Published in final edited form as:

Neuroimage. 2014 January ; 84: . doi:10.1016/j.neuroimage.2013.09.011.

Evaluation of group-specific, whole-brain atlas generation using Volume-based Template Estimation (VTE): application to normal and Alzheimer's populations

Yajing Zhang^{a,b}, Jiangyang Zhang^b, Johnny Hsu^b, Kenichi Oishi^b, Andreia V. Faria^b, Marilyn Albert^c, Michael I. Miller^{a,d}, and Susumu Mori^{b,d,e,*}

^aDepartment of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD, USA

^bThe Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD, USA

^cDepartment of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

^dCenter for Imaging Science, Johns Hopkins University, Baltimore, MD, USA

^eF. M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, MD, USA

Abstract

MRI-based human brain atlases, which serve as a common coordinate system for image analysis, play an increasingly important role in our understanding of brain anatomy, image registration, and segmentation. Study-specific brain atlases are often obtained from one of the subjects in a study or by averaging the images of all participants after linear or non-linear registration. The latter approach has the advantage of providing an unbiased anatomical representation of the study population. But, the image contrast is influenced by both inherent MR contrasts and residual anatomical variability after the registration; in addition, the topology of the brain structures cannot reliably be preserved. In this study, we demonstrated a population-based template-creation approach, which is based on Bayesian template estimation on a diffeomorphic random orbit model. This approach attempts to define a population-representative template without the cross-subject intensity averaging; thus, the topology of the brain structures is preserved. It has been tested for segmented brain structures, such as the hippocampus, but its validity on whole-brain MR images has not been examined. This paper validates and evaluates this atlas generation approach, i.e., Volume-based Template Estimation (VTE). Using datasets from normal subjects and Alzheimer's patients, quantitative measurements of sub-cortical structural volumes, metric distance, displacement vector, and Jacobian were examined to validate the group-averaged shape features of the VTE. In addition to the volume-based quantitative analysis, the preserved brain topology of the VTE allows surface-based analysis within the same atlas framework. This property was demonstrated by analyzing the registration accuracy of the pre- and post-central gyri. The proposed method achieved registration accuracy within 1 mm for these population-preserved cortical structures in an elderly population.

Corresponding author: Susumu Mori, PhD, The Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, 330B Traylor Building, 720 Rutland Avenue Baltimore, MD 21205, Work: 410-614-2702, susumu@mri.jhu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Volume-based Template Estimation (VTE); study-specific atlas; MRI; volume-surface analysis

Introduction

MRI and modern image analysis techniques have greatly enhanced our abilities to describe group variability in normal populations, and to detect abnormalities in diseased brains. An MR-based brain atlas, often accompanied by structural definitions, serving as the template for image mapping and the anatomical framework of statistical analysis (Evans et al., 2012; Faria et al., 2012; Sowell et al., 2003; Thompson et al., 2004; Thompson et al., 2000; Zhang et al., 2010), is a key component in image analysis in various modalities, such as anatomical MRI, diffusion tensor images, and functional MRI.

Typically, there are two types of atlases, each with different characteristics: the population-averaged atlas, and the single-subject (SS) atlas. Both have advantages and drawbacks, which have been the subject of debate. Single-subject atlases, e.g., (Talairach and Tournoux, 1988), have sharp structural definitions and are often preferred as templates for image mapping using high-order nonlinear image warping (Ceritoglu et al., 2009; Oishi et al., 2009). However, single-subject atlases are often criticized for being biased toward the subject of choice and for lacking group-representative anatomical features. Some of these issues could be partially ameliorated by carefully choosing a subject that satisfies certain criteria of global anatomical features, such as an individual with median brain volume (Wu et al., 2011) or the best individual target brain (Kochunov et al., 2001). However, the former is based on global features that do not necessarily conform to local anatomical features, while the latter results in a non-diffeomorphic image mapping.

In comparison, conventional group-averaged atlases have been generated, based on spatial intensity-averaging, and used to represent the ‘average’ shape of a specific group (Evans et al., 1993; Holmes et al., 1998; Mazziotta et al., 2001; Thompson et al., 2000). For example, the ICBM group-averaged atlas (Mazziotta et al., 2001) was created by linearly averaging 152 normal subjects, and serves as a commonly used coordinate system of the normal population in various MR image analysis tools, such as SPM (Friston et al., 2007), and MriStudio (<https://www.mristudio.org/>). A nonlinear version of the ICBM atlas was also introduced recently (Evans et al., 2012; Fonov et al., 2011), and have been incorporated in image analysis tools, such as FSL.

In the population-averaged atlases, the cross-subject averaging process, with regard to image intensity, results in the contrasts being influenced by two dominant factors: inherent MR contrasts based on the relaxation properties of water molecules, and residual anatomical variability among the subjects after registration. The latter factor leads to smoothed anatomical boundaries for regions with a high degree of variability. The degree of the smoothness depends on the number of averaged subjects, the anatomical homogeneity of the population, and the image transformation algorithms; thus, the degree of smoothness is difficult to predict. The structural blurring contains important information about residual anatomical variability after normalization, and could be beneficial for the accuracy of the volume-to-volume image registration by providing natural weighting toward well-defined (population-preserved) structures (Yeo et al., 2008). However, the utilization of a smoothed image as a template for highly nonlinear transformation on sharp individual images may require careful examination (Wu et al., 2011). In addition, the cross-subject intensity-averaging leads to a potential failure to preserve brain topology for certain structures

(Mangin et al., 2010), e.g., the cortex should be topologically one “sheet,” which may no longer be the case after population averaging.

In this study, we tested an approach to creating population-representative templates that does not rely on the cross-subject intensity-averaging process and which also preserves the brain topology. This atlas creation approach was originally proposed as a Bayesian framework of template estimation (Ma et al., 2008). In the modeling, the brain anatomical shapes were treated as a random orbit under the action of the group of diffeomorphisms (Grenander and Miller, 1998), and the variability of anatomical shapes was balanced based on certain geodesic distance measures (Fletcher et al., 2004; Miller et al., 2002; Pennec, 2006). A Bayesian framework was further integrated to estimate the template, based on an initial value. The paper by (Ma et al., 2008) presented a mathematical derivation of this algorithm, which was theoretically derived to ensure the group representative feature, and to preserve image contrast and the topology with respect to the initial template.

However, this atlas generation approach has been validated on segmented structures, such as the hippocampus and the caudate, but it has not been tested for whole-brain, gray-scale images, in which the complicated topology in the cortical area and the interrelated and neighboring brain structures can undermine the performance of this technique. While the theoretical framework is in accordance with the metric definition in (Ma et al., 2008), the optimized result, considering all structures in the whole brain, has not been proven to be consistent with the observational results. Moreover, the important issues, such as validity, stability, and the potential usefulness of the clinical data, have not yet been addressed.

This paper demonstrates a pipeline for whole-brain, group-specific atlas generation, called Volume-based Template Estimation (VTE), using the technique derived from (Ma et al., 2008). We used VTE to generate study-specific atlases for three groups with distinct pathological conditions and ages: 1) a collection of normal adults (22-38 years old); 2) MRI scans from a normal aging population (60-80 years old); and 3) MRI scans from AD patients (69-82 years old). We tested whether the VTE atlases represented group-averaged shapes and reduced anatomical bias. Quantitative measurements of structural volume, metric distance, displacement vector, and Jacobian were used to examine this feature. Use of this study-specific atlas was demonstrated by combining pre-defined parcellation maps to perform automated structural segmentation. To further demonstrate the advantage of the topology-preservation properties of VTE, surface-based quantitative analyses were performed to measure registration accuracy of the pre- and post-central gyri, on the same dataset used for the volume-based analysis. The integrated volume-surface analysis within the same atlas framework provides unique opportunities for quantitative brain anatomy studies.

Materials and Methods

MRI Data

Datasets from healthy young adults, patients with Alzheimer's disease (AD), and age-matched control subjects were included in this study.

Young adult control—MRI data collected from eighteen healthy subjects (28.6 ± 4.6 years old; eleven males, seven females) were included in the study. Local institutional review board approval and written, informed consent were obtained prior to examination. Data were acquired using a 3T MR scanner (Achieva, Philips Healthcare, Best, The Netherlands). A 3D inversion recovery sequence was played out (TR/TE/TI = 6.7/3.1/842 ms) with a $1.0 \times 1.0 \times 1.2 \text{ mm}^3$ resolution over an FOV of $240 \text{ mm} \times 204 \text{ mm} \times 256 \text{ mm}$, in the sagittal plane. The SENSE acceleration factor was 2 in the right-left direction. A multi-

shot fast gradient echo (TFE factor= 240) was used with a 3-s shot interval and the turbo direction being in the slice direction (right-left). The flip angle=8°. No fat saturation was employed. The total scan time was 5 min 56 s. Image acquisition details can be found in (Landman et al., 2011).

Alzheimer's disease patients and age-matched controls—Subjects were excluded from enrollment if they were under the age of 55, had a history of a neurological disease other than AD, or a history of major psychiatric illness. All subjects provided informed consent prior to the initiation of the study in accordance with the requirements of the Johns Hopkins Institutional Review Board. Consent procedures followed the guidelines endorsed by the Alzheimer's Association for participation of cognitively impaired individuals (Alzheimer's Association, 2004). MRI data from 12 patients (75.0 ± 4.2 years old; 11 males, one female) and 12 age-matched normal subjects (75.0 ± 5.9 years old; six males, six females) were used. T1-weighted images from these subjects were acquired on a 3.0 Tesla scanner (Philips Medical Systems, Best, The Netherlands) at the F.M. Kirby Research Center for Functional Brain Imaging at the Kennedy Krieger Institute. The MPRAGE scan was conducted according to the protocol of the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Jack et al., 2008). The degree of clinical severity of each subject was evaluated by a semi-structured interview (Hughes et al., 1982; Morris, 1993). This interview generates both an overall Clinical Dementia Rating (CDR) rating and a measure known as the CDR Sum of Boxes (CDR-SB). More details can be found in (Mielke et al., 2009).

All three data sets were re-sampled to the dimension of $181 \times 217 \times 181$, with a 1mm isotropic voxel size.

Volume-based Template Estimation (VTE) and atlas generation

The whole-brain intensity images, $I_0 \dots I_{N-1} \in \mathcal{I}$, were skull-stripped using SPM tissue segmentation, followed by manual adjustment. To remove the differences in overall brain size and shape, the skull-stripped images were affine-normalized to the ICBM-152 coordinates by the AIR linear tool (<http://www.loni.ucla.edu/Software/AIR>). Intensity values were adjusted using a linear function to ensure similar intensity histograms for all the images.

The VTE atlas generation procedure is schematically shown in Fig.1. The brain anatomical shapes in the dataset were modeled as an orbit under the action of the group of diffeomorphisms (Grenander, 1994). We randomly selected one subject's image (e.g., I_0) as an initial template, T_0 , which served as a prior of the brain anatomy. The nonlinear space of diffeomorphisms was treated as an infinite dimensional Riemannian manifold. With a suitable choice of metric, the geodesic flows were described by a momentum conservation law (Arnol'd, 1978; Holmes et al., 1998; Miller et al., 2006). Once a template is fixed, the geodesic mapping is completely determined by the momentum at the origin (Miller et al., 2006). Therefore, the initial momentum provides a linear representation of the nonlinear diffeomorphic shape space around the template, and makes it possible to apply linear statistical analysis tools. The final atlas, T , was modeled as an unknown deformation, ϕ_μ , which can be determined by the initial momentum, μ . A Bayesian framework was integrated in the random orbit model to estimate the optimal initial momentum, μ^* , and the noise level, σ (inexact matching component). The orbit is generated from the atlas $T \in \mathcal{I}$, and each of the observed data I_i was a conditional Gaussian random field, with mean field $T \circ \varphi_{\mu_i}^{-1}$ and variance σ^2 . Given the observable data, $I_1 \dots I_{N-1}$, and the initial atlas, T_0 , the process was to determine the optimal initial momentum, μ^* , such that the final atlas can be obtained by $T = T_0 \circ \varphi_{\mu^*}^{-1}$.

This procedure was implemented via the “Mode Approximation EM algorithm.” Details of the implementation can be found in a recent publication (Ma et al., 2008). Specifically, in the i -th iteration, the process was summarized to three steps: 1) perform nonlinear image matching (LDDMM, large deformation diffeomorphic metric mapping (Beg et al., 2005)) from the current template image $T^{(i)}$ to each observation image, $I_1 \dots I_{N-1}$, to estimate the deformations, $\varphi_{\mu_1}^{(i)} \dots \varphi_{\mu_{N-1}}^{(i)}$, of the group; 2) estimate the mean of the deformed observation

$$\bar{T}^{(i+1)} = \frac{\sum_{n=1}^{N-1} |J_n^{(i)}| I_n \circ \varphi_{\mu_n}^{(i)}}{\sum_{n=1}^{N-1} |J_n^{(i)}|},$$

images by ,Where $|J_n^{(i)}|$ denotes the Jacobian determinant of the transformation between the current template and the n -th observation image in the i -th iteration. This serves as the estimated optimal ‘location’ in shape space for the next iteration; and 3) perform weighted-LDDMM from the current template, $T^{(i)}$, to the estimated

‘location’, $\bar{T}^{(i+1)}$, with a Jacobian coefficient $J = \sum_{n=1}^{N-1} |J_n|$ as the weighting in order to estimate the template, $T^{(i+1)}$, for the next iteration. The process terminates when the optimal initial momentum, μ^* , is satisfied by the following expression of the corresponding velocity field, v^* , in which μ and v are two related quantities that can be converted to each other (i.e., $\mu = Av$, where A is invertible with inverse $A^{-1} = K$), as shown below:

$$v^* = \arg \min_{v: \dot{\varphi}_t = v_t(\varphi_t)} \left\{ \int_0^1 \|v_t\|_v^2 dt + \frac{1}{\sigma^2} \left\| \bar{T} - T_0 \circ \varphi_{\mu}^{-1} \right\|_2^2 \right\} \quad (a)$$

The estimated atlas, T , (i.e., the VTE atlas), was obtained with the optimal initial momentum by $T = T_0 \circ \varphi_{\mu^*}^{-1}$. Parcellation maps of the initial atlas, T_0 , were naturally propagated to the newly generated VTE atlas space for automated segmentation.

Population-specific T1 atlases: an application

One of the advantages of the VTE-based method is the ability to generate a population-specific atlas that carries the population features. For 1) young adult subjects, 2) the AD patients, and 3) the age-matched control subjects, initial templates were arbitrarily chosen within each dataset, and three population-specific T1 atlas images were generated using the VTE procedure, as described above.

Generation of conventional atlases

To examine the contrast-preserving feature, both a single-subject atlas and group-averaged atlases were generated using the same normal aging population as was used in the VTE atlas for comparison.

Single-subject atlas (SS)—We randomly selected one subject from the aged control subjects, which was used as a single-subject atlas (SS) for comparison.

Group-averaged atlases (GA)—We created a group-averaged atlas, which represented the average shape and size of the aged control group. The aged controls $I_1 \dots I_N$ ($N=12$) were affine-transformed to the ICBM template for the first iteration, and then iteratively registered to the evolving mean template from the previous iteration, to form an affine

group-averaged atlas (AGA): $I_{AGA} = \frac{1}{N} \sum_{i=1}^N \Psi_i(I_i(x))$, where $\Psi_i(\bullet)$ is the i -th affine transformation to transform image I_i to the ICBM-152 coordinates. Similarly, a nonlinear group-averaged atlas (NGA) was generated by averaging, after nonlinear mappings, to the ICBM-152 coordinates. This procedure was iterated until the final image converged. We used LDDMM as the nonlinear mapping tool (Beg et al., 2005)

VTE atlases using different initial templates

While the initial template was chosen randomly, it is important to evaluate the dependency on the choice of initial template. To quantitatively evaluate the anatomical bias of VTE atlases generated from different initial templates, we performed volume-based and surface-based evaluations. For the volume-based analysis, twelve VTEs were created independently from twelve different initial templates and their dissimilarity was evaluated by measuring voxel-by-voxel intensity variability. We set the initial template, T_0 , as each of the images in the aged control dataset, and repeated the VTE process one by one. This yielded twelve VTE atlases from the original dataset ($N=12$). The spatial average of VTE results,

$\bar{I}_{VTE} = \frac{1}{N} \sum_{i=1}^N I_{VTE_i}$, and the standard deviation map $SD_{VTE} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (I_{VTE_i} - \bar{I}_{VTE})^2}$ were calculated. Similarly, the AD-specific VTE atlases were generated using each subject as the initial template, and the mean and SD maps were calculated.

For the surface-based analysis, the GM/WM boundary surfaces were generated from each VTE image using FreeSurfer (Dale et al., 1999). After obtaining each VTE's whole-brain surface, pre- and post-central gyri sub-surfaces were semi-automatically delineated, based on the definitions described elsewhere (Zhong et al., 2010). These structures were chosen because they are anatomically consistent across subjects and readily identifiable on the cortical surface. The gyri delineation method was based on dynamic programming techniques (Khaneja et al., 1998; Ratnanather et al., 2003) and was performed using the software BrainWorks (<http://cis.jhu.edu>). We selected one VTE as a reference and calculated the surface-to-surface distance (SSD) between the reference and the rest of the VTEs, for both pre- and post-central sub-surfaces. The SSD was defined by a Hausdorff distance between the sets of vertices on a pair of sub-surfaces (BrainWorks, <http://cis.jhu.edu>).

Quantitative validation of the group-representative feature of VTE

The structure representativeness of VTE was measured by 1) structural volumes for sub-cortical brain structures; 2) metric distances, which geodesically quantify the amount of deformation between each subject image and the atlas image; and 3) deformation measures, including displacement vector and Jacobian determinant.

Volume measure—The structural volume is an observable measure of whether the structures in the VTE atlas are good representations of the corresponding dataset. In order to make the regional structural volumes comparable, the whole-brain images were first transformed to the ICBM-152 coordinates using affine transformation. Then, 24 sub-cortical brain structures were manually delineated on all subject images for both the AD and age-matched controls. Table 1 shows a list of the segmented sub-cortical structures. VTE atlases were segmented by initially propagating the parcellation of the initial template, followed by manual adjustment. The criteria for manual structural delineation closely followed the method of parcellation in the SS atlas (Oishi et al., 2009). Two raters (J.H. and Y.Z., examined by K.O., a neurologist) performed the manual delineation independently. Volumes

of each segmented structure were calculated using “ROIeditor” (<https://www.mristudio.org/>). For each structural volume, the student's t-test was performed between the mean of the dataset and the VTE atlas.

Metric measure—Given two images, I_i and I_j , the diffeomorphic metric mapping generates a flow from one to the other in a metric shape space. The flow starts from the identity map, Id , and is associated with the velocity vector field, $v_t, t \in [0,1]$. We obtained a metric distance, d_{ij} (Beg et al., 2005; Miller et al., 2002), reflecting the geodesic length of the diffeomorphic mapping between the two images on an overall image scale. This was obtained by:

$$d_{I_i, I_j} = \inf \left\{ d(Id, \varphi), I_j = I_i \circ \varphi^{-1} \right\} = \inf \left\{ \int_0^1 \|v_t\|_v dt \right\} \quad (b)$$

where $\|v_t\|_v$ is an appropriate Sobolev norm on the velocity field. In the age-matched controls, we randomly chose two subjects (labeled as SS#1, SS#2) as examples of two SS atlases, to compare their performances with the VTE atlas. The metric distances between each subject image and the atlas (SS#1, SS#2 or VTE) were calculated by equation (b). The mean and standard deviation of the metric distances to each subject image were calculated as:

$$\bar{d}_{atlas^{(k)}} = \frac{1}{N-1} \sum_{i=1, i \neq atlas^{(k)}}^N d_{atlas^{(k)}, I_i},$$

$$SD(d_{atlas^{(k)}}) = \sqrt{\frac{1}{N-2} \sum_{i=1, i \neq atlas^{(k)}}^N \left(d_{atlas^{(k)}, I_i} - \bar{d}_{atlas^{(k)}} \right)^2}$$

where $k=SS\#1, SS\#2$ or VTE . The subject image that served as the template was excluded from the groups to remove any evaluation bias. Two-sample tests were performed to statistically judge the differences.

Deformation measures—Given the three cases of mapping templates (SS#1, SS#2, VTE), we derived the displacement vectors and the Jacobian determinants to compare deformation properties. For each candidate template, LDDMM was used to map each subject image to the template. For each voxel, the Euclidean distance between the subject and its deformed image was obtained as the magnitude of the displacement vector, $\vec{dv}$. The orientation of this vector was also determined. The average displacement vector (ADV) was

generated as $\vec{dv}_{atlas^{(k)}} = \frac{1}{N-1} \sum_{i=1, i \neq atlas^{(k)}}^N \vec{dv}_{atlas^{(k)}, I_i}$, where k denotes SS#1, SS#2, or VTE. The subject image that served as the template was excluded from the groups to remove any evaluation bias. Furthermore, we calculated the log of the Jacobian determinant,

$\log J_{atlas^{(k)}, I_i}$ of each template to each subject image I_i . Similarly, the average of the log-Jacobian was generated by

$$\overline{\log J_{atlas^{(k)}}} = \frac{1}{N-1} \sum_{i=1, i \neq atlas^{(k)}}^N \log J_{atlas^{(k)}, I_i}.$$

Evaluation of mapping accuracy in largely deformed images

Images with severe morphological deformation were often more difficult to register accurately. To measure their registration accuracy, we registered subjects to 1) a previously published SS atlas (Eve atlas), 2) SS atlases where each subject served as the atlas, and 3) a VTE atlas, using the intensity-based image mappings (LDDMM). Two independent raters (Y.Z. and J.H., examined by K.O., a neurologist) manually delineated the 24 sub-cortical structures (list of structures shown in Table 1) for each AD image, following the same criteria described above (Oishi et al., 2009). The *manual delineation* served as the “gold” standard for the accuracy measurement. The AD-specific VTE atlas was segmented based on the propagated parcellation, followed by a manual correction process. We superimposed the deformed parcellations of the template onto each AD image to segment the twenty-four structures automatically (*automated delineation*), and compared the degree of spatial matching using kappa values. The kappa values were calculated and evaluated according to (Landis and Koch, 1977).

Surface-based analysis of the cortical registrations

To evaluate the registration quality of the cortical areas, surface-based analysis was also performed. The surface generation scheme followed the procedure described above. We used the normal aging population and chose one of the subjects as an initial template to create a VTE atlas. The GM/WM boundary surfaces of all participants were generated, and pre- and post-central gyri sub-surfaces were semi-automatically delineated before and after the registration to the VTE atlas. Surface-to-surface distances (SSD) from the VTE atlas to the original and registered subjects were calculated.

Results

VTE-based T1 atlas and conventional atlases

Fig.2 compares a randomly selected subject, two types of population-averaged atlases (AGA and NGA), and the VTE atlas. Because all group-wise atlases (AGA, NGA, and VTE) were linearly normalized to the ICBM-152 brain, their whole-brain volumes were similar (see Table 2). Visually, VTE, AGA, and NGA shared a similar brain shape, including the shape of the lateral ventricles; but, the image sharpness was markedly different. AGA (Fig.2B) showed blurry images due to spatial smoothing of images. NGA (Fig.2C) had slightly improved image sharpness due to more accurate registration by non-linear transformation, but still showed a certain amount of degradation through the spatial averaging. The VTE atlas (Fig.2D), conversely, preserved the sharp contrast, as in the original single-subject image (Fig. 2A). A large number of gyri and sulci in the cortex that were smeared out in conventional atlases were preserved in the VTE atlas. The trade-off of this advantage is the inevitable template-dependent bias that remained after the estimation, as shown in Fig. 4. The 3D spatial representation of the lateral ventricles in the VTE atlas was superimposed on the lateral ventricles of the dataset (Fig.2E), and lay in the center of the group-averaged outline. The lateral ventricle (LV) volume was 38,320 mm³ for VTE, and 38,720 ± 3,058 mm³ for the population, indicating the good representation of the dataset by VTE for both the location and volume of the lateral ventricle.

VTE atlases for three populations with varied ages and pathological conditions

Using VTE, group-specific T1 atlases for 1) the young adults, 2) Alzheimer's disease (AD) patients, and 3) age-matched controls are shown in Fig.3. The atlas images preserve the original image sharpness. The atlas from the age-matched controls had slightly more enlarged lateral ventricles than those in the young adult group, and the atlas from the AD population further showed dramatically enlarged lateral ventricles. The computation time for each VTE task was approximately 30 hours using a remote server with 24 CPUs in a Linux-based cluster computer.

Effect of initial template for VTE atlas generation

The dependence on the initial template for the VTE method is detailed in Fig.4 and Fig.5. Fig.4A (col. 1-12) shows the original dataset (top row) and VTE atlases (bottom row) using each original subject as the initial template. Fig.4A a-d shows the mean and standard deviation of the initial image set and the VTE set. As expected, the mean VTE (Fig.4A-c) is less blurry than the mean of the dataset (Fig.4A-a; equal to the AGA in Fig.2B). In the standard deviation image of VTE results (Fig.4A-d), variation at sub-cortical regions was close to zero. In cortical areas, the standard deviation was reduced to about one-third the standard deviation range of the original dataset (Fig.4A-b). In Fig.4B, results from the AD population showed results similar to that of the normal elderly population (Fig. 4A).

Fig.5 further quantified the amount of variability observed in the cortical areas using a surface-to-surface distance (SSD). In Fig.5A (A1-A6), one VTE result was selected as a reference (red curve) and the remaining VTE results (cyan curves) were superimposed for the pre- and post-central gyri. All VTE gyri surfaces closely overlapped each other. Fig.5B shows the mean SSD across subjects, which was 1.15 ± 0.30 mm for the pre-central gyri and 0.97 ± 0.16 mm for the post-central gyri. For the two major sulci structures, the non-zero distance values encode the degree of anatomical bias depending on the choice of the initial template. Fig.5C displays the averaged spatial distribution of the SSD on the pre-and post-central gyri surfaces. In most areas of the surface, the SSD values were close to 0.5 mm (dark blue in color), and regions with large SSD (red regions) were concentrated at the edge of the defined surfaces.

Quantitative evaluation of structure representations in VTE atlases

A. Structural volume (ROI-based)—Through manual segmentations of 24 sub-cortical structures, we compared the ROI volumes of the dataset, the VTE results (generated by different initial templates), and an SS atlas. The subject images were under ICBM-152 coordinates and their whole-brain volumes had been normalized (Table 2). In the aged control population, the 24 mean volumes from the VTE atlases (Fig. 6A, black bars) approximated the mean volumes of the dataset (white bars). Statistical tests were performed between the mean of the dataset and that of VTE atlases for each structural volume, and no significant difference in their means was reported. The standard deviation (STD) for VTE atlases was only 53% of that of the dataset. For the sub-cortical tissue structures, the coefficient of variance (standard-deviation/mean) for VTE atlases was $9.0 \pm 2.6\%$. This indicates a reduced dependence on the choice of the initial template for the generation of the VTE atlas. In the sub-divisions of the lateral ventricle structure, where there is a larger variability among subjects (Fig.6A rightmost ten columns), the VTE results also provided a good estimate of the average of the dataset. A small mis-match between the mean of the dataset and the mean of the VTE atlases existed in some structures, but were within one standard deviation of the VTE atlas set. The volumes of a single subject (bars with dots) are also shown in Fig. 6A, and provided a straightforward comparison of biased volume representations for the dataset. Structures such as the caudate, the thalamus, and the amygdala had volumes 15%-25% larger than the dataset mean, while several ventricle-

related structures had volumes 20%-70% smaller than the dataset mean. The volume measurements for the AD population yielded similar results (see Fig.6B). The results indicated an unbiased volume representation from the VTE atlas.

B. Metric distance (whole-image-based)—Based on the random orbit model in diffeomorphic mapping, the metric distance summarizes and quantifies the amount of nonlinear deformation between two whole-brain images as a scalar. Fig. 7A-7C demonstrates the individual metric distances (MD) from a template to the aged control subjects. Fig. 7D provides a statistical summary of the metric distance comparison. On average, the metric distances from the VTE to the dataset (Fig.7C) were smaller than those from the SS to the dataset (Fig.7A or 7B). The mean metric distance was significantly smaller for VTE than for either of the SS cases ($p < 0.05$).

C. Displacement and Jacobian coefficient (voxel-based)—We used both displacement vectors and Jacobian determinants to show the information from the deformation between the template and targeted subjects. Fig.8 compares the magnitude of the averaged displacement vector (ADV) (Fig. 8A) and the averaged log-Jacobian (Fig. 8B) when using SS#1, SS#2, and VTE as templates. In Fig.8A, the ADV magnitude using VTE (Fig.8A, col. 2, 4) shows much smaller values (close to zero) than SS atlases in most regions (Fig.8A, col.1,3). This indicates that VTE was close to the group center and required the least voxel displacements to register the population. For the volumetric changes, large log-Jacobian absolute values using SS atlases (Fig.8B, col.1), such as those observed at the frontal horn of the lateral ventricles, indicated that this particular SS atlas (SS#1) had a smaller frontal part of the ventricles compared to all the other subjects in the population, and therefore, was biased from the center of the population. Log-Jacobian using VTE had values close to “0,” indicating that the VTE was close to the group center.

Evaluation of the registration performance using the VTE atlas

Fig. 9 shows kappa values evaluating the registration accuracy of the AD patient images to 1) each subject image (as a within-group, single-subject template), 2) the VTE atlas, and 3) the Eve atlas (Oishi et al., 2009). The kappa values showed a large variation across different templates and different structures. For example, for most structures, the kappa values from the registration using Sub#12 as a template were higher than those when using Sub#2 as template. Overall, the VTE atlas yielded the best registration performance for the examined brain structures in terms of kappa values. Looking at individual structures, the VTE atlas showed higher averaged kappa values than the single-subject templates.

Results of cortical surface registration

Surface-to-surface distance was used as a quantitative measure of registration accuracy for the pre- and post-central gyri. In Fig.10, two panels show the measurements for the pre- and post-central gyri. The mean SSD was 1.92 ± 0.56 mm for the pre-central gyrus after linear alignment, and was improved to 0.99 ± 0.16 mm after registration to VTE; the mean SSD for the post-central gyrus was 1.89 ± 1.09 mm, and was improved to 0.79 ± 0.10 mm (Fig. 10E). The improvement in the results was found to be significant ($p \ll 0.05$). In Fig.10-FG, after registration by VTE, the SSD on most areas of the surface was decreased to a scale of 0.5 mm (dark blue), and the area with large SSD values was distributed at the margin of the surface.

Discussion

We have validated a group-specific atlas generation method (a.k.a. VTE) based on a template estimation algorithm described earlier (Ma et al., 2008), and tested it with datasets

from healthy young adult subjects, Alzheimer's disease patients, and age-matched normal subjects. The VTE algorithm was fundamentally based on a fluid dynamic model and the action of the group of diffeomorphisms (Grenander and Miller, 1998). The variability of anatomical shapes was balanced, based on certain geodesic distance measures (Fletcher et al., 2004; Miller et al., 2002; Pennec, 2006). This type of framework has been demonstrated to preserve original image contrast and to provide an unbiased shape (Joshi et al., 2004), but the result is neither guaranteed to be in the same shape space as the initial atlas (Joshi et al., 2004), nor is it robust enough to shape outliers (Avants and Gee, 2004). Several follow-up papers investigated and utilized this method in different ways. A paper by (Beg and Khan, 2006) proposed a method with a similar cost function and optimization procedure, which has been applied to heart and caudate segmentation. (Geng et al., 2009) generated an implicit reference image, based on a cost function that considers the displacement field as a kind of metric measure. (Wu et al., 2011) developed a mean-image estimator, which helped to obtain a sharper mean template during the metric averaging. The robustness of a metric-based averaging framework with outlier subjects has been investigated and improved by (Fletcher et al., 2009).

A Bayesian framework was combined with metric averaging for atlas estimation by iteratively deforming an initial template to the 'center' of the population (Allasonnière et al., 2007; Ma et al., 2008; Ma et al., 2010; Zhang et al., 2011), with the metric defined as the length of the geodesic curve in diffeomorphic mapping (Miller et al., 2002; Younes, 1998). The result remains in the orbit of the initial atlas in shape space, with image sharpness preserved. By integrating a Bayesian framework and introducing a noise level between the target and deformed template images (which was often lacking in earlier models (Joshi et al., 2004)), both the optimal initial momentum and the noise level can be estimated, making the algorithm more robust. The procedure has also been utilized to generate a few sub-cortical volume templates individually for a shape-based classification application (Qiu et al., 2010). However, it has not been tested on whole-brain images.

In this study, our results are based on a whole-brain scale and were tested with three populations with different anatomical features. An anatomical parcellation map, stored in the initial template, was translated into a VTE atlas, providing automated atlas-based segmentation. In addition, the topology-preserving property of VTE allowed us to perform volume- and surface-based analysis under the same atlas framework.

Validation and evaluation of the proposed VTE atlas

The results of our atlas generation method confirmed that there was a good representation of the anatomy within a group. In the scope of the whole-brain image, the representativeness was defined as the ability to represent a less biased anatomical feature, such as structural volumes and the amount of deformation needed to transform subjects to the atlas. In our studies, the mean volumes of VTE results approximated the mean volumes of the dataset well. There were small differences (within 10% of the mean volume), possibly because of optimization on a whole-brain VTE, which takes full account of all structures implicitly, rather than each structure independently (Qiu et al., 2010). The neighboring structures could be an important anatomical constraint, which might compromise the ability of certain structures to reach their individual, theoretically unbiased shapes. In addition to volume measures, metric distance derived from the diffeomorphic mapping, displacement, and Jacobian from the deformation were also used to evaluate the group-representative feature. All results show minimized cost for the VTE compared to the SS atlas.

The real usefulness of atlases should be judged by actual applications; thus, we also evaluated the registration performance of the VTE. When anatomical differences between two images become larger, as occurs during pathological progressions (e.g., Alzheimer's

disease-induced tissue shrinkage), the image matching by the SS atlas might no longer reach a good level of accuracy. VTE provides a tool with which to minimize the total deformation of the group and facilitates “easier” image matching. To confirm this, we performed automated segmentation on 12 manually pre-parcellated AD subjects, and performed a leave-one-out analysis. We found that, using an external SS atlas from a young adult population gave the worst registration accuracy (Fig.9, pink cross markers), since its structure is vastly different from the AD population. Furthermore, VTE yielded the best overall registration performance, supporting our hypothesis.

Among conventional group-averaged atlases (Fig.2B-2C), the AGA has significant spatial smoothing, and the degree of blurring effect increases along with the increasing number of subjects. The NGA and iterative NGA (Fonov et al., 2011) have sharper anatomical definition than the AGA. Between NGA and VTE, the residual registration errors have a very different impact: the cross-subject averaging process in the NGA introduces a certain degree of spatial blurring, while, in the VTE, the errors lead to anatomical bias, specific to the choice of initial template, while preserving the contrast.

With regard to the impact on registration accuracy, a direct comparison between the VTE and the NGA might be a point of interest. The comparison results are shown in the Supplemental Materials, in which a study-specific NGA was created as described in the Methods section, and used as a registration template. This type of comparison between two different templates is, in general, difficult because the results are heavily influenced by the way the anatomical structures are defined in the templates. For VTE analysis, we used one of the 12 hand-segmented data as the initial template and used the remaining 11 hand-segmented data as targets for accuracy measurements. The results are, thus, the combination of the true registration accuracy and reproducibility of the hand-segmentation across 12 subjects. For the NGA, we needed to re-define the 24 structures in the NGA with different noise and contrast properties. Some structures with relatively large anatomical variability among the population, such as the ventricles, have a high level of freedom for manual definition due to a blurred anatomical boundary. We attributed the lower registration quality with the NGA, especially for the posterior lateral ventricles, to the difficulty of manual boundary definition. In this analysis, we also tried an alternative approach, in which the structural labels in the NGA were created from the probabilistic map generated from the 11 pre-segmented data. The results showed comparable registration accuracy between the VTE and the NGA across all structures. This analysis, however, gave a slight advantage to the NGA because the averaging of labels from 11 data could have reduced the random errors of manual delineation, while, in real studies, such probabilistic labels from all data are not available. Most likely, structural labeling is performed only once for an initial template (on VTE) or for the averaged image (on NGA).

Effects from the initial template

Under a Bayesian framework, a prior template was used to constrain the topology to the human brain anatomy. There are many ways of choosing the initial template. A median image from the group may be a good choice for such an initial value (Wu et al., 2011). Considering additional information, such as parcellation information, choosing a pre-defined atlas with rich parcellation as the initial template may be more beneficial.

For different choices of initial template, if the selections possess different topological structures of the brain, which is often the case, then, VTE results would inevitably adopt certain structural differences due to the topology-preserving characteristics of the diffeomorphic mapping. This would make the results biased toward the geometry of the initial template, especially for the cortical structures. In our studies, the VTE results provided a stable shape for sub-cortical structures, regardless of the large variability in

initial templates. The brain architecture converged and was not biased when different initial values were chosen. In the cortical area, we quantified the bias using surface-to-surface distance (SSD), as shown in Fig. 5, which was on a scale of 1 mm.

Surface-based analysis using a VTE atlas

There is a potential problem when performing surface-based analysis using group-averaged atlases, because the diffeomorphism is not guaranteed during the spatial averaging, and therefore, the group-averaged atlases may lack the inter-subject correspondence in the cortical folding (Evans et al., 2012; Mangin et al., 2010). A different strategy is to generate cortical surfaces from original subjects and to build the averaged surface template (Lyttelton et al., 2007; Van Essen, 2005), regardless of the constraints from volumetric intensity-matching throughout the brain volume (Evans et al., 2012). This lack of constraints in the process of “averaging on surfaces” makes the result different from the result of “extracting the surface from the average volumetric template,” as we have performed with the VTE atlas. Therefore, the “average of surfaces” atlas may not be compatible with the coordinates generated by the volumetric-averaged atlas. Rich information from the cortex (e.g., local functional features) and information from the deep brain (e.g., fiber connectivity) are not straightforwardly integrated.

Our atlas provides a framework by which to integrate volume (for internal structures such as sub-cortical regions) and surface (for cortex) analysis within the coordinates. Since the VTE was obtained through a diffeomorphic mapping from the initial template with pre-determined topology, the cortex of the VTE atlas remained one manifold, which made the extraction of surface coordinates and the analysis convenient. In our study, we quantified the registration errors of the pre- and post-central gyri as a mean SSD of 0.99 mm and 0.79 mm, respectively. The registration error (SSD) distributed on each vertex of the surface patch was also mapped in the VTE atlas coordinates (Fig.10-FG). This type of measurement would be difficult if an NGA were used as a registration reference.

Limitations of the VTE approach

It should be noted that the VTE approach has its own limitations. First, the accuracy and robustness of the approach hinge on the quality of image mapping. This could be especially troublesome if a patient group with severe abnormalities, such as stroke or brain tumor patients, had to be mapped to a template from a normal population.

Second, a potential bias may arise in VTE results from regions with large anatomical variability or multiple subtypes of morphology. In the cortical area, where the gyri and sulci from each subject vary dramatically, the VTE result is expected to have a certain bias adopted from the initial template. This is a limitation that is inevitable for any topology-preserving mappings. In practice, a VTE atlas is especially beneficial when the study is focused on the core white matter and sub-cortical structures. For application to the two major cortical regions, we have quantified this bias effect at a scale of 1 mm. Users can choose the VTE method if this tolerance is acceptable for the application.

The other region that may be problematic is the posterior horn of the lateral ventricles (Djamanakova et al., 2013). For example, when the structure has three subtypes in a population while using a uniform atlas for registration (e.g., see Fig.11), mis-segmentation is expected in the subjects who possess subtypes that are different from that of the atlas. This could be the main reason for the extremely low kappa values in the structure delineated by the red arrow in Fig.9. Despite this, a VTE atlas is the optimal choice to achieve a balanced and the best-averaged kappa values. In addition, other types of methods may provide better representations of the population. For example, an atlas stratification method can determine

the number of atlases (modes) needed for the whole population (Blezek and Miller, 2007). (Jia et al., 2010) used self-organized registration, in which the local data distribution within the neighboring subjects was explored and used to guide the registration. Furthermore, if the goal is purely to segment the ventricle-related structures accurately, and no common coordinate is needed in a particular study, a multi-atlas framework (Aljabar et al., 2009; Heckemann et al., 2006; Isgum et al., 2009; Jia et al., 2012; Lötjönen et al., 2010; Rohlfing et al., 2004; Wang et al., 2011) may also be appropriate to solve this problem.

The computation time for the VTE atlas was comparable to that of other iterative atlas generation methods. In the VTE procedure, the result converges quickly within 10 iterations for whole-brain shapes. The overall rate-limiting step is the computation speed of the individual nonlinear mapping step (LDDMM), which is typically 10 mins for the image size we used in this paper, using remote computing by Diffeomap (<https://www.mristudio.org/>).

In conclusion, we have evaluated a whole-brain, group-specific atlas generation method (VTE) using three populations with different degrees of atrophy. The results showed that the VTE atlases presented group representative shapes while preserving image contrast and anatomical topology. The amount of bias due to the choice of the initial template was $9.0 \pm 2.6\%$ for the sub-cortical tissue structures, and within 1 mm for motor and sensory cortical surface distances. The improved registration accuracy using the VTE was confirmed, compared to study-specific, single-subject templates. Because of its topology-preserving nature, the VTE can be used as a template for volume- and surface-based integrative analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This publication was made possible by grants R01AG020012 (SM), R01EB000975 (MIM), R01HD065955 (KO), R21AG033774 (KO) from the National Institutes of Health, and P41EB015909 from the National Institute of Biomedical Imaging and Bioengineering. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NIH. The terms of this arrangement are being managed by Johns Hopkins University in accordance with its conflict of interest policies. The authors thank Dr. Laurent Younes and Dr. Tilak Ratnanather for their helpful suggestions.

References

- Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage*. 2009; 46:726–738. [PubMed: 19245840]
- Allasonnière S, Amit Y, Trouvé A. Towards a Coherent Statistical Framework for Dense Deformable Template Estimation. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*. 2007; 69:3–29.
- Arnol'd, VI. *Mathematical methods of classical mechanics*. Second. Vol. 1989. Springer; 1978.
- Avants B, Gee JC. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *NeuroImage*. 2004; 23(1):S139–S150. [PubMed: 15501083]
- Beg, MF.; Khan, A. Computing an average anatomical atlas using LDDMM and geodesic shooting. *Biomedical Imaging: Nano to Macro, 2006 3rd IEEE International Symposium on*; 2006. p. 1116–1119.
- Beg MF, Michael IM, Alain T, Laurent Y. Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *Int J Comput Vision*. 2005; 61:139–157.
- Blezek DJ, Miller JV. Atlas stratification. *Medical Image Analysis*. 2007; 11:443–457. [PubMed: 17765003]

- Ceritoglu C, Oishi K, Li X, Chou MC, Younes L, Albert M, Lyketsos C, van Zijl PCM, Miller MI, Mori S. Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging. *NeuroImage*. 2009; 47:618–627. [PubMed: 19398016]
- Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*. 1999; 9:179–194. [PubMed: 9931268]
- Djamanakova A, Faria AV, Hsu J, Ceritoglu C, Oishi K, Miller MI, Hillis AE, Mori S. Diffeomorphic brain mapping based on T1-weighted images: Improvement of registration accuracy by multichannel mapping. *J Magn Reson Imaging*. 2013; 37:76–84. [PubMed: 22972747]
- Evans, AC.; Collins, DL.; Mills, SR.; Brown, ED.; Kelly, RL.; Peters, TM. 3D statistical neuroanatomical models from 305 MRI volumes. *Nuclear Science Symposium and Medical Imaging Conference*, 1993., IEEE Conference Record; 1993. p. 1813-1817.
- Evans AC, Janke AL, Collins DL, Baillet S. Brain templates and atlases. *NeuroImage*. 2012
- Faria AV, Joel SE, Zhang Y, Oishi K, van Zijl PCM, Miller MI, Pekar JJ, Mori S. Atlas-based analysis of resting-state functional connectivity: Evaluation for reproducibility and multi-modal anatomy-function correlation studies. *NeuroImage*. 2012
- Fletcher PT, Conglin L, Pizer SM, Sarang J. Principal geodesic analysis for the study of nonlinear statistics of shape. *Medical Imaging, IEEE Transactions on*. 2004; 23:995–1005.
- Fletcher PT, Venkatasubramanian S, Joshi S. The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage*. 2009; 45:S143–S152. [PubMed: 19056498]
- Fonov V, Evans AC, Botteron K, Almli CR, McKinsty RC, Collins DL. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*. 2011; 54:313–327. [PubMed: 20656036]
- Friston, KJ.; Ashburner, JT.; Kiebel, SJ.; Nichols, TF.; Penny, WD. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier; Amsterdam, The Netherlands: 2007.
- Geng X, Christensen GE, Gu H, Ross TJ, Yang Y. Implicit reference-based group-wise image registration and its application to structural and functional MRI. *NeuroImage*. 2009; 47:1341–1351. [PubMed: 19371788]
- Grenander, U. *General Pattern Theory*. Oxford Univ Press; 1994.
- Grenander U, Miller MI. Computational anatomy: An emerging discipline. *Quarterly of applied mathematics*. 1998; 56:617–694.
- Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*. 2006; 33:115–126. [PubMed: 16860573]
- Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. Enhancement of MR Images Using Registration for Signal Averaging. *Journal of Computer Assisted Tomography*. 1998; 22:324–333. [PubMed: 9530404]
- Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry*. 1982; 140:566–572. [PubMed: 7104545]
- Isgum I, Staring M, Rutten A, Prokop M, Viergever MA, van Ginneken B. Multi-Atlas-Based Segmentation With Local Decision Fusion - Application to Cardiac and Aortic Segmentation in CT Scans. *Medical Imaging, IEEE Transactions on*. 2009; 28:1000–1010.
- Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, L Whitwell J, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DLG, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Gunnar K, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*. 2008; 27:685–691. [PubMed: 18302232]
- Jia H, Wu G, Wang Q, Shen D. ABSORB: Atlas Building by Self-organized Registration and Bundling. *NeuroImage*. 2010; 51:1057–1070. [PubMed: 20226255]
- Jia H, Yap PT, Shen D. Iterative multi-atlas-based multi-image segmentation with tree-based registration. *NeuroImage*. 2012; 59:422–430. [PubMed: 21807102]
- Joshi S, Davis B, Jomier M, Gerig G. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*. 2004; 23:S151–S160. [PubMed: 15501084]
- Khanuja N, Miller MI, Grenander U. Dynamic programming generation of curves on brain surfaces. *Ieee Transactions on Pattern Analysis and Machine Intelligence*. 1998; 20:1260–1265.

- Kochunov P, Lancaster JL, Thompson P, Woods R, Mazziotta J, Hardies J, Fox P. Regional spatial normalization: toward an optimal target. *J Comput Assist Tomogr*. 2001; 25:805–816. [PubMed: 11584245]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
- Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IAL, Farrell JAD, Bogovic JA, Hua J, Chen M, Jarso S, Smith SA, Joel S, Mori S, Pekar JJ, Barker PB, Prince JL, van Zijl PCM. Multi-parametric neuroimaging reproducibility: A 3-T resource study. *NeuroImage*. 2011; 54:2854–2866. [PubMed: 21094686]
- Lötjönen JMP, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, Rueckert D. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage*. 2010; 49:2352–2365. [PubMed: 19857578]
- Lytellon O, Boucher M, Robbins S, Evans A. An unbiased iterative group registration template for cortical surface analysis. *NeuroImage*. 2007; 34:1535–1544. [PubMed: 17188895]
- Ma J, Miller MI, Trouv A, Younes L. Bayesian template estimation in computational anatomy. *NeuroImage*. 2008; 42:252–261. [PubMed: 18514544]
- Ma J, Miller MI, Younes L. A Bayesian Generative Model for Surface Template Estimation. *International Journal of Biomedical Imaging*. 2010; 2010
- Mangin JF, Jouvent E, Cachia A. In-vivo measurement of cortical morphology: means and meanings. *Current Opinion in Neurology*. 2010; 23:359–367. [PubMed: 20489617]
- Mazziotta JC, Toga AW, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 2001; 356:1293–1322.
- Mielke MM, Kozauer NA, Chan KCG, George M, Toroney J, Zerrate M, Bandeen-Roche K, Wang MC, vanZijl P, Pekar JJ, Mori S, Lyketsos CG, Albert M. Regionally-specific diffusion tensor imaging in mild cognitive impairment and Alzheimer's disease. *NeuroImage*. 2009; 46:47–55. [PubMed: 19457371]
- Miller MI, Trouv A, Younes L. On the Metrics and Euler-Lagrange Equations of Computational Anatomy. *Annual Review of Biomedical Engineering*. 2002; 4:375–405.
- Miller MI, Trouv A, Younes L. Geodesic Shooting for Computational Anatomy. *Journal of Mathematical Imaging and Vision*. 2006; 24:209–228. [PubMed: 20613972]
- Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*. 1993; 43:2412–2414. [PubMed: 8232972]
- Oishi K, Faria A, Jiang H, Li X, Akhter K, Zhang J, Hsu JT, Miller MI, van Zijl PCM, Albert M, Lyketsos CG, Woods R, Toga AW, Pike GB, Rosa-Neto P, Evans A, Mazziotta J, Mori S. Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: Application to normal elderly and Alzheimer's disease participants. *NeuroImage*. 2009; 46:486–499. [PubMed: 19385016]
- Pennec X. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal of Mathematical Imaging and Vision*. 2006; 25:127–154.
- Qiu A, Brown T, Fischl B, Ma J, Miller MI. Atlas Generation for Subcortical and Ventricular Structures With Its Applications in Shape Analysis. *Image Processing, IEEE Transactions on*. 2010; 19:1539–1547.
- Ratnanather JT, Barta PE, Honeycutt NA, Lee N, Morris HM, Dziorny AC, Hurdal MK, Pearlson GD, Miller MI. Dynamic programming generation of boundaries of local coordinatized submanifolds in the neocortex: application to the planum temporale. *NeuroImage*. 2003; 20:359–377. [PubMed: 14527596]
- Rohlfing T, Brandt R, Menzel R, Maurer CR Jr. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*. 2004; 21:1428–1442. [PubMed: 15050568]

- Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW. Mapping cortical change across the human life span. *Nature Neuroscience*. 2003; 6:309–315.
- Talairach, J.; Tournoux, P. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. Thieme; New York: 1988.
- Thompson PM, Hayashi KM, de Zubicaray GI, Janke AL, Rose SE, Semple J, Hong MS, Herman DH, Gravano D, Doddrell DM, Toga AW. Mapping hippocampal and ventricular change in Alzheimer disease. *NeuroImage*. 2004; 22:1754–1766. [PubMed: 15275931]
- Thompson PM, Woods RP, Mega MS, Toga AW. Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. *Human Brain Mapping*. 2000; 9:81–92. [PubMed: 10680765]
- Van Essen DC. A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *NeuroImage*. 2005; 28:635–662. [PubMed: 16172003]
- Wang, H.; Jung Wook, S.; Das, S.; Pluta, J.; Altinay, M.; Yushkevich, P. Regression-based label fusion for multi-atlas segmentation. *Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on; 2011. p. 1113–1120.
- Wu G, Jia H, Wang Q, Shen D. SharpMean: Groupwise registration guided by sharp mean image and tree-based registration. *NeuroImage*. 2011; 56:1968–1981. [PubMed: 21440646]
- Yeo BTT, Sabuncu MR, Desikan R, Fischl B, Golland P. Effects of registration regularization and atlas sharpness on segmentation accuracy. *Medical Image Analysis*. 2008; 12:603–615. [PubMed: 18667352]
- Younes L. Computable Elastic Distances between Shapes. *SIAM Journal on Applied Mathematics*. 1998; 58:565–586.
- Zhang, Y.; Zhang, J.; Ma, J.; Oishi, K.; Faria, AV.; Miller, MI.; Mori, S. *Proc Intl Soc Mag Reson Med*. Vol. 19. Montréal, Canada: 2011. Creation of a population-representative brain atlas with clear anatomical definition; p. 135
- Zhang Y, Zhang J, Oishi K, Faria AV, Jiang H, Li X, Akhter K, Rosa-Neto P, Pike GB, Evans A, Toga AW, Woods R, Mazziotta JC, Miller MI, van Zijl PCM, Mori S. Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *NeuroImage*. 2010; 52:1289–1301. [PubMed: 20570617]
- Zhong J, Phua DY, Qiu A. Quantitative evaluation of LDDMM, FreeSurfer, and CARET for cortical surface mapping. *NeuroImage*. 2010; 52:131–141. [PubMed: 20381626]

Highlights

- We introduced an atlas generation method on whole-brain MR images.
- The dependence of initial values for this atlas generation method is examined
- The atlas is validated to be group representative and improve registration accuracy
- This atlas can improve image registration accuracy on sub-cortical brain structures
- This atlas enables a volume-surface framework for image analysis

Figure 1.

A schematic illustration of the VTE-based, group-specific atlas generation method. A single subject (preferably with a pre-defined parcellation map) serves as the initial template. In each step, the current template is mapped to each subject image to update the group mean diffeomorphism, and obtain the ‘centered’ template gradually. At the end, the averaged metric distances are optimized. The parcellation map of the initial template is carried into the VTE atlas coordinates through the transformation over the entire process.

Figure 2.

Comparison of a single-subject (SS) atlas, group-averaged atlases (AGA, NGA), and the VTE atlas for the aged control subjects ($n=12$). Axial and coronal views of the images are shown in (A-D). The SS was randomly chosen from the dataset. The AGA (affine group average) and the NGA (nonlinear group average) were the group average after affine normalization and after nonlinear normalization, respectively. The VTE atlas preserves the same image sharpness as the original single-subject image. The structural shape in the VTE was close to the shape shown in the AGA, which represents a ‘center shape’ of the dataset in a smoothed image. The SS atlas was biased from this center shape. E) shows the 3D visualization of the segmented lateral ventricles of each subject ($n=12$), the group average outline (GA), and the VTE. The lateral ventricle of VTE is positioned at the center of the group-averaged outline.

Figure 3.

VTE-based, population-specific T1 atlases for three groups of data with different ages and pathological conditions. A) shows the VTE from a healthy young adult population, 28.6 ± 4.6 years of age; B) shows the VTE from a normal aging population, 75.0 ± 5.9 years of age; and C) shows the VTE from an Alzheimer's patient population, 75.0 ± 4.2 years of age. The population-specific atlases reflect the unique anatomical features (e.g., lateral ventricles) of each group with sharp image contrast.

Figure 4.

The VTE results from different initial templates for aged controls (A) and AD patients (B).

A: The aged control subjects (top row) and the VTE atlases, with each subject serving as the initial template (bottom row), are shown in columns 1-12. The intensity average and standard deviation images of both the subjects and the VTE images are shown in columns 13-14. B: AD subjects (top row) and their corresponding VTE results (bottom row). In each population, regardless of the large variation in the morphology of each subject's image, all the VTE results have very similar shapes, except for small differences in the cortical areas. The average of the VTEs is less smeared than the average of the dataset. The maximum standard deviation of the VTE results is within one-third of the standard deviation of the population.

Figure 5.

The pre- and post-central gyri sub-surfaces of the VTE atlases generated from different initial templates. A1-A6) show the spatial alignment of gyri sub-surfaces from 12 VTE results, with one VTE (red) as a reference. B1-B2) show the surface-to-surface distances (SSD) from the reference VTE to the rest of the VTEs. The median value of each subject's SSD profile was calculated, and the mean of this median SSD across subjects was 1.15 ± 0.30 mm for the pre-central gyri and 0.97 ± 0.16 mm for the post-central gyri. The histogram peaks for the pre- and post-central gyri were located at 0.58 mm and 0.51 mm, respectively. Fig.5C displays the spatial distribution of the SSD on the sub-surfaces. The color encodes the distance value, with a range of 0-5 mm.

Figure 6.

The volume comparison of sub-cortical structures for the normal-age population (A), and for the AD population (B). The three columns of bar plots in each structure denote: 1) the dataset ($n=12$, white bar); 2) VTE atlases generated by each original image as the initial template ($n=12$, black bars); and 3) an SS atlas randomly chosen from the dataset (bar with dots). In both populations, the mean volume of VTE atlases (black bars) approximates the group means (white bar), with a much-reduced spatial variance than that of the dataset. For all structures, statistical tests confirmed the hypothesis that the two means had no significant differences ($p > 0.05$). The volume of a random single subject (bar with dots) appears heavily biased from the group means (white bar).

Figure 7.

Metric distances between each template and the subjects. A-C) show individual metric distances from a template to the aged control dataset. The length of each line segment was proportional to its corresponding metric distance value. The templates were SS#1, SS#2, and VTE for (A, B, and C), respectively. The number in black surrounding the rings represents the subject indices, and the number (in cyan) on each ring shows the scale of the metric distance value. D) shows the mean and standard deviation of metric distances measured from A-C. T-tests showed significant differences between the VTE and each of the SS atlases.

Figure 8.

Voxel-wise comparison of the magnitude of the averaged displacement vector (ADV) and the log-Jacobian in the normal aging population. In A), the warm color represents a higher value of averaged displacement magnitude on the shown location. The spatial pattern of the average displacement varies, depending on the choice of a template. The averaged displacement for the VTE atlas showed very small values (closer to zero) in most regions, and the values were on a scale of 2 mm for cortical areas. In B), the log-Jacobian encodes volumetric expansion ($\log-J > 0$) or shrinkage ($\log-J < 0$) on the given voxel. A large log-Jacobian absolute value was observed at the frontal horn of the lateral ventricles in columns 1 and 3. The log-Jacobian value for the VTE as a template was close to “0.”

Figure 9.

The kappa values of 24 sub-cortical brain structures in the image registration of AD subjects ($n=12$) using VTE (black triangle markers), the Eve atlas (pink cross markers), and each AD subject as the template. Each different colored curve with markers represents a specific template image that was used to register the group data; the kappa value, averaged across the group, is shown. The Eve atlas obtained the lowest kappa value for registration in the following structures: right caudate; left caudate; right hippocampus; left hippocampus; 3rd ventricle; lateral ventricle; and the inferior horn of the lateral ventricles. The kappa values from the VTE registration were higher than the average performance of 12 single-subject templates for each structure. In most structures, VTE achieved better registration than a randomly chosen SS template.

Figure 10.

The VTE-based registrations on the pre- (left panel) and post-central (right panel) gyri. In each panel, A-B) the stereotaxic views of the gyri surfaces for original subjects (A) and registered subjects by VTE (B) are shown. Red curves show the reference gyri surfaces; curves in cyan show each individual surface. The cortical surfaces were generated using FreeSurfer regular routines. The original subjects were in ICBM-152 coordinates. C) shows the SSD from the VTE gyri to each original subject (blue curves) and each registered subject (pink curves). D) shows a comparison of SSD before and after registration. E-F) shows the SSD distribution of spatial alignment before and after registration.

Figure 11.

Three sub-types of the lateral ventricle tail and its effect on registration. The posterior horn of the lateral ventricle in the VTE was segmented as two isolated regions (Type III), and the three subjects with different subtypes of topology in the lateral ventricle tail were automatically segmented by the VTE atlas (bottom row), while their manual delineations were obtained. In Type I and Type II subjects, the automated segmentation showed incorrect topological features with automated segmentation. Despite this, VTE provided the highest registration performance (kappa) compared to the other SS templates for this structure (kappa values compared in Fig.9, red arrow).

Table 1
List of manually segmented sub-cortical structures

Abbreviations	Full names of the structure
Put_R	Right putamen
Put_L	Left putamen
Caud_R	Right caudate
Caud_L	Left caudate
GP_R	Right globus pallidus
GP_L	Left globus pallidus
Thal_R	Right thalamus
Thal_L	Left thalamus
Amyg_R	Right amygdala
Amyg_L	Left amygdala
Hippo_R	Right hippocampus
Hippo_L	Left hippocampus
LV_frontal_R	Right frontal part of lateral ventricle
LV_frontal_L	Left frontal part of lateral ventricle
LV_body_R	Right body part of lateral ventricle
LV_body_L	Left body part of lateral ventricle
LV_atri_R	Right atrium of lateral ventricle
LV_atri_L	Left atrium of lateral ventricle
LV_post_R	Right posterior part of lateral ventricle
LV_post_L	Left posterior part of lateral ventricle
LVinfR	Right inferior part of lateral ventricle
LV_inf_L	Left inferior part of lateral ventricle
3 rd V	The third ventricle
4 th V	The fourth ventricle
Lateral V	The lateral ventricle

Total brain volumes of the atlases (mm³). The atlases were generated based on the normal-aging dataset (N=12), with images affine-normalized to the ICBM-152 coordinates. After normalization, the volumes of the atlases approximated the mean brain volume of the dataset. The brain volumes in the AGA and the NGA that were larger than the group mean were probably attributable to the “virtual convolution”, which somewhat enlarges the template compared with most individual brains, as noted for MNI305 T1.

Table 2

Images	Single Subject	AGA	NGA	VTE	Dataset
Volumes (mm ³)	2032512	2117092	2106730	1983060	1973250 ± 66094