Effects of Registration Regularization and Atlas Sharpness on Segmentation Accuracy

B.T. Thomas Yeo[a,1], **Mert R. Sabuncu**[a,1], **Rahul Desikan**[b], **Bruce Fischl**[a,c,d,2], and **Polina Golland**[a,2]

a*Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA*

b*Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, USA*

c*Department of Radiology, Harvard Medical School, Charlestown, MA, USA*

d*Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA*

Abstract

In non-rigid registration, the tradeoff between warp regularization and image fidelity is typically determined empirically. In atlas-based segmentation, this leads to a probabilistic atlas of arbitrary sharpness: weak regularization results in well-aligned training images and a sharp atlas; strong regularization yields a "blurry" atlas.

In this paper, we employ a generative model for the joint registration and segmentation of images. The atlas construction process arises naturally as estimation of the model parameters. This framework allows the computation of unbiased atlases from manually labeled data at various degrees of "sharpness", as well as the joint registration and segmentation of a novel brain in a *consistent* manner.

We study the effects of the tradeoff of atlas sharpness and warp smoothness in the context of cortical surface parcellation. This is an important question because of the increasingly availability of atlases in public databases and the development of registration algorithms separate from the atlas construction process. We find that the optimal segmentation (parcellation) corresponds to a unique balance of atlas sharpness and warp regularization, yielding statistically significant improvements over the FreeSurfer parcellation algorithm. Furthermore, we conclude that one can simply use a single atlas computed at an optimal sharpness for the registration-segmentation of a new subject with a pre-determined, fixed, optimal warp constraint. The optimal atlas sharpness and warp smoothness can be determined by probing the segmentation performance on available training data. Our experiments also suggest that segmentation accuracy is *tolerant* up to a small mismatch between atlas sharpness and warp smoothness.

Correspondence to: B.T. Thomas Yeo.

Email addresses: ythomas@csail.mit.edu (B.T. Thomas Yeo), msabuncu@csail.mit.edu (Mert R. Sabuncu), rahul@nmr.mgh.harvard.edu (Rahul Desikan), fischl@nmr.mgh.harvard.edu (Bruce Fischl), polina@csail.mit.edu (Polina Golland).
[1]Thomas Yeo and Mert Sabuncu contributed equally to this work.
[2]Bruce Fischl and Polina Golland contributed equally to this work.

Keywords

Generative Model; Registration; Segmentation; Parcellation; Multiple Atlases; Markov Random Field; Regularization

1 Introduction

In this work, we propose a generative model for the joint registration and parcellation of cortical surfaces. We formulate the atlas construction process as estimation of the generative model parameters. This provides a consistent framework for constructing the atlas as well as the registration and segmentation of a novel subject. We explore the effects of registration regularization on the atlas construction and the new image segmentation. We conclude that optimal segmentation corresponds to a unique balance of atlas sharpness and warp regularization. However, the segmentation accuracy is *tolerant* up to a small mismatch between atlas sharpness and warp smoothness.

1.1 Probabilistic Atlases

Probabilistic atlases are powerful tools in segmentation [3,7,11,16,17,35,44]. The simplest probabilistic segmentation atlas provides only the prior probability of labels at a spatial position and no information about the expected appearance of the image. One uses this type of atlas to initialize and to guide the segmentation of a new image [44]. More sophisticated algorithms deform the atlas to the new image and transfer label probabilities from the atlas to the image [3,21]. These label probabilities are then combined with the inten-sity of the new image to produce the final segmentation. We note that the phrase "image intensities" is used in a generic sense to indicate all image-derived features, such as MR intensity or cortical geometry. In this approach, the relationship between labels and image intensities is estimated in the new image and not from the training images used to compute the atlas.

More complex probabilistic atlases provide statistics on the relationship between the segmentation labels and the image intensities [11,12,16,17,35]. The intensity model in the atlas is used to bring the atlas and new image into the same space. The label probabilities and intensity model are then used to segment the new image. Such a registration process can be done sequentially [11,16,17] or jointly [35]. Since this complex atlas contains more information than the simpler atlas, it can conceivably be a more powerful tool for segmentation. However, extra care is needed if the modalities of the new image and training images are different.

In this work, we focus on probabilistic segmentation atlases that model both labels and intensities. An initial step in probabilistic atlas computation is the spatial normalization of the training images. The features used for co-registering images are usually derived from the images themselves [1,4,13,20,23,25,34,43] or from the segmentation labels [10,29,45,46,51]. After normalization to a common space, we can compute the variability in intensity values (and/or labels) in the atlas space. Our approach uses both labels and intensities to co-register the training images. The need to utilize both labels and intensities arises naturally from the proposed generative model.

Joint registration-segmentation algorithms are generally more effective than sequential registration-segmentation as registration and segmentation benefit from additional knowledge of each other [3,35,47–50]. In our case, the requirement to jointly register and segment is also a natural consequence of the generative model.

1.2 Warp Regularization

Spatial normalization of the training images can be achieved with different registration algorithms that vary in the flexibility of warps they allow. Both low-dimensional warps, e.g., affine [35], and richer, more flexible warps can be employed, e.g., splines [4,43,46], dense displacement fields [10,20,13,17,45] and velocity fields [23,29]. More restricted warps yield blurrier atlases that capture inter-subject variability of structures, enlarging the basin of attraction for the registration of a new subject. However, label prediction accuracy is limited by the sharpness of the atlas. Recently, it has been shown that combining information on the warp flexibility and the residual image of the registration process improve the classification accuracy of schizophrenic and normal subjects [30]. Finding the optimal warp regularization tradeoff has received attention in recent years. Twining *et al*. [43] and Van Leemput [45] propose frameworks to find the least complex models that explain the image intensity and segmentation labels of the training images.

In contrast, we explicitly explore the relationship between the warp flexibility and the atlas sharpness in the context of atlas-based segmentation. Rather than finding one optimal warp regularization when computing the atlas, we compute atlases with varying degrees of warp regularization and thus obtain atlases of varying degrees of sharpness. We use these to study the effect of atlas sharpness and warp regularization on segmentation accuracy. In particular, we compare three specific approaches: (1) progressive (i.e., using increasingly flexible warps) registration-segmentation of a new brain with increasingly sharp atlases; (2) progressive registration with an atlas of a particular sharpness; (3) registration with an atlas of a particular sharpness using a pre-determined, fixed constraint on the warp.

1.3 Cortical Parcellation

We investigate the question of regularization in the context of automatic parcellation of cortical surfaces. Automatic labeling of surface models of the cerebral cortex is important for identifying regions of interest for clinical, functional and structural studies [11,39]. In particular, it has been shown that cortical surface registration can increase the power of functional alignment or activation consistency [14], as well as the alignment of cytoarchitectonics [18]. Recent efforts range from identification of sulcal or gyral ridge lines [41,42] to segmentation of sulcal or gyral basins [11,17,27,28,31,37,39]. Similar to these prior studies, we are interested in the parcellation of the entire cortical surface, i.e., the automated labeling of each point of the surface of a new subject.

Because the local geometries of different sulci and gyri might be similar, learning the geometries of the different sulcal and gyral labels using information from a new image only is difficult. Instead, we will follow the approach of [17,11] and utilize an atlas that encodes both the spatial distributions of the labels and their spatially varying geometries.

In the context of cortical parcellation, our experiments suggest that the optimal parcellation in all three registration schemes for exploring atlas sharpness and warp regularization corresponds to a unique balance of atlas sharpness and warp regularization. Furthermore, the optimal parameters yield statistically significant improvements over the FreeSurfer parcellation algorithm [17]. There is no difference in the optimal parcellation accuracy achieved by the three schemes.

While one might expect that outlier images require weaker regularization to warp closer to the population "average" to achieve better segmentation accuracy, our experiments show that it is sufficient to use a single atlas computed at an optimal sharpness for the registration-segmentation of a new subject with a pre-determined, fixed, optimal warp constraint.

The optimal atlas sharpness and warp smoothness can be determined by probing the segmentation performance on available training data. We find that the optimal warp smoothness for the new subject should be the same as the warp smoothness used for creating the atlas. While this might be an obvious consequence of our explicit generative model, in practice, it is common to use or develop a registration algorithm separate from atlas construction. This is especially true with the increasing availability of atlases in publicly available databases, such as MNI305 [12]. It is therefore important to determine whether using different warp smoothness and atlas sharpness is necessarily detrimental to segmentation in practice. Our experiments suggest that segmentation accuracy is tolerant up to a small mismatch between atlas sharpness and warp smoothness.

A preliminary version of this work was published at the International Conference on Medical Image Computing and Computer Assisted Intervention [52]. This article expands the conference paper with a more detailed theoretical development and more extensive experimental work. In particular, the theory for the atlas construction process was only briefly discussed in [52], but is covered in depth in this paper.

Our contributions are as follows:

1. We propose a generative model for the joint registration and segmentation of images. The atlas construction process is formulated as a parameter estimation problem. This provides a consistent framework for both estimating the atlas, as well as for the registration and segmentation of a new image.

2. We explore the space of atlas sharpness and warp regularization for registering and parcellating cortical surfaces. We find that the optimal parcellation corresponds to a unique balance of atlas sharpness and warp regularization. The robustness of the optimal atlas sharpness and warp regularization across trials suggests that we can use the same parameters for future cortical surface parcellation.

3. We show improved parcellation results over the state-of-the-art FreeSurfer parcellation algorithm [17].

In the next section, we introduce the generative model, describe the resulting atlas estimation process, and the registration and segmentation of a novel image. Section 3 introduces the cortical surface parcellation problem and describes further modeling assumptions made in the context of this problem. We present experimental results in Section 4.

2 Theory

Given a training set of N images $I_{1:N} = \{I_1, \ldots, I_N\}$ with label maps $L_{1:N} = \{L_1, \ldots, L_N\}$, joint registration-segmentation aims to infer the registration parameters R and segmentation labels L of a new image I. To achieve this goal, we first learn the parameters of the generative model (Figure 1) from the training images. These parameters correspond to the atlas A and smoothness parameter S. Here, we assume the generative model for the warp R is parameterized by S. According to the generative model, the parameters R and A are independent. However, conditioned on the observed image, R and A become dependent. Estimating the parameters S and A involves co-registration of the training images into a common space. We emphasize that our co-registration process uses both the labels and image intensities of the training images, as described in the next section.

2.1 Generative Model for Registration and Segmentation

We consider the generative model of Figure 1. L' is a label map in the atlas space generated probabilistically by the atlas A. For example, L' could be the tissue type at each MRI pixel, generated by a tissue probability map. Given the label map, the atlas then generates image I'.

For example, at each pixel, we can generate an MR intensity conditioned on the tissue type and spatial location. Finally, we generate a random warp field R controlled by the smoothness parameter S. For example, we can generate a random displacement field, with neighboring displacements encouraged to be "close". For example, S can be the spacing of spline control points or the penalty in a cost function that discourages large or discontinuous deformation fields. The random warp R is then applied to the label map L' and image I' to create the observed label map L and observed image I, i.e., $I(R(x)) = I'(x)$ and $L(R(x)) = L'(x)$. Thus a location x in the atlas space is mapped to a location $R(x)$ in the native (or raw) image space. We defer a detailed instantiation of the model to Section 3.

During co-registration, a small value of smoothness parameter S leads to less constrained warps, resulting in better alignment of the training images[3]. This results in a sharper atlas. On the other hand, a larger smoothness parameter yields more regularized warps and a blurrier atlas.

2.2 Atlas Building: Estimating Parameters of Generative Model

To estimate the parameters of the generative model, we maximize the likely-hood of the observed images $I_{1:N}$ and $L_{1:N}$ over the values of the non-random smoothness parameter S and atlas A.

$$(S^*,A^*)=\underset{S,A}{\operatorname{argmax}} \log p(I_{1:N},L_{1:N};S,A) \tag{1}$$

$$=\underset{S,A}{\operatorname{argmax}} \log \int p(I_{1:N},L_{1:N},R_{1:N};S,A)d R_{1:N}. \tag{2}$$

Here $p(a; b)$ indicates the probability of random variable a parameterized by a non-random parameter b while $p(a/b)$ indicates the probability of random variable a conditioned on a random variable b.

In this case, we need to marginalize over the registration warps $R_{1:N}$, which is difficult. Previously demonstrated methods for atlas construction use various approximations to evaluate the integral. In this paper, we adopt the standard approximation that replaces the integral with the value of the integrand estimated at the maximum likelihood estimate of deformation. In contrast, [38] uses a sampling method while [45] uses the Laplace approximation, which essentially models the distribution to be a Gaussian centered at the maximum likelihood estimate. It is unclear that these more complex methods, while theoretically more sound, lead to (practically) better approximation. Based on this approximation, we seek

$$(S^*,A^*,R_{1:N}^*)=\underset{S,A,R_{1:N}}{\operatorname{argmax}} \log p(I_{1:N},L_{1:N},R_{1:N};S,A) \tag{3}$$

$$=\underset{S,A,R_{1:N}}{\operatorname{argmax}} \sum_{n=1}^{N} \log p(I_n,L_n,R_n;S,A) \tag{4}$$

$$=\underset{S,A,R_{1:N}}{\operatorname{argmax}} \sum_{n=1}^{N} \left[\log p(R_n;S)+\log p(I_n,L_n|R_n;A) \right]. \tag{5}$$

---

[3]By a "better alignment of images", we mean that the warped images look more similar, i.e., the similarity measure is improved. However, an improved similarity measure does not necessarily imply deformations with better label alignment. In fact, we show in the paper that the best segmentation occurs when warps are not overly flexible.

The second equality comes from the fact that the training images are independent of each other given the atlas A and smoothness parameter S. The last equality is obtained by applying Bayes' rule and the independence relations specified by the graphical model in Figure 1.

Optimizing the above expression yields the atlas A^* and smoothness parameter S^*. As mentioned before, a smaller value of the smoothness parameter S results in a sharper atlas. Since we are interested in how atlas sharpness affects segmentation accuracy, instead of estimating one single optimal S^*, we construct a series of atlases corresponding to different values of the regularization parameter S. In particular, we discretize S into a finite set $\{S_k\} = \{S_1 > S_2 > \ldots > S_K\}$. For each value S_k, we seek the optimal atlas and set of deformations:

$$(A^*, R^*_{1:N}) = \underset{A, R_{1:N}}{\arg\max} \sum_{n=1}^{N} \left[\log p(R_n; S_k) + \log p(I_n, L_n | R_n; A) \right]. \tag{6}$$

We shall refer to the atlas computed using a particular S_k as $A_{\alpha=S_k}$. We use alternating optimization to maximize Eq. (6). In each step, we fix the set of registration warps $R_{1:N}$ and estimate the atlas $A_{\alpha=S_k}$:

$$A^*_{S_k} = \underset{A_{S_k}}{\arg\max} \sum_{n=1}^{N} \log p(I_n, L_n | R_n; A_{S_k}). \tag{7}$$

We then fix the atlas $A_{\alpha=S_k}$, and optimize the registration warps $R_{1:N}$ by optimizing each warp independently of others:

$$R^*_n = \underset{R_n}{\arg\max} \log p(R_n; S_k) + \log p(I_n, L_n | R_n; A_{S_k}), \forall n = 1, \ldots, N. \tag{8}$$

This process is repeated until convergence. Convergence is guaranteed since this is essentially a coordinate-ascent procedure operating on a bounded function.

We can think of Eq. (8) as the atlas co-registration objective function by treating $\log p(I_n, L_n | R_n; A_{S_k})$ as the data fidelity function and $\log p(R_n; S_k)$ as the regularization term. We effectively register each image independently to the atlas $A_{\alpha=S_k}$. This iterated process of updating the atlas is similar to [23], except we include training labels in the co-registration. We will show a concrete instantiation of this formulation in Section 3.

In practice, we first create atlas A_∞ based on a simple rigid-body registration and use it to initialize the atlas A_{S_1}, where S_1 is large enough such that the resultant warp is almost rigid. We then use atlas A_{S_1} to initialize the atlas A_{S_2} where $S_1 > S_2$, and so on. The result is a set of atlases $\{A_\alpha\} = \{A_{S_1} \ldots A_{S_K}\}$. With enough samples, the finite set $\{S_k, A_{\alpha=S_k}\}$ accurately represents the underlying continuous space of atlases at different levels of sharpness.

2.3 Registration and Segmentation of a New Image

Given an atlas A and smoothness parameter S, the registration and segmentation of a new image I can be computed using a Maximum-A-Posteriori (MAP) estimate. This involves finding the mode of the posterior distribution of the new image labels L and registration R given the observed image I and atlas A and smoothness parameter S:

$$(L^*, R^*) = \underset{L, R}{\arg\max} \log p(L, R | I; S, A) \tag{9}$$

$$= \underset{L, R}{\arg\max} \log p(I, L, R; S, A) \tag{10}$$

$$= \underset{L, R}{\arg\max} \log p(R; S) + \log p(I, L | R; A). \tag{11}$$

The second equality follows from the definition of the conditional probability. The last equality follows from the independence relations specified by the graphical model in Figure 1. In prior work, this MAP approach is favored by some [47–49], while others suggest estimating the MAP solution for the registration warp alone [3,35]:

$$R^* = \underset{R}{\text{argmax}} \; p(I,R;S,A) \tag{12}$$

$$= \underset{R}{\text{argmax}} \; \log \sum_L p(L,I,R;S,A), \tag{13}$$

and recovering the segmentation labels as a final step after recovering the optimal registration R^*. Prior work in joint registration and segmentation did not consider atlas construction in the same framework [3,35,47–49]. Furthermore, S is usually set by an expert rather than estimated from the data.

We previously reported results based on the latter two-step approach [52]. In this version, we use the former MAP framework since it requires fewer assumptions for practical optimization. As we show in Section 3, optimizing Eq. (11) using a soft-MAP coordinate ascent approach using the Mean Field approximation [22,26] results in the same update rule used by our previously demonstrated method [52].

The differences between the new image registration specified by Eq. (11) and the atlas co-registration objective function in Eq. (8) comes from the unavoidable fact that for the new image, the label map is not observed. To optimize Eq. (11), we use a coordinate ascent scheme. In step t, we fix the registration parameters $R^{(t)}$ and estimate the labels $L^{(t+1)}$:

$$L^{(t+1)} = \underset{L}{\text{argmax}} \; \log p(R^{(t)};S) + \log p(I,L|R^{(t)};A) \tag{14}$$

$$= \underset{L}{\text{argmax}} \; \log p(I,L|R^{(t)};A) \tag{15}$$

$$= \underset{L}{\text{argmax}} \; \log p(L|I,R^{(t)};A). \tag{16}$$

Eq. (16) optimizes the log posterior probability of the label map L given the image I, atlas A and current estimate of the registration parameters $R^{(t)}$. Next, we fix the label estimate $L^{(t+1)}$ and re-estimate the registration parameters $R^{(t+1)}$:

$$R^{(t+1)} = \underset{R}{\text{argmax}} \; \log p(R;S) + \log p(I,L^{(t+1)}|R;A). \tag{17}$$

We can think of Eq. (17) as the new image registration objective function by treating $\log p(I, L^{(t+1)}|R;A)$ as the data fidelity term and $\log p(R; S)$ as the regularization term.

To maintain generality, we allow the use of a smoothness parameter S_k for the registration and segmentation of a new subject with an atlas A_α where S_k may not be equal to α. In other words, we can, for instance, compute an atlas using affine transformations, while using a flexible deformation model for the registration of a new subject. Strictly speaking, S_k should theoretically be equal to α in Eq. (11) from the perspective of probabilistic inference. Here, we examine whether using different warp smoothness and atlas sharpness is necessarily detrimental to segmentation in practice.

More specifically, we investigate three strategies for exploring the space of atlas sharpness and warp smoothness to register and segment a new image as illustrated in Figure 2.

1. Multiple Atlas, Multiple Smoothness (MAMS): A multiscale approach where we optimize Eq. (16) and Eq. (17) w.r.t. R and L with the blurry atlas A_{S_1} and warp regularization S_1. The resulting alignment is used to initialize the registration with a sharper atlas A_{S_2} and a correspondingly flexible warp regularization S_2, and so on (Figure 2a).

2. Single Atlas, Multiple Smoothness (SAMS): We use a fixed atlas of a fixed sharpness scale A_{S_k} to compute R and L according to Eq. (16) and Eq. (17) using a progressively decreasing warp smoothness S (Figure 2b).

3. Single Atlas, Single Smoothness (SASS):We optimize Eq. (16) and Eq. (17) w.r.t. R and L with a fixed atlas A_{S_k} and warp regularization S_m, where S_k might not be equal to S_m (Figure 2c).

So far, the derivations have been general without any assumption about the atlases A_α, the prior $p(R; S_k)$ or the image-segmentation fidelity $p(I, L|R;A_\alpha)$. In the next section, we instantiate this approach for a concrete example of cortical surface registration and parcellation.

3 Cortical Surface Parcellation

We now demonstrate the framework developed in the previous section for the joint registration and parcellation of surface models of the cerebral cortex. These surfaces are represented by triangular meshes with a spherical coordinate system that minimizes metric distortion [9,13]. The aim is to register an unlabeled cortical surface to a set of manually labeled surfaces and to classify each vertex of the triangular mesh in terms of the anatomical units. To construct the generative model for this problem, we need to define the prior on the registration parameters $p(R; S)$ and the model for label and image generation $p(I, L|R;A_\alpha)$. In general, our model decisions were inspired by previous work on cortical surface parcellation [13,15,17].

3.1 Generative Model for Registration and Segmentation

We model the warp regularization with an MRF parameterized by S:

$$p(R;S) \triangleq \frac{F(R)}{Z_1(S)} \exp\left\{ -S\left[\sum_i \sum_{j \in N_i}\left(\frac{d_{ij}^R - d_{ij}^0}{d_{ij}^0}\right)^2 \right]\right\}$$

(18)

where d_{ij}^R is the distance between vertices i and j under registration R, d_{ij}^0 is the original distance, N_i is a neighborhood of vertex i and $Z_1(S)$ is the partition function. Our regularization penalizes percentage metric distortion weighted by a scalar S that reflects the amount of smoothness (rigidity) of the final warp. We choose a percentage metric distortion instead of an absolute metric distortion [13] to ensure that the tradeoff between the regularization and the data fidelity term is invariant to scale in our multi-resolution framework. This is especially important in our work since we explore a large range of warp smoothness parameter S. $F(\cdot)$ ensures invertibility. It is zero if any triangle is folded by warp R and one otherwise. We represent the warp R as a displacement field on the sphere. Therefore, a term like $F(R)$ is necessary. One could replace $F(R)$ with a more sophisticated invertibility prior (e.g., [2,33]) or restrict the space of feasible warps to be the space of diffeomorphisms [23,19].

We first decompose the label and image prior

$$p(I,L|R;A_\alpha)=p(L|R;A_\alpha)p(I|L,R;A_\alpha).$$

(19)

and impose an MRF prior on the parcellation labels

$$p(L|R;A_\alpha) \triangleq \frac{1}{Z_2(A_\alpha)}\exp\left\{\sum_i U_i(A_\alpha)L(R(x_i))+\right.$$

$$\left.+\sum_i \sum_{j\in N_i} L^T(R(x_i))V_{ij}(A_\alpha)L(R(x_j))\right\}$$

(20)

Here, we use the vectorized MRF notation of [26]. Assuming the total number of labels is *M*, $L(R(x_i))$ is a column vector of size $M \times 1$ that sums to 1. Each component of $L(R(x_i))$ is an indicator variable. In particular, suppose the image has label *m* at location $R(x_i)$, then $L(R(x_i))$ is zero for all entries, except for the *m*-th entry, which is equal to 1. It is a common practice to relax the constraint so that *L* still sums to 1 but the entries might take on fractional values to indicate uncertainty in the segmentation results. $U_i(A_\alpha)$ is a 1x*M* local potential vector that captures the frequency of label $L(R(x_i))$ at vertex *i* of the atlas mesh. The *M*x*M* compatibility matrix $V_{ij}(A_\alpha)$ reflects the likelihood of two labels at vertices *i* and *j* being neighbors. $Z_2(A_\alpha)$ is the partition function dependent on the atlas A_α, ensuring the given MRF is a valid probability distribution.

We further assume that the noise added to the mean image intensity at each vertex location is independent, given the label at that location.

$$p(I|L,R;A_\alpha) \triangleq \prod_i W_i(I(R(x_i));A_\alpha)L(R(x_i)),$$

(21)

where $W_i(I(R(x_i));A_\alpha)$ is a 1x*M* observation potential vector defined at each atlas vertex *i*. The *m*-th entry corresponds to the likelihood of observing a particular image intensity or vectors of intensity (in this case, the local surface geometries) at location $R(x_i)$ given a particular label *m*. We assume W_i follows a Gaussian distribution, e.g., given that the post-central gyrus is at location $R(x_i)$ of the image, we expect the local curvature and/or sulcal depth $I(R(x_i))$ to follow a Gaussian distribution whose parameters are estimated from the training images.

The collection of model components $\{U_i, V_{ij}, W_i\}$ define the atlas A_α. Eq. (20) defines an isotropic prior on the labels, which is a simpler model than that used by modern approaches. The FreeSurfer parcellation algorithm uses a spatially varying and anisotropic MRF model [17], whose parameters change dynamically with the geometries of the subject being segmented. An anisotropic MRF improves the parcellation accuracy because the boundaries of certain gyral regions are predicted robustly by the variation in curvature. For example, the boundary between the pre-central and post-central gyrus is the central sulcus. Along the boundary, there is high curvature, while across the boundary, the curvature drops off sharply.

We made the explicit choice of warping an image (interpolating an image) to the atlas space. The alternative of warping the atlas (interpolating the atlas) to image space would require us to interpolate the MRF, which is non-trivial. Interpolating *U* would result in a change in the partition function, which is exponentially hard to compute. In addition, if we were to use the dynamic model of FreeSurfer, since we have made the choice of warping the subject, this would mean that the model parameters and hence the partition function of the MRF would change during the registration step. FreeSurfer does not have this problem because it does not perform joint registration and segmentation. Therefore, we assume *V* to be spatially stationary and isotropic and drop the subscripts *i, j*.

3.2 Atlas Building: Estimating Parameters of Generative Model

Substituting Eq. (18), Eq. (20) and Eq. (21) into the atlas co-registration objective function in Eq. (8), we obtain:

$$
\begin{aligned}
R_n^* \;=\;& \underset{R_n}{\mathrm{argmax}}\; \log p(R_n;S_k) + \log p(I_n, L_n \big| R_n; A_{S_k}) \\[4pt]
=\;& \underset{R_n}{\mathrm{argmax}}\; \log F(R_n) - S_k \sum_i \sum_{j \in N_i} \left(\frac{d_{ij}^{R_n} - d_{ij}^0}{d_{ij}^0} \right)^2 + \\[4pt]
& + \sum_i U_i(A_\alpha) L(R_n(x_i)) + \sum_i \sum_{j \in N_i} L^T(R_n(x_i)) V(A_\alpha) L(R_n(x_i)) + \\[4pt]
& + \sum_i W_i(I(R_n(x_i)); A_\alpha) L(R_n(x_i)) + \text{const.}
\end{aligned}
$$

(22)

where the first term prevents folding triangles, the second term penalizes metric distortion, the third and fourth terms are the Markov prior on the labels and the last term is the likelihood of the surface geometries given the segmentation labels of the *n*-th training image.

We can then fix R_n^* and estimate the atlas parameters $A_\alpha = \{U_i, V, W_i\}$ using Eq. (7). In practice, we use the naive approach of frequency counts [17] to estimate the clique potentials U, V and the maximum likelihood estimation of the Gaussian likelihood terms W. Appendix B provides the implementation details.

3.3 Registration and Segmentation of a New Image

Similarly, we substitute Eq. (18), Eq. (20) and Eq. (21) into the update rules for the new subject segmentation in Eq. (16) and registration in Eq. (17).

Warping the subject to the atlas, optimization in Eq. (16) with a fixed $R^{(t)}$ involves estimating the segmentation labels at positions $R^{(t)}(\{x_i\})$ of the subject, where $\{x_i\}$ are vertices of the atlas mesh. We will denote this segmentation estimate by $\hat{L}^{(t+1)}$ Eq. (16) becomes

$$
\hat{L}^{(t+1)} = \underset{\hat{L}}{\mathrm{argmax}}\; \log p(\hat{L} \big| I(R^{(t)}(\{x_i\}); A_\alpha)).
$$

(23)

Even with fixed $R^{(t)}$, solving the MAP segmentation Eq. (23) is NP-hard. We adopt the Mean Field approximation [22,26]. We then use the complete approximate distribution provided by the Mean Field solver in optimizing Eq. (17). This approximation effectively creates a soft segmentation estimate $\hat{L}_i^{(t+1)}$ at each location $R^{(t)}(x_i)$ of the new subject. $\hat{L}_i^{(t+1)}$ is a column vector of size $M \times 1$. The *m*-th component of $\hat{L}_i^{(t+1)}$ is the probability of finding label *m* at location $R^{(t)}(x_i)$ of the new subject. To estimate the label $\hat{L}(x)$ at an arbitrary location x in the subject space, we interpolate from $\hat{L}^{(t+1)}$ defined at $R^{(t)}(\{x_i\})$ onto location x.

This optimization procedure leads to fewer local minima since it avoids commitment to the initial estimates obtained through hard thresholding that might be very far from a good solution if the new image is originally poorly aligned to the atlas. Appendix A provides an outline for computing $\hat{L}^{(t+1)}$ via the Mean Field approximation. Warping the subject to the atlas, Eq. (17) becomes:

$$
\begin{aligned}
R^{(t+1)} \;=\;& \underset{R}{\mathrm{argmax}}\; \log F(R) - S_k \sum_i \sum_{j \in N_i} \left(\frac{d_{ij}^{R} - d_{ij}^0}{d_{ij}^0} \right)^2 + \\[4pt]
& + \sum_i U_i(A_\alpha) \hat{L}(R(x_i)) + \sum_i \sum_{j \in N_i} \hat{L}^T(R(x_i)) V(A_\alpha) \hat{L}(R(x_j)) + \\[4pt]
& + \sum_i W_i(I(R(x_i)); A_\alpha) \hat{L}(R(x_i)) + \text{const.}
\end{aligned}
$$

(24)

Further implementation details can be found in Appendix B.

4 Experiments and Discussion

We consider 39 surfaces that represent the gray-white matter interface of the left and right hemispheres automatically segmented from 3D MRI using FreeSurfer [9]. This data set exhibits significant anatomic variability since it contains young, middle-aged, elderly subjects and Alzheimer's patients. The surfaces are topologically corrected [15,40] and a spherical coordinate system is established by minimizing metric distortion [13]. For each hemisphere, the 39 cortical surfaces are first rigidly aligned on the sphere, which corresponds to rotation only. The surfaces are manually parcellated by a neuroanatomical expert into 35 labels [11]. Figure 3 illustrates the manual parcellation for one subject. A complete list of the parcellation units is included in Table 1.

We perform cross-validation by leaving out subjects 1 through 10 in the atlas construction, followed by the joint registration-segmentation of left-out subjects. We repeat with subjects 11 through 20, 21 through 30 and finally 31 through 39. We select S to be the set {100, 50, 25, 12.5, 8, 5, 2.5, 1, 0.5, 0.25, 0.1, 0.05, 0.01}. We find that in practice, $S = 100$ corresponds to allowing minimal metric distortion and $S = 0.01$ corresponds to allowing almost any distortion. The intervals in the set S are chosen so that each decrease in the value of S roughly corresponds to an average of 1mm increased displacement in registration.

Since we treat the subject mesh as the moving image, both registration and parcellation are performed on the fixed atlas mesh. The segmentation is interpolated from the atlas mesh onto the subject mesh to obtain the final segmentation. We compute segmentation quality by comparing this final segmentation with the "ground truth" manual parcellation.

To speed up the algorithm, we construct the atlas on a sub-divided icosahedron mesh with about 40k vertices. Typically, each subject mesh has more than 100k vertices. The segmentation labels inferred on the low resolution atlas mesh are therefore computed on a coarser grid than that of the manual parcellation. Yet, as we discuss in the remainder of this section, the proposed implementation on average outperforms the FreeSurfer parcellation algorithm [17].

Despite working with a lower resolution icosahedron mesh, registration at each smoothness scale still takes between 20 minutes to 1.5 hours per subject per atlas. Registration with weakly constrained warps ($S \leq 0.1$) requires more time because of the need to preserve the invertibility of the warps. The entire set of experiments took approximately 3 weeks to complete on a computing cluster, using up to 80 nodes in parallel.

4.1 Exploration of Smoothness S and Atlas A_α

In this section, we discuss experimental results for the exploration of the smoothness parameter S and atlases A_α. We measure the segmentation accuracy using the Dice coefficient, defined as the ratio of cortical surface area with correct labels to the total surface area, averaged over the test set.

Figure 4 shows the segmentation accuracy for SAMS ($A_\alpha = A_1$) and MAMS as we vary S. The average Dice peaks at approximately $S = 1$ for all cross-validation trials, although individual variation exists, as shown in Figure 5[4]. For a particular value of S, outliers warp more because the tradeoff between the data-fidelity and regularization is skewed towards the former. However, it is surprising to find that the optimal S is mostly constant across subjects. It also appears that peaks of the segmentation accuracy plots are relatively broad, implying that good parcellation results can be obtained for a range of S between 1 and 2.5.

---

[4]The optimal value of $S = 1$ is coincidental in the sense that it depends on the unit chosen for metric distortion.

For large S (highly constrained warps), MAMS consistently outperforms SAMS. Because the surfaces are initially misaligned, using a sharp atlas (in the case of SAMS, atlas A_1) results in poor segmentation accuracy due to a mismatch between the image features and the atlas statistics. As we decrease the smoothness S, SAMS allows for more flexible warps towards the population average than MAMS since it uses a sharper atlas. The similarity measure is therefore given higher weight to overcome the regularization. This results in better segmentation accuracy than MAMS. Eventually, SAMS and MAMS reach similar maximal values at the same optimal smoothness S. Beyond the optimal S, however, both MAMS and SAMS exhibit degraded performance. This is probably due to overfitting and local optima created by more flexible warps.

We also examine the Dice measure of each parcellation structure as a function of the warp smoothness S. In general, the curves are noisier but follow those of Figure 4. Figure 6(a) shows a typical curve that peaks at $S = 1$, while Figure 6(b) shows a curve that peaks at $S = 5$. However, in general, for most structures, the optimal smoothness S occurs at approximately $S = 1$ (Figure 7), which is not surprising since for most subjects, the optimal overall Dice (computed over the entire surface) also occurs at $S = 1$ (Figure 5).

We now explore the effects of both warp smoothness S and atlas sharpness α on parcellation accuracy. Figure 8 shows a plot of Dice averaged over all 39 subjects. The performances of MAMS, SAMS and SASS at ($S = 1$, $\alpha = 1$) are indistinguishable. As an example, we see that for both hemispheres, SAMS with $\alpha = 0.01$ (green line) starts off well, but eventually overfits with a worse peak at $S = 1$ ($p < 10^{-5}$ for one-sided paired-sampled t-test, statistically significant even when corrected for multiple comparisons). Similarly for SASS, the best values of α and S are both equal to 1. We also show SASS with $\alpha = 0.01$ and $S = 1$ in Figure 8. In general, there is no statistical difference between MAMS, SAMS or SASS at their optima: $\alpha = 1$, $S = 1$.

Originally, MAMS and SAMS were introduced to reduce local optima in methods, such as SASS. It is therefore surprising that the performance of all three methods is comparable. While using the correct smoothness and atlas sharpness is important, our "annealing" process of gradual reduction of smoothness (MAMS and SAMS) does not seem to increase the extent of the basins of attraction in the context of cortical parcellation. One possible reason is that on the cortical surfaces, two adjacent sulci might appear quite similar locally. Smoothing these features might not necessarily remove the local minima induced by such similarity. Incorporating multiscale features [32,53,54] with multiple smoothness offers a promising approach for avoiding such local optimal issues on the cortex.

The fact that the optimal smoothness parameter S^* corresponds to the optimal atlas sharpness parameter α^* is not surprising. According to the graphical model in Figure 1 and as mentioned in the derivations in Section 2.2 and Section 2.3, theoretically, we do expect $S^* = \alpha^*$. Learning this optimal S^* in the atlas construction process is a future avenue of investigation.

Alternatively, we can also determine the best S and A_α for a new image registration by optimizing the objective function in Eq. (11). Unfortunately, there are technical difficulties in doing this. First, we notice that the objective function in Eq. (11) increases with decreasing S. This model contains no Occam's razor regularization terms that penalize overfitting due to flexible warps. This is mainly because Eq. (11) omits certain difficult-to-compute normalization terms, such as the partition function that depends on U, V and W and thus dependent on S and α. These terms are ignored for fixed values of S and α. We can use various approximation strategies to compute the normalization terms. But it is not clear whether these approximations yield sufficient accuracy to determine the optimal values for S and α. On the other hand, empirically we find that $S = \alpha = 1$ consistently yields the optimal (or very close to the optimal) segmentation performance. This suggests that one can probe the training data

using a MAMS-type strategy to determine the optimal values of warp smoothness and atlas sharpness, and then use the SASS strategy for future registration and segmentation of new subjects. Furthermore, our experiments suggest that segmentation accuracy is tolerant up to a small mismatch between atlas sharpness and warp smoothness.

Further work will involve the application of our framework to other data sets (including volumetric data) and experiments with other models of data fidelity and regularization. We expect the optimal smoothness and atlas sharpness to be different but it would be interesting to verify if these values are consistent across subjects. It is possible that in the volumetric case where the intensity features are more predictive of labels, a MAMS-type strategy may outperform SASS, particularly in cases where the anatomy is dramatically different (e.g., ventricles in Alzheimer patients).

In this work, the optimal smoothness parameter found is global to the entire surface. Previous work demonstrated the advantage of using spatially varying smoothness parameters [8]. Unfortunately, the approach of exhaustive search presented here cannot be directly applied, since it would involve searching over a much larger space of parameters.

4.2 Comparison with FreeSurfer

We now compare the performance of our algorithm with the FreeSurfer parcellation algorithm, described in [17] and extensively validated in [11].

It is unclear which algorithm has a theoretical advantage. The FreeSurfer algorithm is essentially a "Single Atlas, Single Smoothness" (SASS) method that uses a sequential registration-segmentation approach and a more complicated anisotropic MRF model that has been specifically designed and fine-tuned for the surface parcellation application. Our model lacks the anisotropic MRF and introducing it would further improve its performance. On the other hand, FreeSurfer uses Iterated Conditional Modes [5] to solve the MRF, while we use the more powerful Mean Field approximation [22,26]. FreeSurfer also treats the subject mesh as a fixed image and the parcellation is performed directly on the subject mesh. Therefore, unlike our approach, no interpolation is necessary to obtain the final segmentation.

As shown in Figure 8, the optimal performances of MAMS, SAMS and SASS are statistically significantly better (even when corrected for multiple comparisons) than the FreeSurfer, with p-value 1×10^{-8} (SASS) for the left hemi- sphere and 8×10^{-4} (SASS) for the right hemisphere.

Because Dice computed over the entire surface can be deceiving by suppressing small structures, we show in Figure 9 the percentage improvement of SASS over FreeSurfer on inflated cortical surfaces. Qualitatively, we see that SASS performs better than FreeSurfer since there appears more orange-red regions than blue regions. The fact that the colorbar has significantly higher positive values than negative values indicates that there are parcellation regions with significantly greater improvements compared with regions that suffer significant penalties.

More quantitatively, Figure 10 and 11 display the average Dice per structure for FreeSurfer, MAMS, SAMS and SASS at ($S = 1$, $\alpha = 1$) for the left and right hemispheres respectively. Standard errors of the mean are displayed as red bars. The numbering of the structures correspond to Table 1. The structures with the worst Dice are the frontal pole, corpus callosum and entorhinal cortex. These structures are small and relatively poorly defined by the underlying cortical geometry. For example, the entorhinal cortex is partially defined by the rhinal sulcus, a tiny sulcus that is only visible on the pial surface. On the other hand, the corpus callosum is mapped from the white matter volume onto the cortical surface. Its boundary is thus defined by the surrounding structures, rather than by the cortical geometry.

For each structure, we perform a one-sided paired-sampled t-test between SASS and FreeSurfer, where each subject is considered a sample. We use the False Discovery Rate (FDR) [6] to correct for multiple comparisons. In the left hemisphere, SASS achieves statistically significant improvement over FreeSurfer for 17 structures (FDR < 0.05), while the remaining structures yield no statistical difference. In the right hemisphere, SASS achieves improvement for 11 structures (FDR < 0.05), while the remaining structures yield no statistical difference. The p-values for the left and right hemispheres are pooled together for the False Discovery Rate analysis.

A major factor influencing the accuracy of the automatic parcellation is the manual segmentation. In well-defined regions such as pre- and post-central gyri, accuracy is more than 90% and within the range of inter-rater variability. In the ambiguous regions, such as the frontal pole, inconsistent manual segmentation leads to less consistent training and worse segmentation performance.

5 Conclusions

In this paper, we proposed a generative model for the joint registration and segmentation of images. The atlas construction process is formulated as estimation of the graphical model parameters. The framework incorporates consistent atlas construction, multiple atlases of varying sharpness and MRF priors on both registration warps and segmentation labels. We show that atlas sharpness and warp regularization are important factors in segmentation and that the optimal smoothness parameters are stable across subjects in the context of cortical parcellation. The experiments imply that the optimal atlas sharpness and warp smoothness can be determined by cross-validation. Furthermore, segmentation accuracy is tolerant up to a small mismatch between atlas sharpness and warp smoothness. With the proper choice of atlas sharpness and warp regularization, even with a less complex MRF model, the joint registration-segmentation framework achieves better segmentation accuracy than the state-of-the-art FreeSurfer algorithm [11,17].

There are various promising directions for future work. We believe that incorporating multiscale features [32,53,54] can further improve the registration and parcellation of the cortical surfaces. Learning the optimal spatially varying smoothness parameters in the training set is also a problem we are currently working on. Finally, we plan to apply this framework to other problems, including volumetric segmentation.

A Mean Field Derivation Outline

The mean-field approximation uses a variational formulation [22], where we seek to minimize the KL-divergence (denoted by $D(\cdot \| \cdot)$) between $q(\hat{L}) = \prod_i \mathbf{b}_i(\hat{L}_i)$ and $p(\hat{L}|I(R(t)(\{x_i\}));A_\alpha)$:

$$\left\{\mathbf{b}_i^*\right\} = \underset{\{\mathbf{b}_i\}}{\operatorname{argmax}} \; D(q(\widehat{L}) \| p(\widehat{L}|I(R^{(t)}(\{x_i\}));A_\alpha)) \tag{A.1}$$

This results in a fixed-point iterative solution. Since this is a fairly standard derivation [22], we only provide the final update:

$$\mathbf{b}_i(m) \propto e^{U_i(m) + \log p(I(R^{(t)}(x_i))|\widehat{L}_i=m;A_\alpha) + \sum_{j \in N_i} \sum_{\widehat{L}_j} \mathbf{b}_j(\widehat{L}_j)\left[V(m,\widehat{L}_j)+V(\widehat{L}_j,m)\right]} \tag{A.2}$$

where $\mathbf{b}_i$ is normalized to be a valid probability mass function.

B Implementation

We now present some implementation details for completeness.

We first discuss the estimation of the atlas A_α defined by $\{U_i, V, W_i\}$ in Eq. (20) and Eq. (21) from the maximum likelihood function Eq. (7). In our model, estimating U_i and V is hard in practice, since evaluating Eq. (7) requires computing the NP-hard partition function. Instead, we use frequency counts to estimate the clique potentials, similar to FreeSurfer [17].

- In our implementation, the singleton potential U_i is a row vector of length M and L_n is a column indicator vector. We set

$$U_i = \log \frac{1}{N} \sum_n L_n^T (R_n(x_i)),$$

(B.1)

where $(\cdot)^T$ indicates transpose.

- The pairwise potential V is a MxM matrix. Following [17], we set

$$V = \log \frac{1}{2NE} \sum_n \sum_i \sum_{j \in N_i} L_n(R_n(x_i)) L_n^T(R_n(x_i)),$$

(B.2)

where E is the number of edges in the atlas mesh. More rigorous methods of optimizing the clique potentials through iterative proportional fitting [24] would further improve the clique potential estimates.

- The likelihood potential W_i is a row vector of length M defined at each vertex. The m-th entry of W_i corresponds to the likelihood of observing a particular image intensity or vectors of intensity (in our case, the local surface geometries) at location $R(x_i)$ given a particular label m. While we might be observing multiple geometric features at any vertex, the likelihood of these features is combined into the row vector W_i. We use maximum likelihood estimates of the mean and variance of the Gaussian distribution of cortical geometries conditioned on the segmentation labels to parameterize this distribution. In this work, we use the mean curvature of the original cortical surface and average convexity, which is a measure of sulcal depth [13], as intensity features.

At spatial locations where there is no training data for a particular label, it is unclear what the value of the entry in W should be since it is spatially varying. We simply assume a mean of zero and a large variance, essentially being agnostic in the value of intensity we observe. A more sophisticated method would involve the use of priors on the atlas parameters, so that the atlas parameters become random. In that case, when there is no observations, the maximum likelihood estimates of the atlas parameters become the priors.

Secondly, we discuss the registration of a training image to an atlas (Eq. (22)) and the new image registration (Eq. (24)).

- The registration warp R is a map from a 2-Sphere to a 2-Sphere. We represent R as a dense displacement field. In particular, each point x_i has an associated displacement vector u_i tangent to the point x_i on the sphere. $R(x_i)$ maps x_i to $x_i + u_i$ normalized to be a point on the sphere.

- To interpolate (for example the mean curvature of a cortical surface onto $R(x_i)$), we first find the intersection between the vector $R(x_i)$ and each planar triangle of the spherically mapped cortical surface. We then use barycentric interpolation to interpolate the values at the vertices of the mesh onto $R(x_i)$.

- The above two bullets completely specify the computation of the atlas co-registration objective function Eq. (22) and the new subject registration function Eq. (24). This allows us to compute the gradients of the objective function via the chain rule.

- We use conjugate gradient ascent with parabolic line search [36] on a coarse-to-fine grid. The coarse-to-fine grid comes from the representation of the atlas as a subdivided icosahedron mesh.

- The final segmentation is obtained by selecting for each vertex the label with the highest posterior probability.

- To satisfy $F(R)$, the regularization that induces invertibility, we ensure that no step in the line search results in folded triangles. Unfortunately, in practice, this results in many small steps. It is much more efficient to perform the line search without considering $F(R)$, and then unfold the triangles using $F(R)$ after the line search. In general, we find that after unfolding, the objective function is still better than the previous iteration. This unfolding process can be expensive for small smoothness parameter S ($S \leq 0.1$), resulting in long run times of about 1.5 hours per subject per atlas for $S \leq 0.1$.

## Acknowledgments

References

1. Allassonnière S, Amit Y, Trouvé A. Towards a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistiscal Society B 2007;68:3–29.

2. Ashburner J, Andersson J, Friston K. High-dimensional Image Registration Using Symmetric Priors. NeuroImage 1999;9:619–628. [PubMed: 10334905]

3. Ashburner J, Friston K. Unified Segmentation. NeuroImage 2005;26:839–851. [PubMed: 15955494]

4. Bhatia KK, Hajnal JV, Puri BK, Edwards AD, Rueckert D. Consistent Groupwise Non-Rigid Registration for Atlas Construction. International Symposium on Biomedical Imaging. 2004

5. Besag J. On the Statistical Analysis of Dirty Pictures. Journal of Royal Statistical Society B 1986;48 (3):259–302.

6. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of Royal Statistical Society B 1995;57(1):289–300.

7. Collins DL, Zijdenbos AP, Baaré WFC, Evans AC. ANIMAL+INSECT: Improved Cortical Structure Segmentation. Information Processing in Medical Imaging 1999;1613:210–223.

8. Commowick, O.; Stefanescu, R.; Fillard, P.; Arsigny, V.; Ayache, N.; Pennec, X.; Malandain, G. Incorporating Statistical Measures of Anatomical Variability in Atlas-to-Subject Registration for Conformal Brain Radiotherapy; International Conference on Medical Image Computing and Computer-Assisted Intervention; 2005. p. 927-934.LNCS

9. Dale A, Fischl B, Sereno M. Cortical Surface-Based Analysis I: Segmentation and Surface Reconstruction. NeuroImage 1999;9(2):179–194. [PubMed: 9931268]

10. De Craene, M.; du Bois d'Aische, A.; Macq, B.; Warfield, S. Multi-subject Registration for Unbiased Statistical Atlas Construction; International Conference on Medical Image Computing and Computer-Assisted Intervention; 2004. p. 655-662.LNCS

11. Desikan R, S'egonne F, Fischl B, Quinn B, Dickerson B, Blacker D, Buckner R, Dale A, Maguire P, Hyman B, Albert M, Killiany R. An automated labeling system for subdividing the human cerebral

cortex on MRI scans into gyral based regions of interest. NeuroImage 2006;31:968–980. [PubMed: 16530430]

12. 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. p. 1813-1817.IEEE Conference Record

13. Fischl B, Sereno M, Dale A. Cortical Surface-Based Analysis II: Inflation, Flattening, and a Surface-Based Coordinate System. NeuroImage 1999;9(2):195–207. [PubMed: 9931269]

14. Fischl B, Sereno M, Tootell R, Dale A. High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping 1999;8(4):272–284. [PubMed: 10619420]

15. Fischl B, Liu A, Dale A. Automated Manifold Surgery: Constructing Geometrically Accurate and Topologically Correct Models of the Human Cerebral Cortex. IEEE Transactions on Medical Imaging 2001;20(1):70–80. [PubMed: 11293693]

16. Fischl B, Salat D, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale A. Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. Neuron 2002;33(3):341–355. [PubMed: 11832223]

17. Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat D, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale A. Automatically Parcellating the Human cerebral Cortex. Cerebral Cortex 2004;14:11–22. [PubMed: 14654453]

18. Fischl B, Rajendran N, Busa E, Augustinack J, Hinds O, Yeo BTT, Mohlberg H, Amunts K, Zilles K. Cortical Folding Patterns and Predicting Cytoarchitecture. Cerebral Cortex. 2007

19. Glaunès J, Vaillant M, Miller M. Landmark Matching via Large Deformation Diffeomorphisms on the Sphere. Journal of Mathematical Imaging and Vision 2004;20:179–200.

20. Guimond A, Meunier J, Thirion J-P. Average Brain Models: A Convergence Study. Computer Vision and Image Understanding 2000;77(2):192–210.

21. Heckemann RA, Hajnal J, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 2006;33(1):115–126. [PubMed: 16860573]

22. Jaakkola, T. Advanced Mean Field Methods: Theory and Practice. MIT Press; 2000. Tutorial on variational approximation methods.

23. Joshi S, Davis B, Jomier M, Gerig G. Unbiased Diffeomorphic Atlas Construction for Computational Anatomy. NeuroImage 2004;23:151–160.

24. Jirousek R, Preucil S. On the effective implementation of the iterative proportional fitting procedure. Computational Statistics and Data Analysis 1995;19:177–189.

25. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J. A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development The International Consortium for Brain Mapping (ICBM). NeuroImage 1995;2(2):89–101. [PubMed: 9343592]

26. Kapur, T.; Grimson, WEL.; Kikinis, R.; Wells, W. Enhanced Spatial Priors for Segmentation of Magnetic Resonance Imagery; International Conference on Medical Image Computing and Computer Aided Intervention; 1998. p. 457-468.

27. Klein A, Hirsch J. Mindboggle: a scatterbrained approach to automate brain labeling. NeuroImage 2005;(24):261–280. [PubMed: 15627570]

28. Lohmann G, VonCramon DY. Automatic labelling of the human cortical surface using sulcal basins. Medical Image Analysis 2000;4:179–188. [PubMed: 11145307]

29. Lorenzen P, Prastawa M, Davis B, Gerig G, Bullitt E, Joshi S. Multi-Modal Image Set Registration and Atlas Formation. Medical Image Analysis 2006;10(3):440–451. [PubMed: 15919231]

30. Makrogiannis S, Verma R, Karacali B, Davatzikos C. A Joint Transformation and Residual Image Descriptor for Morphometric Image Analysis using an Equivalence Class Formulation. Mathematical Methods in Biomedical Image Analysis. 2006

31. Mangin, J-F.; Frouin, V.; Bloch, I.; Régis, J.; Lopez-Krahe, J. Automatic construction of an attributed relational graph representing the cortex topography using homotopic transformations; Proceedings of SPIE; 1994. p. 110-121.
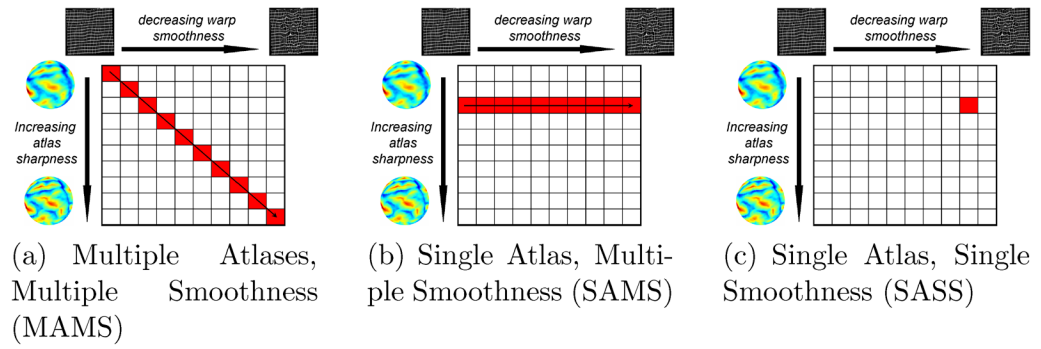
32. Nain D, Haker S, Bobick A, Tannenbaum A. Multiscale 3-D Shape Representation and Segmentation Using Spherical Wavelets. IEEE Transactions on Medical Imaging 2007;26(4):598–618. [PubMed: 17427745]

33. Nielsen, M.; Johansen, P.; Jackson, AD.; Lautrup, B. Brownian Warps: A Least Committed Prior for Non-rigid Registration; International Conference on Medical Image Computing and Computer-Assisted Intervention; 2002. p. 557-564.LNCS

34. Paus T, Tomaiuolo F, Otaky N, MacDonald D, Petrides M, Atlas J, Morris R, Evans A. Human Cingulate and Paracingulate Sulci: Pattern, Variability, Asymmetry, and Probabilistic Map. Cerebral Cortex 1996;6:207–214. [PubMed: 8670651]

35. Pohl K, Fisher J, Grimson WEL, Kikinis R, Wells W. A Bayesian model for joint segmentation and registration. NeuroImage 2006;31:228–239. [PubMed: 16466677]

36. Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. Numerical Recipes in C: The Art of Scientific Computing. 2nd ed.. Cambridge, U.K: Cambridge University Press;

37. Rettmann M, Han X, Xu C, Prince J. Automated Sulcal Segmentation Using Watersheds on the Cortical Surface. NeuroImage 2002;15:329–344. [PubMed: 11798269]

38. Richard F, Samson A, Cuenod C. A SAEM algorithm for the estimation of template and deformation parameters in medical image sequences. preprint MAP5. 2007 Nov;

39. Rivière D, Mangin J-F, Papadopoulos-Orfanos D, Martinez J-M, Frouin V, Régis J. Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. Medical Image Analysis 2002;6:77–92. [PubMed: 12044997]

40. Ségonne F, Pacheco J, Fischl B. Geometrically Accurate Topology-Correction of Cortical Surfaces Using Nonseparating Loops. IEEE Transactions on Medical Imaging 2007;26(4):518–529. [PubMed: 17427739]

41. Tao X, Prince J, Davatzikos C. Using a Statistical Shape Model to Extract Sulcal Curves on the Outer Cortex of the Human Brain. Transactions on Medical Imaging 2002;21:513–524.

42. Tu Z, Zheng S, Yuille A, Reiss A, Dutton R, Lee A, Galaburda AM, Dinov I, Thompson P, Toga A. Automated Extraction of the Cortical Sulci Based on a Supervised Learning Approach. IEEE Transactions on Medical Imaging 2007;26(4):541–552. [PubMed: 17427741]

43. Twining C, Cootes T, Marsland S, Petrovic V, Schestowitz R, Taylor C. A unified information theoretic approach to groupwise non-rigid registration and model building. Information Processing in Medical Imaging 2005;9:1–14. [PubMed: 17354680]

44. Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based tissue classification of MR images of the brain. IEEE Transactions on Medical Imaging 1999;18(10):897–908. [PubMed: 10628949]

45. Van Leemput, K. Probabilistic Brain Atlas Encoding Using Bayesian Inference; International Conference on Medical Image Computing and Computer Aided Intervention; 2006. p. 704-711.

46. Weisenfeld N, Warfield S. Simultaneous Alignment and Central Tendency Estimation for Brain Atlas Construction. Workshop on Statistical Registration: Pair-wise and Group-wise Alignment and Atlas Formation. 2007

47. Wyatt P, Noble A. MAP MRF Joint Segmentation and Registration. Medical Image Analysis 2003;7 (4):539–552. [PubMed: 14561557]

48. Xiaohua, C.; Brady, M.; Rueckert, D. Simultaneous Segmentation and Registration for Medical Image; International Conference on Medical Image Computing and Computer Aided Intervention; 2004. p. 663-670.

49. Xiaohua C, Brady M, Lo J, Moore N. Simultaneous Segmentation and Registration of Contrast-Enhanced Breast MRI. Information Processing in Medical Imaging 2005;3565:126–137. [PubMed: 17354690]

50. Yezzi A, Zollei L, Kapur T. A variational framework for integrating segmentation and registration through active contours. Medical Image Analysis 2003;7(2):171–185. [PubMed: 12868620]

51. Yeo, BTT.; Sabuncu, M.; Mohlberg, H.; Amunts, K.; Zilles, K.; Golland, P.; Fischl, B. What Data to Co-register for Computing Atlases; Proceedings of the International Conference on Computer Vision, IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis; 2007.

52. Yeo, BTT.; Sabuncu, M.; Desikan, R.; Fischl, B.; Golland, P. Effects of Registration Regularization and Atlas Sharpness on Segmentation Accuracy; International Conference on Medical Image Computing and Computer Aided Intervention; 2007. p. 683-691.

53. Yu P, Grant E, Qi Y, Han X, Ségonne F, Pienaar R, Busa E, Pacheco J, Makris N, Buckner R, Golland P, Fischl B. Cortical Surface Shape Analysis Based on Spherical Wavelets. IEEE Transaction on Medical Imaging 2007;26(4):582–597.

54. Yu, P.; Yeo, BTT.; Grant, E.; Fischl, B.; Golland, P. Cortical Folding Development Study based on Over-Complete Spherical Wavelets; Proceedings of MMBIA: IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis; 2007.
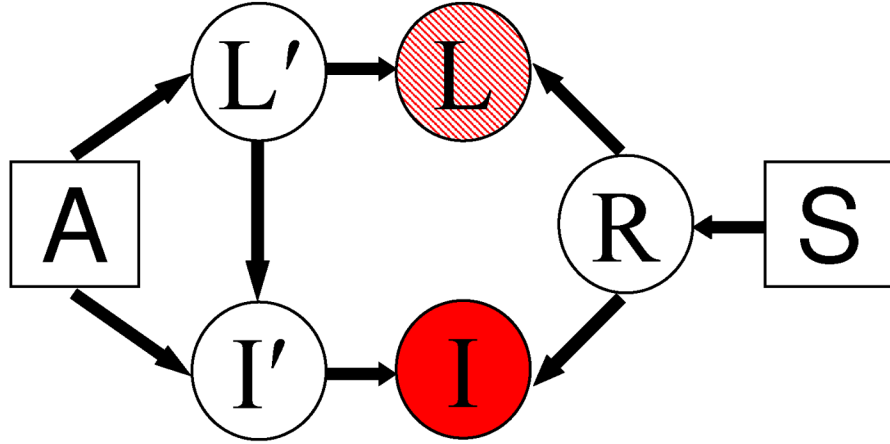
Fig. 1.
Generative model for registration and segmentation. *A* is an atlas used to generate the label map *L′* in some universal atlas space. The atlas *A* and label map *L′* generate image *I′*. *S* is the smoothness parameter that generates random warp field *R*. This warp is then applied to the label map *L′* and image *I′* to create the label map *L* and the image *I*. We assume the label map *L* is available for the training images, but not for the test image. The image *I* is observed in both training and test cases.

(a) Multiple Atlases, Multiple Smoothness (MAMS)

(b) Single Atlas, Multiple Smoothness (SAMS)

(c) Single Atlas, Single Smoothness (SASS)

Fig. 2.
Strategies for exploring space of atlas sharpness and warp smoothness of a new image.
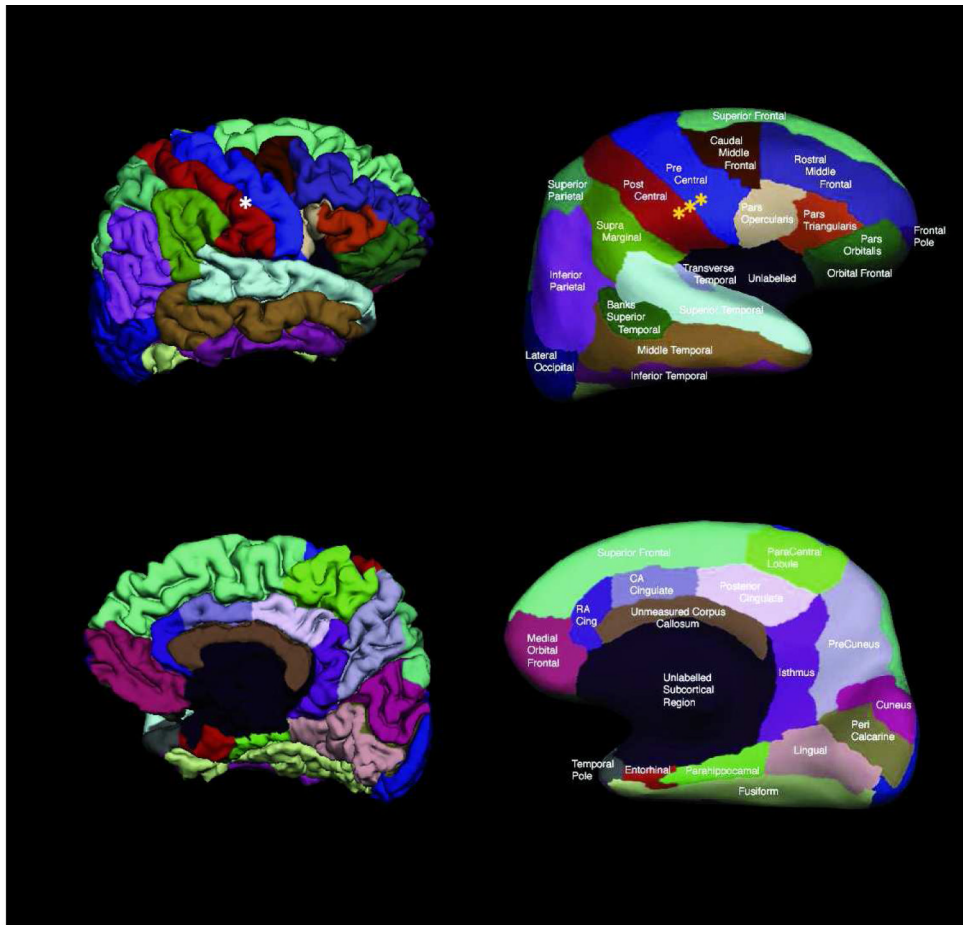
Fig. 3.
Example of manual parcellation shown on a partially inflated cortical surface. In our data set, the neuroanatomist preferred gyral labels to sulcal labels. There are also regions where sulci and gyri are grouped together as one label, such as the superior and inferior parietal complexes.

(a) Left Hemisphere

(b) Right Hemisphere
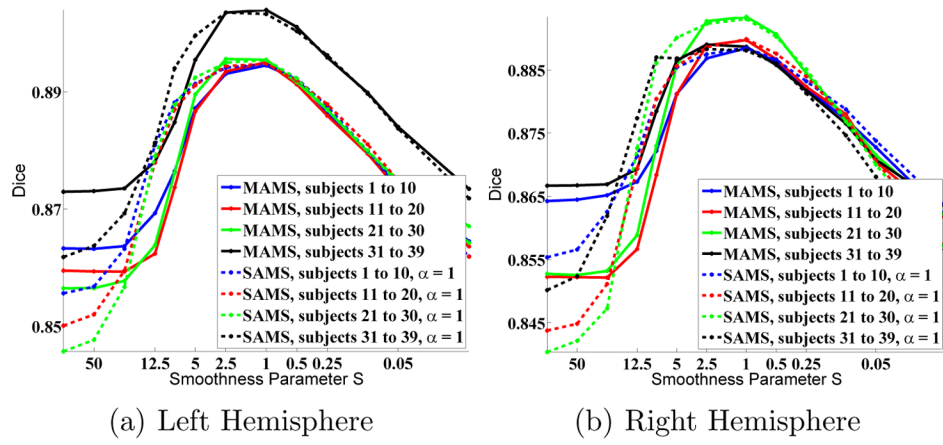
Fig. 4.
Parcellation accuracy as a function of warp smoothness. S is plotted on a log scale.

(a) Left Hemisphere   (b) Right Hemisphere
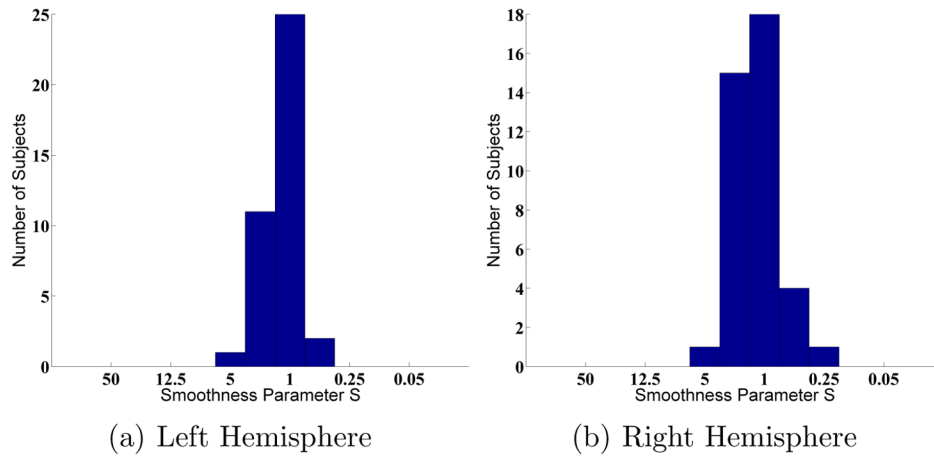
Fig. 5.
Histogram of optimal warp smoothness S across subjects (MAMS).

(a) Right Inferior Temporal Gyrus   (b) Right Temporal Pole
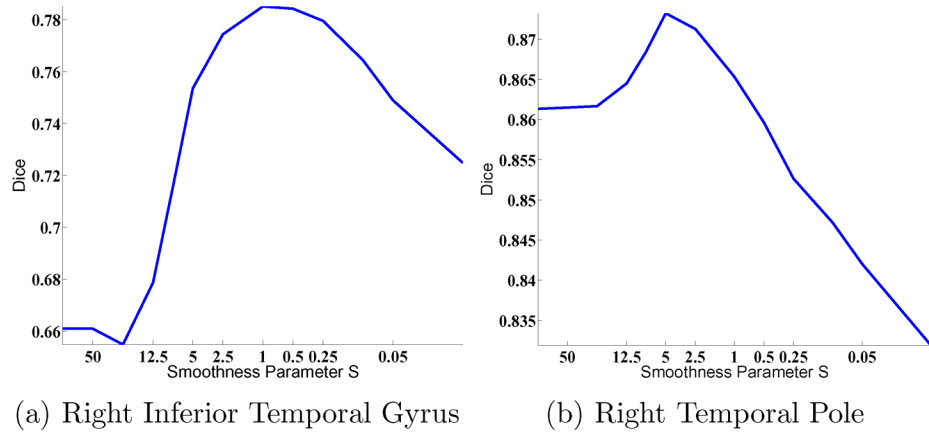
Fig. 6.
(a) Typical plot of Dice against smoothness *S*. (b) A noisy plot of Dice against smoothness *S*.

(a) Left Hemisphere   (b) Right Hemisphere
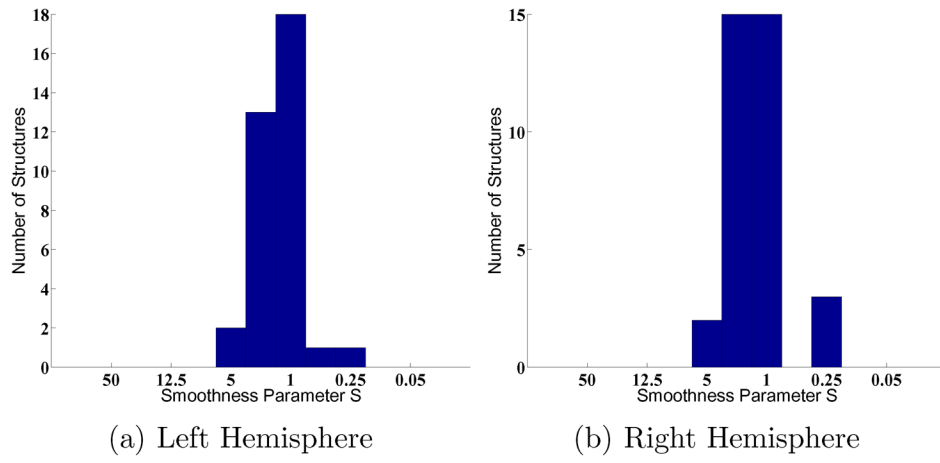
Fig. 7.
Histogram of optimal S across structures (MAMS).

(a) Left Hemisphere   (b) Right Hemisphere
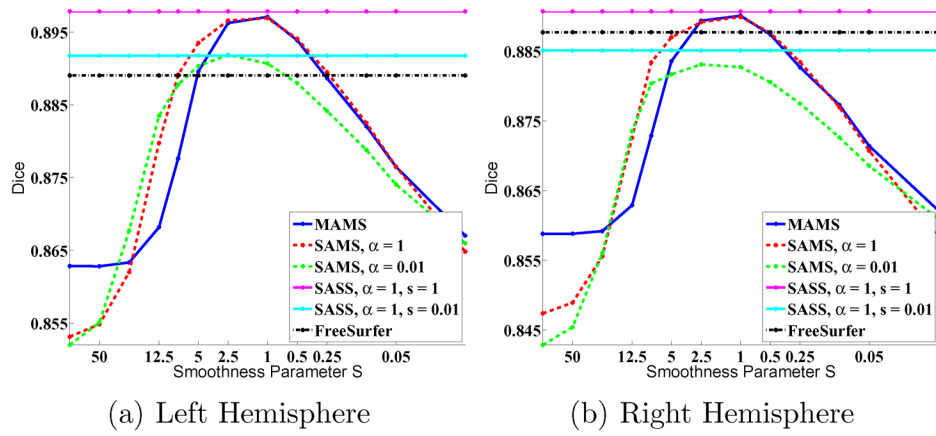
Fig. 8.
Overall Dice versus smoothness. S is plotted on a log scale.

(a) left lateral   (b) Right lateral

(c) Left medial   (d) Right medial
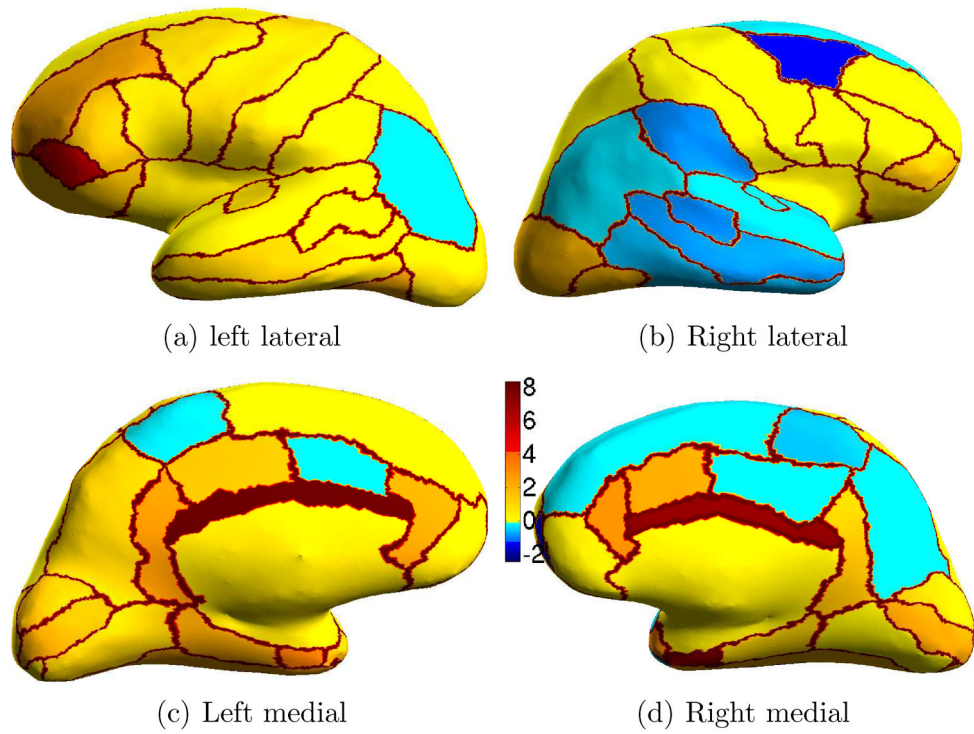
Fig. 9.
Percentage improvement of SASS over FreeSurfer. The boundaries between parcellation regions are set to reddish-brown so that the different regions are more visible.

(a) Left hemi structures: 1-9

(b) Left hemi structures: 10-18

(c) Left hemi structures: 19-27

(d) Left hemi structures: 28-35
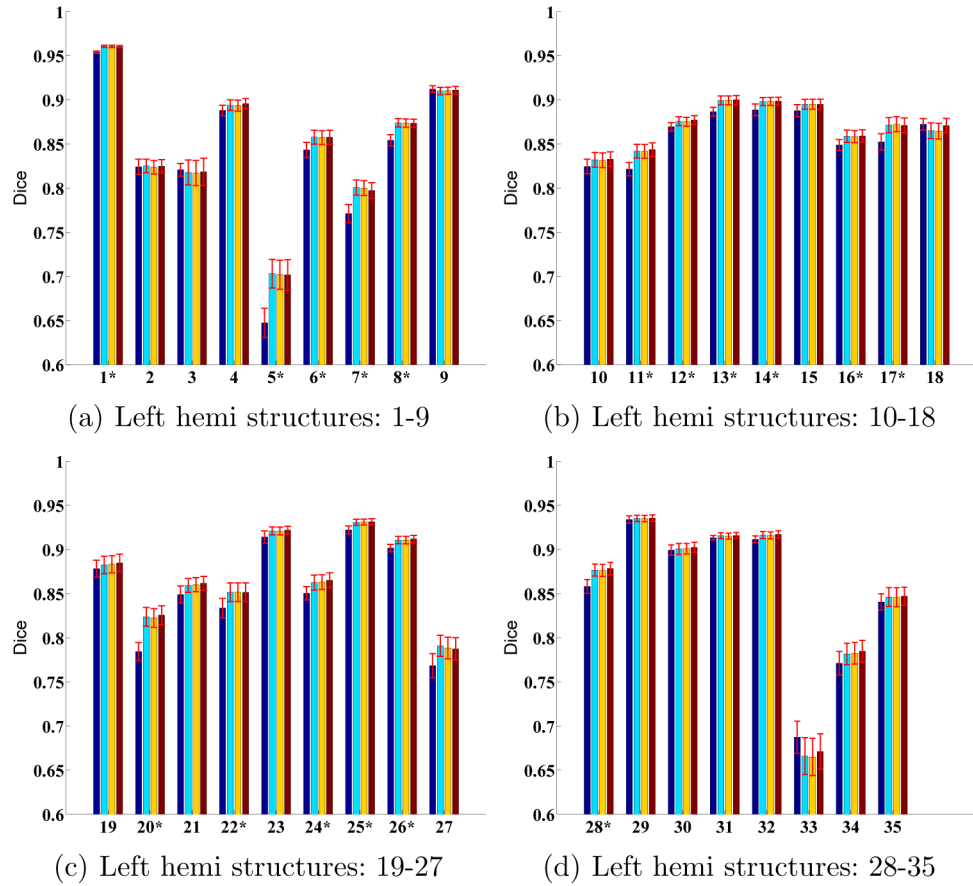
Fig. 10.
Structure-specific parcellation accuracy for the left hemisphere. First column (dark blue) corresponds to FreeSurfer. Second (light blue), third (yellow) and fourth (brown) columns correspond to MAMS, SAMS and SASS respectively. ($S = 1$, $\alpha = 1$). * indicates structures where SASS shows statistically significant improvement over FreeSurfer. There is no structure that becomes worse.

(a) Right hemi structures: 1-9

(b) Right hemi structures: 10-18

(c) Right hemi structures: 19-27

(d) Right hemi structures: 28-35
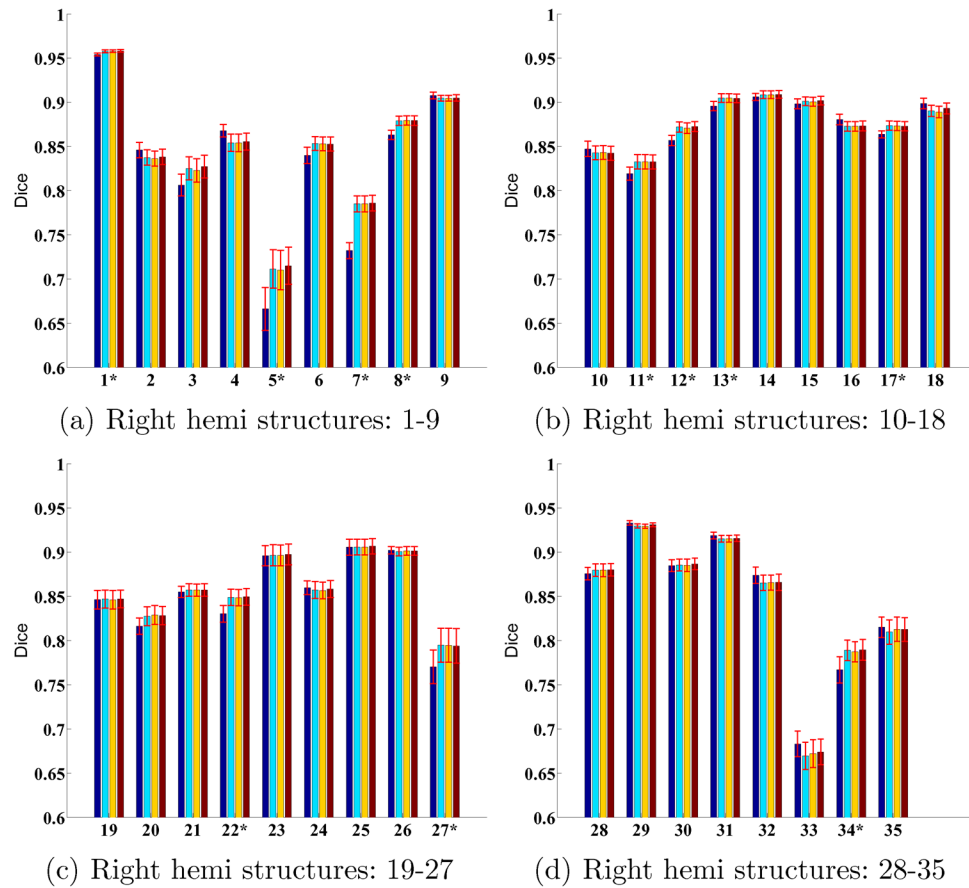
Fig. 11.
Structure-specific parcellation accuracy for the right hemisphere. First column (dark blue) corresponds to FreeSurfer. Second (light blue), third (yellow) and fourth (brown) columns correspond to MAMS, SAMS and SASS respectively. ($S = 1$, $\alpha = 1$). * indicates structures where SASS shows statistically significant improvement over FreeSurfer. There is no structure that becomes worse.

Table 1

List of Parcellation Structures

1.	Sylvian Fissure / Unknown	2.	Bank of the Superior Temporal Sulcus
3.	Caudal Anterior Cingulate	4.	Caudal Middle Frontal Gyrus
5.	Corpus Callosum	6.	Cuneus
7.	Entorhinal	8.	Fusiform Gyrus
9.	Inferior Parietal Complex	10.	Inferior Temporal Gyrus
11.	Isthmus Cingulate	12.	Lateral Occipital
13.	Lateral Orbito Frontal	14.	Lingual
15.	Medial Orbito Frontal	16.	Middle Temporal Gyrus
17.	Parahippocampal	18.	Paracentral
19.	Parsopercularis	20.	Parsorbitalis
21.	Parstriangularis	22.	Peri-calcarine
23.	Post-central Gyrus	24.	Posterior Cingulate
25.	Pre-central Gyrus	26.	Pre-cuneus
27.	Rostral Anterior Cingulate	28.	Rostral Middle Frontal
29.	Superior Frontal Gyrus	30.	Superior Parietal Complex
31.	Superior Temporal Gyrus	32.	Supramarginal
33.	Frontal Pole	34.	Temporal Pole
35.	Transverse Temporal		