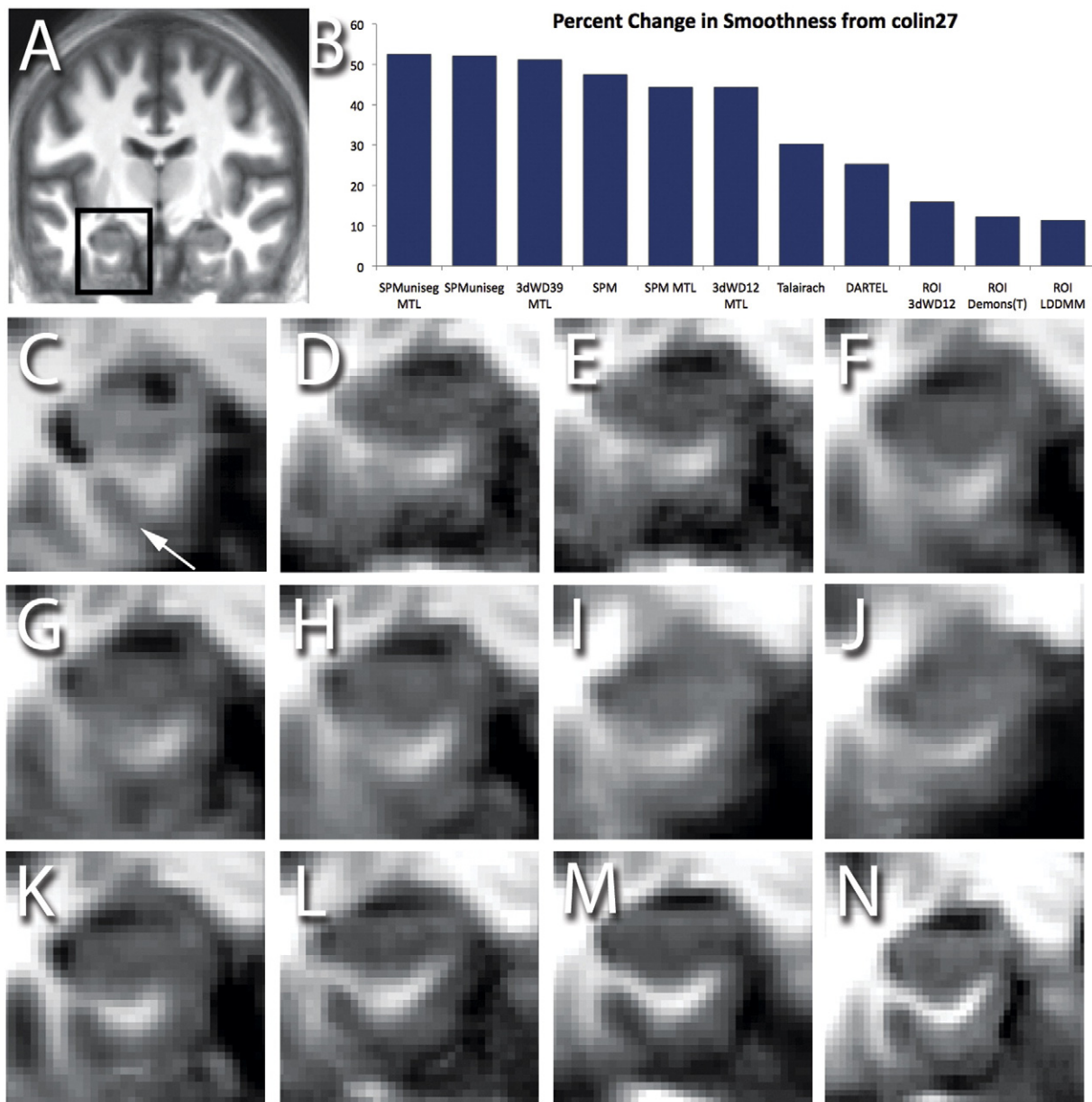

Fig. 2. Results of 3dFWHM smoothness assessment. (A). A typical average MRI scan illustrating the region in the left medial temporal lobe that is depicted in the rest of the figures; (B) A graph showing percent change in smoothness from the colin27 single subject template. Methods with "MTL" used an MTL weighted mask. We can see that the ROI-based methods produced the least increase in smoothness (the least blur). In addition, within the grayscale methods, the fully deformable method DARTEL had the least increase in smoothness; (C) Single subject template (colin27; arrow points to the location of the clearly visible collateral sulcus; (D) Talairach alignment; (E) Traditional SPM alignment; (F) Traditional SPM with MTL weighting mask; (G) SPM unified segmentation; (H) SPM unified segmentation with MTL weighting mask; (I) AFNI 3dWarpDrive with 12 parameters and an MTL weighting mask; (J) AFNI 3dWarpDrive with 39 parameters and an MTL weighting mask; (K) DARTEL alignment; (L) AFNI 3dWarpDrive with 12 parameters directly on ROI segmentations; (M) ROI-LDDMM alignment; (N) ROI-Demons (Thirion) alignment. Note the visibility of the collateral sulcus in each of these methods as an indicator of the underlying blur.

instantiating the ROI-AL principles (those based on aligning the ROIs) performed the best. By aligning each participant's hippocampus to a model hippocampus, each participant's perirhinal cortex to a model perirhinal cortex, etc, each participant's MTL structures lined up well with every other participant's MTL structures. In some ways, this should come as little surprise, but even simple algorithms such as the combination of affines (e.g. Law et al., 2005) outperformed the most sophisticated whole-brain grayscale approaches. When the full implementation of ROI-AL is not used but either the regional constraint is applied (e.g. with weighting masks), or a fully deformable method like DARTEL is used, performance can still be improved over standard whole-brain techniques.

Within the ROI-based methods, ROI-Demons(T) using the original Thirion's algorithm for optimization was consistently better than all other methods at aligning regions in the medial temporal lobe, followed closely by ROI-LDDMM. Vercauteren's diffeomorphic ROI-Demons(V) actually performed a bit worse than ROI-LDDMM. Among the diffeomorphic ROI-based methods, there were several factors affecting performance (see Discussion), thus each method can be optimized substantially. For example, while the current version of ROI-LDDMM was not able to match ROI-Demons(T), versions currently in development can perform as well or in some cases better than ROI-Demons.

For some of the regions, e.g. the hippocampus proper, we can also see that DARTEL performed quite well (see Fig. 1). It is clear from these data, however, that traditional methods that use structural whole brain scans for alignment (aside from those using a large or infinite number of degrees of freedom) are only achieving about a 50% overlap across participants in the medial temporal lobe.

Comparison of smoothness of the average images produced by each method (Table 1 and Fig. 2) showed similar results. Within the MTL, the ROI-based approaches outperformed all others and produced the sharpest averaged structural image. Here, ROI-LDDMM produced the smallest change in smoothness from a single subject scan (~11%), followed closely by ROI-Demons(T) (~12%), followed by the multi-affine approach using ROI-AFNI 3dWD with ROI segmentations and 12 parameters (~16%). As with the overlap metric, DARTEL came closest to the ROI-based approaches (~25%). Other methods such as Talairach and SPM's normalization produced images that had about a 30–50% increase in blur from a single subject template.

The diffeomorphic techniques were also tested on their ability to align hippocampal subfields. For the ROI-based approaches, this was done both using the standard approach (single ROI label for the hippocampus) and by assigning each subfield a distinct label and

conducting the alignment based on the subfields as well. Here, we again observed the benefit afforded by the ROI-AL principles as the ROI-based approaches performed the best. We also observed that while some amount of subfield level alignment came along with the smooth diffeomorphic alignment of the hippocampus, the inclusion of separate labels offered substantial benefit. Specifically, the ROI-based demons and LDDMM with subfield segmentations outperformed their counterparts with just the MTL segmentations by almost 30%. In turn, the ROI-based approaches with just the MTL segmentations improved upon the traditional Talairach alignment by an additional 20% (Fig. 3). DARTEL alignment of hippocampal subfields was considerably better than Talairach alignment, however it was still not as accurate as the ROI-based diffeomorphic methods.

As one might expect, some of the methods described are more computationally intensive than others and require more processing time and resources. All of the methods were benchmarked and tested using a dual processor 2 GHz Intel Core Duo Apple iMac equipped with 2 GB of RAM and running Mac OS X version 10.5.4, except for LDDMM analyses which were run on a 2.4 GHz quad-core Intel® Xeon® E7330 node with 4 GB of RAM at the Johns Hopkins University Center for Imaging Science (JHU-CIS) supercomputing facility. The average processing times reported here are based on the approximate time it takes to align a single subject to a template. Template creation time was not included since this is a one-time only procedure that will vary with the number of subjects included in the template. Talairach alignment in AFNI takes only a few seconds per subject once the landmarks are placed (the entire procedure takes on the order of about a minute). AFNI's 3dWarpDrive takes on the order of a minute to align a single subject to a template. ROI-Demons running in MedINRIA takes about 2–5 minutes to align a single subject. SPM's alignment using traditional normalization (with or without weighting) takes on the order of 2–20 minutes depending on the starting estimates. A little bit of manual realignment prior to normalization significantly improves the starting estimates and reduces processing time. SPM's unified segmentation methods and DARTEL take on the order of 1.5–2 hours per subject. Finally, LDDMM-based methods take 4–6 hours per subject and require software and hardware that are currently only available to very few sites (JHU-CIS being one of them).

Discussion

The results presented herein demonstrate that the ROI-AL principles introduced in our earlier work are highly effective at the
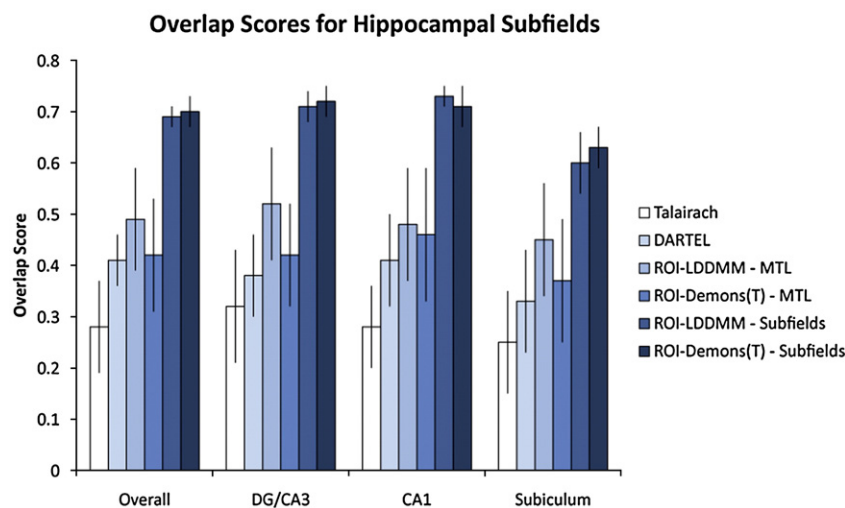

Fig. 3. Subfield alignment performance using Talairach, DARTEL, MTL-based ROI-LDDMM, MTL-based ROI-Demons(T), subfield-based ROI-LDDMM, and subfield-based ROI-Demons (T). MTL-based methods treat the hippocampus as a single ROI, while the subfield methods split up the hippocampus into a DG/CA3 region, a CA1 region, and a subiculum region). ROI-Demons with subfield segmentations yields the highest overlap scores, followed closely by ROI-LDDMM. Error bars are ±1 standard deviation.

alignment of specific brain areas. The first principle states that in the case of limited degrees of freedom, the available alignment power should be applied only to the areas of the brain in which we are interested. For example, if one is specifically interested in the hippocampus, one should not waste any degrees of freedom aligning other regions such as structures in the parietal or occipital lobe. Critically, alignment of parts of the brain outside the areas of interest should not come at the expense of the areas in which one is interested. If you have no real limitation on degrees of freedom (which is the case in methods that utilize the large deformation framework) then this issue is moot. The second principle of the ROI-AL approach is that where one knows something about the underlying anatomical or functional structure of the areas of interest, the alignment algorithm should be allowed to benefit from that information. In our typical application using 3D segmentations as the basis for alignment, this principle encompasses the first as well. However, other applications of this principle are possible. For example, one could place landmarks within or around the borders of the regions of interest when performing landmark-based alignment to help enforce this constraint (Csernansky, et al., 2005; Wang et al., 2003). Likewise, one can also use the ROI approach on 2-dimensional flat maps or spherical representations of the cortical surface (Van Essen, 2004; Fischl et al., 1999) or 3D surface deformation information (e.g. Vaillant et al., 2007, Qiu et al., 2008) to arrive at similar outcomes.

In general, we have shown that applying both of the ROI-AL principles leads to significant improvement in the quality of the alignment. If one is unable to inform the alignment technique with regional information, then a large deformation diffeomorphic technique such as DARTEL offers some advantage over traditional approaches. In the past, we have shown that MRI-LDDMM, another large deformation diffeomorphic approach also outperforms traditional approaches but also falls short of a full ROI-AL implementation's performance (Miller et al. 2005).

It is important to note that the choice of registration method will entirely depend on the application. For example, using a ROI-based method to align MTL segmentations may yield excellent overlap within MTL regions and fail miserably at aligning the rest of the brain where the alignment rapidly becomes only as good as the initial alignment passed into the technique (i.e. Talairach in our case). If one would like to look at whole brain effects here, then alignment using a whole brain method such as DARTEL or MRI-LDDMM may be more appropriate. For studies that are interested in both a whole brain analysis and a regional analysis, two different alignment techniques should be employed. A ROI-based method should be used to align the ROIs and a whole brain method should be used to align the whole brain either in separate analyses or by using a sophisticated whole brain technique's alignment as the starting point for a ROI-based fine tuning (e.g. MRI-LDDMM followed by ROI-LDDMM). For studies that are interested in both whole brain effects and specific regional effects but are not able to manually segment or label ROIs, a method such as DARTEL is likely optimal.

In this study, we have presented results from the ROI-AL framework using the medial temporal lobe structures as a test case. However it is important to note that these principles can be generalized to any other region of the brain. There is nothing special about the medial temporal lobe per se. For example, the same principles can and have been applied to studies of the striatum (Nomura et al., 2007, Mattfeld and Stark, submitted for publication). Also, functional ROI definitions such as the delineation of early visual regions by retinotopic mapping might benefit from these methods. With adequate control over the smoothness of the deformations, any region of the brain can be aligned using this method.

This leads us to the next fundamental issue that the current work highlights. There is a necessary tradeoff between smoothness of the deformations and accuracy of the alignment. Some regions are more complex than others and thus may need more control over constraints on smoothness during alignment to ensure the biological plausibility of the deformations. Consider for example our subfield alignment. Although ROI-Demons(T) outperformed ROI-LDDMM at aligning the MTL structures, when its warp parameters were applied directly to hippocampal subfields it did not perform as well as ROI-LDDMM (Fig. 3). The reason is quite simple. ROI-LDDMM employs a more conservative smoothness constraint making sure that warps within structures are not too extreme. Although ROI-Demons could align hippocampus on top of hippocampus very well, it warped the structure in ways that misaligned its composite subfields. The result of using a much less conservative smoothness constraint in ROI-Demons(T) was that the moving image was able to undergo a slightly odd stretch to minimize the sum of squared error. Of course, once either method is given the subfield information, smoothness does not play as large a role, as indicated by their almost identical performance at aligning the subfields, which was 20% better than alignment using just the MTL structures. It is important to note that one can optimize ROI-LDDMM's parameters to improve performance (we have done so in several tests) by varying the smoothness constraints on the deformation fields. Likewise, one can also increase the smoothness of the fields generated by ROI-Demons(T).

Another practical implication of the smoothness vs. accuracy tradeoff is that the use of relaxed smoothness constraints may only be effective in situations like those presented by our small set of ROIs. Using an aggressive ROI-approach, like our ROI-Demons(T) implementation, we can align a single gyrus (the parahippocampal gyrus) exceptionally well, but this same aggressive approach cannot be readily extended to many gyri. We can get good results in our regions of interest because we are afforded the luxury of being much more liberal with the smoothness constraint. However, when we attempt to use the same liberal thresholds for aligning whole brain scans, the resulting vector fields have highly improbable shearing distortions. Here, in order to model the much more complex structure of the cortical sheet, one has to be much more conservative and apply a higher smoothness constraint. Methods that enforce a higher smoothness constraint such as Vercauteren's diffeomorphic ROI-Demons(V) and ROI-LDDMM result in very smooth deformations but do not result in the best alignment in the MTL.

It is important to note that adjusting any of LDDMM's many parameters would yield different results and can be used to improve its performance (e.g., adjusting the amount of Gaussian smoothing to the vector fields, the use of multiple input channels, and/or the use of several alignment iterations at various levels of smoothness). In fact, by tuning these parameters, LDDMM can be made to outperform ROI-Demons(T) (Michael Miller and Can Ceritoglu, personal communication, July 3, 2008). In all likelihood, the deformations created by such an approach would be more biologically plausible than the more brute-force, single pass of ROI-Demons(T), as the subfield analysis suggested. Also, LDDMM has been successful in modeling clinically meaningful deformations from hippocampal atrophy in Alzheimer's disease (Wang et al., 2003) to cardiac anatomy (Beg et al., 2004). In order to fully assess the biological plausibility of the deformations we would need to determine how each method behaves when we use more complex anatomical challenges, e.g. warping diseased brains with gross shape changes, or modeling subtle deformations in small structures with aging. Although these challenges were not directly assessed in this study, they will be the topic of future investigation.

In this study, we presented two robust metrics for evaluating the quality of registration algorithms. Our gold standard is the overlap score metric, since it provides an accurate measure of how well regions of interest line up on top of one another. If one is interested in functional differences across participants within a region like the hippocampus, one must be able to at least successfully align that region across participants. The smoothness metric provides convergent results and is a reliable way to approximate the amount of tissue type mismatch in the average images. It has also been used in the past

to evaluate registration algorithms' performance (Wu et al., 2006). In addition to the overlap scores and the smoothness metrics, we also plotted the histograms of the cropped average structural images produced by each method. We fit two Gaussians to the gray matter and white matter distributions to see how well each tissue type was represented (CSF was not included due to the lack of appreciable amounts of CSF within the medial temporal lobe masks). This was used as an additional index of image quality, and as expected, the results were largely convergent with the smoothness metric's results. It is worthy of note that this last measure however can be influenced easily by image enhancement techniques such as bias or inhomogeneity correction, which is part of the workflow for some methods (e.g. SPM's unified segmentation and DARTEL) but not others. If one decides to use such a metric, the final deformations must be applied to the raw images and not to the bias-corrected images in order to assess the quality of the histogram. In summary, all of our metrics converged upon the same conclusion that combining the ROI approach with a fully deformable method yields the best results.

Conclusions

To conclude, we have shown that while regions in the medial temporal lobe are typically poorly aligned, the ROI-AL approaches can substantially improve the alignment. As we have previously demonstrated (Miller et al. 2005), this increase in alignment accuracy has a clear effect on the detection power in fMRI studies of the medial temporal lobe. We would encourage researchers who are interested in specific regions of the brain to apply the ROI-AL principles to their alignment techniques. To this end, we have made all of the tools necessary for ROI-AL using the Demons implementation, including tutorials, sample data, models, and instructions on how to obtain source code and executable binaries for the programs available for download from our website at http://darwin.bio.uci.edu/~cestark/roial/roial.html.

References

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. NeuroImage 38, 95–113.
Ashburner, J., Friston, K.J., 1999. Nonlinear spatial normalization using basis functions. Hum. Brain Mapp. 7 (4), 254–266.
Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–851.
Ashburner, J., Neelin, P., Collins, D.L., Evans, A.C., Friston, K.J., 1997. Incorporating prior knowledge into image registration. NeuroImage 6, 344–352.
Bakker, A., Kirwan, C.B., Miller, M.I., Stark, C.E.L., 2008. Pattern separation in the human Hippocampal CA3 and dentate gyrus. Science 319 (5870), 1640–1642.
Beg, M.F., Helm, P.A., McVeigh, E., Miller, M.I., Winslow, R.L., 2004. Computational cardiac anatomy using MRI. Magn. Reson. Med. 52 (5), 1167–1174.
Beg, M.F., Miller, M.I., Trouve, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. 61 (2), 139–157.
Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comp. and Biomed. Res. 29, 162–163.
Csernansky, J.G., Wang, L., Swank, J., Miller, J.P., Gado, M., McKeel, D., Miller, M.I., Morris, J.C., 2005. Preclinical detection of Alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. NeuroImage 25 (3), 783–792.
Duvernoy, H., 1998. The Human Hippocampus, 2nd ed. Springer, New York.
Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284.
Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. 33 (5), 636–647.
Insausti, R., Juottonen, K., Soininen, H., Insausti, A.M., Partanen, K., Vainio, P., Laakso, M., Pitkanen, A., 1998. MR imaging volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. Am. J. Neuroradiol. 19, 659–671.
Kirwan, C.B., Jones, C.K., Miller, M.I., Stark, C.E.L., 2007. High-resolution fMRI investigation of the medial temporal lobe. Hum. Brain Mapp. 28, 959–966.
Law, J.R., Flanery, M.A., Wirth, S., Yanike, M., Smith, A.C., Frank, L.M., Suzuki, W.A., Brown, E.N., Stark, C.E.L., 2005. fMRI activity during the gradual acquisition and expression of paired-associate memory. J. Neurosci. 25, 5720–5729.
Mattfeld, A., Stark, C.E.L., submitted for publication. An fMRI investigation of striatal activity during the learning and expression of arbitrary paired-associate memory.
Miller, M.I., Beg, M.F., Ceritoglu, C., Stark, C.E.L., 2005. Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. Proc. Natl. Acad. Sci. U. S. A. 102 (27), 9685–9690.
Nomura, E.M., Maddox, W.T., Filoteo, J.V., Ing, A.D., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., Reber, P.J., 2007. Neural correlates of rule-based and information-integration visual category learning. Cereb. Cortex 17 (1), 37–43.
Qiu, A., Vaillant, M., Barta, P., Ratnanathar, J.T., Miller, M.I., 2008. Region-of-interest-based analysis with application of cortical thickness variation of left planum temporale in schizophrenia and psychotic bipolar disorder. Hum. Brain Mapp. 29 (8), 973–985.
Stark, C.E.L., Okado, Y., 2003. Making memories without trying: medial temporal lobe activity associated with incidental memory formation during recognition. J. Neurosci. 23 (17), 6748–6753.
Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme, New York.
Thirion, J.P., 1998. Image matching as a diffusion process: an analogy with Maxwell's demons. Med. Image Anal. 2 (3), 243–260.
Vaillant, M., Qiu, A., Glaunes, J., Miller, M.I., 2007. Diffeomorphic metric surface mapping in subregion of the superior temporal gyrus. NeuroImage 34 (3), 1149–1159.
Van Essen, D.C., 2004. Surface-based approaches to spatial localization and registration in the primate cerebral cortex. NeuroImage, 23 (S1), S97–S107.
Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2007. Non-parametric diffeo-morphic image registration with the demons algorithm. In: Ayache, N., Ourselin, S., Maeder, A. (Eds.), Medical Image Computing and Computer-Assisted Intervention — MICCAI. Springer-Verlag, Berlin/Heidelberg, Germany, pp. 319–326.
Wang, L., Swank, J.S., Glick, I.E., Gado, M.H., Miller, M.I., Morris, J.C., Csernansky, J.G., 2003. Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. NeuroImage 20, 667–682.
Wu, M., Carmichael, O., Lopez-Garcia, P., Carter, C.S., Aizenstein, H.J., 2006. Quantitative comparison of AIR, SPM, and the fully deformable model for atlas-based segmentation of functional and structural MR images. Hum. Brain Mapp. 27, 747–754.
Yassa, M.A., Stark, C.E.L., 2008. Multiple signals of recognition memory in the medial temporal lobe. Hippocampus 18, 945–954.
Xiong, J., Gao, J.H., Lancaster, J.L., Fox, P.T., 1995. Clustered pixels analysis for functional MRI activation studies of the human brain. Hum. Brain Mapp. 3, 287–301.