Spherical Demons: Fast Diffeomorphic Landmark-Free Surface Registration

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/tmi.2009.2030797">http://dx.doi.org/10.1109/tmi.2009.2030797</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Fri Jan 11 17:35:29 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/59540">http://hdl.handle.net/1721.1/59540</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Spherical Demons: Fast Diffeomorphic Landmark-Free Surface Registration

B. T. Thomas Yeo*, Mert R. Sabuncu, Tom Vercauteren, Nicholas Ayache, Bruce Fischl, and Polina Golland

Abstract—We present the Spherical Demons algorithm for registering two spherical images. By exploiting spherical vector spline interpolation theory, we show that a large class of regularizers for the modified Demons objective function can be efficiently approximated on the sphere using iterative smoothing. Based on one parameter subgroups of diffeomorphisms, the resulting registration is diffeomorphic and fast. The Spherical Demons algorithm can also be modified to register a given spherical image to a probabilistic atlas. We demonstrate two variants of the algorithm corresponding to warping the atlas or warping the subject. Registration of a cortical surface mesh to an atlas mesh, both with more than 160 k nodes requires less than 5 min when warping the atlas and less than 3 min when warping the subject on a Xeon 3.2 GHz single processor machine. This is comparable to the fastest non-diffeomorphic landmark-free surface registration algorithms. Furthermore, the accuracy of our method compares favorably to the popular FreeSurfer registration algorithm. We validate the technique in two different applications that use registration to transfer segmentation labels onto a new image 1) parcellation of in vivo cortical surfaces and 2) Brodmann area localization in ex vivo cortical surfaces.

Index Terms—Cortical registration, demons, diffeomorphism, spherical registration, surface registration, vector field interpolation.

I. INTRODUCTION

MOTIVATED by many successful applications of the spherical representation of the cerebral cortex, this paper addresses the problem of registering two spherical images. Cortical folding patterns have been shown to correlate with both cytoarchitectural [25], [68] and functional regions [64], [27]. In group studies of cortical structure and function, determining corresponding folds across subjects is therefore important. There has been much effort focused on registering cortical surfaces in 3-D [22], [23], [30], [58]. Since cortical areas—both structure and function—are arranged in a mosaic across the cortical surface, an alternative approach is to model the surface as a 2-D closed manifold in 3-D and to warp the underlying spherical coordinate system [27], [50], [59], [60], [64], [67]. Warping the spherical coordinate system establishes correspondences across the surfaces without actually deforming the surfaces in 3-D.

Deformation Model: There is frequently a need for invertible deformations that preserve the topology of structural or functional regions across subjects. Unfortunately, this causes many spherical warping algorithms to be computationally expensive. Previously demonstrated methods for cortical registration [27], [60], [67] rely on soft regularization constraints to encourage invertibility. These require unfolding the mesh triangles, or limit the size of optimization steps to achieve invertibility [27], [67]. Elegant regularization penalties that guarantee invertibility exist [5], [46] but they explicitly rely on special properties of the Euclidean image space that do not hold for the sphere.

An alternative approach to achieving invertibility is to work in the group of diffeomorphisms [4], [7], [9], [22], [31], [43], [66]. In this case, the underlying theory of flows of vector fields can be extended to manifolds [11], [44], [47]. The large deformation diffeomorphic metric mapping (LDDMM) [7], [9], [22], [31], [43] is a popular framework that seeks a time-varying velocity field representation of a diffeomorphism. Because LDDMM optimizes over the entire path of the diffeomorphism, the resulting method is slow and memory intensive. By contrast, Ashburner [4] and Hernandez et al. [33] consider diffeomorphic transformations parameterized by a single stationary velocity field. While restricting the space of solutions reduces the memory needs relative to LDDMM, these algorithms still have to consider the entire trajectory of the deformation induced by the velocity field when computing the gradients of the objective function, leading to a long run time. We note that recent algorithmic advances [34], [43] promise to improve the speed and relieve the memory requirements of both LDDMM and the stationary velocity approach.
In this work, we adopt the approach of the Diffeomorphic Demons algorithm [66], demonstrated in the Euclidean image space, which constructs the deformation space that contains compositions of diffeomorphisms, each of which is parameterized by a stationary velocity field. Unlike the Euclidean Diffeomorphic Demons, the Spherical Demons algorithm utilizes velocity vectors tangent to the sphere and not arbitrary 3-D vectors. This constraint need not be taken care of explicitly in our algorithm since we directly work with the tangent spaces. In each iteration, the method greedily seeks the locally optimal diffeomorphism to be composed with the current transformation. As a result, the approach is much faster than LDDMM [7], [9], [22], [31] and its simplifications [4], [33]. A drawback is that the path of deformation is no longer a geodesic in the group of diffeomorphisms.

**Image Similarity Versus Regularization Tradeoffs:** Another challenge in registration is the tradeoff between the image similarity measure and the regularization in the objective function. Since most types of regularization favor smooth deformations, the gradient computation is complicated by the need to take into account the deformation in neighboring regions. For Euclidean images, the popular Demons algorithm [57] can be interpreted as optimizing an objective function with two regularization terms [14], [66]. The special form of the objective function facilitates a fast two-step optimization where the second step handles the warp regularization via a single convolution with a smoothing filter.

Using spherical vector spline interpolation theory [31] and other differential geometric tools, we show that the two-stage optimization procedure of Demons can be efficiently approximated on the sphere. We note that the problem is not trivial since tangent vectors at different points on the sphere are not directly comparable. We also emphasize that this decoupling of the image similarity and the warp regularization could also be accomplished with a different space of admissible warps, e.g., spherical thin plate splines [72].

**Interpolation:** Yet another reason why spherical image registration is slow is because of the difficulty in performing interpolation on a spherical grid, unlike a regular Euclidean grid. In this paper, we use existing methods for interpolation, requiring about 1 s to interpolate data from a spherical mesh of 160 k vertices onto another spherical mesh of 160 k vertices. Recent work on using different coordinate charts of the sphere [63] promises to further speed up our implementation of the Spherical Demons algorithm.

While most discussion in this paper concentrates on pairwise registration of spherical images, the proposed Spherical Demons algorithm can be modified to incorporate a probabilistic atlas. We derive and implement two variants of the algorithm for registration to an atlas corresponding to whether we warp the atlas or the subject. On a Xeon 3.2 GHz single processor machine, registration of a cortical surface mesh to an atlas mesh, both with more than 160 k nodes, requires less than 5 min when warping the atlas and less than 3 min when warping the subject. Note that the registration runtime reported includes registration components dealing with rotation, which takes up one quarter of the total runtime. The total runtime is comparable to other nonlinear landmark-free cortical surface registration algorithms whose runtime ranges from minutes [23], [60] to more than an hour [27], [67]. However, the other fast algorithms suffer from folding spherical triangles [60] and intersecting triangles in 3-D [23] since only soft constraints are used. No runtime comparison can be made with spherical registration algorithm of the LDDMM type because to the best of our knowledge, no landmark-free LDDMM spherical registration algorithm that handles cortical surfaces has been developed yet.

Unlike landmark-based methods for surface registration [8], [22], [31], [50], [58], [64], we do not assume the existence of corresponding landmarks. Landmark-free methods have the advantage of allowing for a fully automatic processing and analysis of medical images. Unfortunately, landmark-free registration is also more challenging, because no information about correspondences are provided. The difficulty is exacerbated for the cerebral cortex since different sulci and gyri appear locally similar. Nevertheless, we demonstrate that our algorithm is accurate in both cortical parcellation and cyto-architectonic localization applications.

The Spherical Demons algorithm for registering cortical surfaces presented here does not take into account the metric properties of the original cortical surface. FreeSurfer [27] uses a regularization that penalizes deformation of the spherical coordinate system based on the distortion computed on the original cortical surface. Thompson et al. [59] suggest the use of Christoffel symbols [39] to correct for the metric distortion of the initial spherical coordinate system during the registration process. However, it is unclear whether correcting for the metric properties of the cortex is important in practice, since we demonstrate that the accuracy of the Spherical Demons algorithm compares favorably to that of FreeSurfer. A possible reason is that we initialize the registration with a spherical parametrization that minimizes metric distortion between the spherical representation and the original cortical surface [27].

This paper is organized as follows. In the next section, we discuss the classical Demons algorithm [57] and its diffeomorphic variant [66]. In Section III, we present the extension of the Diffeomorphic Demons algorithm to the sphere. We conclude with experiments in Section IV and further discussion in Section V. The appendices provide technical and implementation details of the Spherical Demons algorithm and the extension to probabilistic atlases. This paper extends a previously presented conference article [69] and contains detailed derivations and discussions that were left out in the conference version. We note that our Spherical Demons code is freely available. To summarize:

1. We demonstrate that the Demons algorithm can be efficiently extended to the sphere.
2. We demonstrate that the use of a limited class of diffeomorphisms combined with the Demons algorithm yields a speed gain of more than an order of magnitude compared with other landmark-free invertible spherical registration methods, such as [27] and [67].

There are two versions of the code (Matlab and ITK) available at http://www.sites.google.com/site/yeoyeo02/software/sphericaldemonsrelease. The Matlab code is used in the experiments of this paper. The preliminary ITK code [35]–[37] can also be found at http://wwwhdl.handle.net/10380/3117.
3) We validate our algorithm by demonstrating an accuracy comparable to that of the popular FreeSurfer algorithm [27] in two different applications.

II. BACKGROUND—DEMONS ALGORITHM

Algorithm 1. Demons Algorithm

Data: A fixed image $F$ and moving image $M$.

Result: Transformation $\Gamma$ so that $M \circ \Gamma$ is “close” to $F$.

Set $\mathcal{Y}^0 = \text{identity transformation}$ (or some a-priori transformation, e.g., from a previous registration)

repeat

Step 1. Given $\mathcal{Y}^{(t)}$,

Minimize the first two terms of (3):

$$u^{(t)} = \arg\min_u \left[ \Sigma^{-1} \left( F - M \circ \{ \mathcal{Y}^{(t)} \circ u \} \right) \right]$$

$$+ \frac{1}{\sigma_x^2} \text{dist} \left( \mathcal{Y}^{(t)}, \{ \mathcal{Y}^{(t)} \circ u \} \right)$$

(1)

where $u$ is any admissible transformation. Compute $\Gamma^{(t)} = \mathcal{Y}^{(t)} \circ u^{(t)}$.

Step 2. Given $\Gamma^{(t)}$,

Minimize the last two terms of (3):

$$\mathcal{Y}^{(t+1)} = \arg\min_{\mathcal{Y}} \frac{1}{\sigma_x^2} \text{dist} \left( \mathcal{Y}, \Gamma^{(t)} \right)$$

$$+ \frac{1}{\sigma_T^2} \text{Reg} \left( \mathcal{Y} \right)$$

(2)

until convergence;

We choose to work with the modified Demons objective function [14], [66]. Let $F$ be the fixed image, $M$ be the moving image and $\Gamma$ be the desired transformation that deforms the moving image $M$ to match the fixed image $F$. Throughout this paper, we assume that $F$ and $M$ are scalar images, even though it is easy to extend the results to vector images [70]. We introduce a hidden transformation $\mathcal{Y}$ and seek

$$\left( \mathcal{Y}^*, \Gamma^* \right) = \arg\min_{\mathcal{Y},\Gamma} \left[ \Sigma^{-1} (F - M \circ \Gamma) \right]^2$$

$$+ \frac{1}{\sigma_x^2} \text{dist} \left( \mathcal{Y}, \Gamma \right) + \frac{1}{\sigma_T^2} \text{Reg} \left( \mathcal{Y} \right)$$

(3)

In this case, the fixed image $F$ and warped moving image $M \circ \Gamma$ are treated as $N \times 1$ vectors. Typically, $\text{dist} \left( \mathcal{Y}, \Gamma \right) = \| \mathcal{Y} - \Gamma \|^2$, encouraging the resulting transformation $\Gamma$ to be close to the hidden transformation $\mathcal{Y}$ and $\text{Reg} \left( \mathcal{Y} \right) = \| \nabla (\mathcal{Y} - \Gamma) \|^2$, i.e., the regularization penalizes the gradient magnitude of the displacement field $\mathcal{Y} - \Gamma$. $\sigma_x$ and $\sigma_T$ provide a tradeoff among the different terms of the objective function. $\Sigma$ is typically a diagonal matrix that models the variability of a feature at a particular voxel. It can be set manually or estimated during the construction of an atlas.

This formulation facilitates a two-step optimization procedure that alternately optimizes the first two (first and second) and last two (second and third) terms of (3). Starting from an initial displacement field $\mathcal{Y}^0$, the Demons algorithm iteratively seeks an update transformation to be composed with the current estimate, as summarized in Algorithm 1.

In the original Demons algorithm [57], the space of admissible warps includes all 3-D displacement fields, and the composition operator $\circ$ corresponds to the addition of displacement fields. The resulting transformation might therefore be not invertible. In the Diffeomorphic Demons algorithm [66], the update $u$ is a diffeomorphism from $\mathbb{R}^3$ to $\mathbb{R}^3$ parameterized by a stationary velocity field $\mathcal{v}$. Note that $\mathcal{v}$ is a function that associates a tangent vector with each point in $\mathbb{R}^3$. Under certain mild smoothness conditions, a stationary velocity field $\mathcal{v}$ is related to a diffeomorphism through the exponential mapping $u = \exp(\mathcal{v})$. In this case, the stationary ODE $\partial x(t)/\partial t = \mathcal{v}(x(t))$ with the initial condition $x(0) \in \mathbb{R}^3$ yields $\exp(\mathcal{v})(x(0))$ as a solution at time $1$, i.e., $x(1) = \exp(\mathcal{v})(x(0)) \in \mathbb{R}^3$. In this case, $\exp(\mathcal{v})(x(0))$ maps point $x(0)$ to point $x(1)$.

The Demons algorithm and its variants are fast because for certain forms of $\text{dist} \left( \mathcal{Y}, \Gamma \right)$ and $\text{Reg} \left( \mathcal{Y} \right)$, Step 1 reduces to a nonlinear least-squares problem that can be efficiently minimized via Gauss–Newton optimization and Step 2 can be solved by a single convolution of the displacement field $\Gamma$ with a smoothing kernel. The proof of the reduction of Step 2 to a smoothing operation is illuminating and related to a smoothing operation is only an approximation on the sphere. We will see in Section III-D that the reduction of Step 2 to a smoothing operation is only an approximation on the sphere.

III. SPHERICAL DEMONS

In this section, we demonstrate suitable choices of $\text{dist} \left( \mathcal{Y}, \Gamma \right)$ and $\text{Reg} \left( \mathcal{Y} \right)$ that lead to efficient optimization of the modified Demons objective function in (3) on the unit sphere $S^2$. We construct updates $u$ as diffeomorphisms from $S^2$ to $S^2$ parameterized by a stationary velocity field $\mathcal{v}$. We emphasize that unlike Diffeomorphic Demons [66], $\mathcal{v}$ is a tangent vector field on the sphere and not an arbitrary 3-D vector field. A glossary of common terms used throughout the paper is found in Table I.

A. Choice of $\text{dist} \left( \mathcal{Y}, \Gamma \right)$

Suppose the transformations $\Gamma$ and $\mathcal{Y}$ map a point $x_n \in S^2$ to two different points $\Gamma(x_n) \in S^2$ and $\mathcal{Y}(x_n) \in S^2$, respectively. An intuitive notion of distance between $\Gamma(x_n)$ and $\mathcal{Y}(x_n)$ would be the geodesic distance between $\Gamma(x_n)$ and $\mathcal{Y}(x_n)$. Therefore, we could define $\text{dist} \left( \mathcal{Y}, \Gamma \right) = \sum_{n=1}^{N} \text{geodesic}(\mathcal{Y}(x_n), \Gamma(x_n))$. For reasons that will become clear in Section III-D, we prefer to define $\text{dist} \left( \mathcal{Y}, \Gamma \right)$ in terms of a tangent vector representation of the transformations $\Gamma$ and $\mathcal{Y}$, illustrated in Fig. 1, where the
TABLE I
Glossary of Terms Used Throughout the Paper

| $F, M$ | Fixed image $F$, moving image $M$. |
| $\Sigma, \sigma$ | Typically a diagonal matrix that models variability of feature values at a particular vertex. |
| $\Gamma, \bar{\Gamma}$ | Parameters of Demons cost function in Eq. (3). |
| $\tilde{\Gamma}, \bar{\tilde{\Gamma}}$ | Transformations from $S^2$ to $S^2$. $\Gamma$ is the transformation we are seeking. $\bar{\Gamma}$ is the smooth hidden transformation close to $\Gamma$. Discrete tangent vector representation of the deformations (see Fig. 1 and Eq. (5)). For example, given the tangent vector $\bar{\Gamma}_n$ at $x_n \in S^2$, one can compute $\bar{\Gamma}(x_n)$. |
| $\hat{v}$ | We parameterize diffeomorphic transformations from $S^2$ to $S^2$ by a composition of diffeomorphisms, each parameterized by a stationary velocity field $v$. $\hat{v}$ is the velocity vector at $x_n$. The diffeomorphism parameterized by the stationary velocity field $v$ is the solution of a stationary ODE at time 1. |
| $u(\cdot)$ | $\hat{v}$ and $\tilde{v}$ are orthonormal vectors tangent to the sphere at $x_n$. |
| $E_n \triangleq [\tilde{e}_1^T \tilde{e}_2^T]$ | Coordinate chart defined in Eq.(10): $\Psi_n(x') = \frac{x_n + E_n x''}{\|E_n x''\|}$, $\Psi_n$ is a diffeomorphism between $\mathbb{R}^2$ and a hemisphere centered at $x_n \in S^2$. $\tilde{e}_n$ is an arbitrary tangent vector at the origin of $\mathbb{R}^2$. At $x_n$, the velocity vector $\hat{v}_n = E_n \tilde{e}_n$ via the coordinate chart $\Psi_n$ (see Eq.(14)). |

Fig. 1. Tangent vector representation of transformation $\Gamma$. See text for more details.

length of the tangent vector encodes the amount of deformation.

Let $T_{x_n} S^2$ be the tangent space at $x_n$. We define $\tilde{\Gamma}_n \in T_{x_n} S^2$ to be the tangent vector at $x_n$ pointing along the great circle connecting $x_n$ to $\Gamma(x_n)$. In this work, we set the length of $\tilde{\Gamma}_n$ to be equal to the sine of the angle between $x_n$ and $\Gamma(x_n)$. With this particular choice of length, there is a one-to-one correspondence between $\Gamma(x_n)$ and $\tilde{\Gamma}_n$, assuming the angle between $x_n$ and $\Gamma(x_n)$ is less than $\pi/2$, which is a reasonable assumption even for relatively large deformations. The choice of this length leads to a compact representation of $\tilde{\Gamma}_n$ via vector products. We define $G_n$ to be the $3 \times 3$ skew-symmetric matrix representing the cross-product of $x_n$ with any vector

$$G_n = \begin{pmatrix} 0 & -x_n(3) & x_n(2) \\ x_n(3) & 0 & -x_n(1) \\ -x_n(2) & x_n(1) & 0 \end{pmatrix}$$

where $x_n(i)$ is the $i$th coordinate of $x_n$. Thus, $x_n \times \Gamma(x_n) = G_n \Gamma(x_n)$. Then on a unit sphere, we obtain

$$\tilde{\Gamma}_n = -x_n \times (x_n \times \Gamma(x_n)) = -G_n \Gamma(x_n).$$

A more intuitive choice for the length of $\tilde{\Gamma}_n$ might be the geodesic distance between $x_n$ and $\Gamma(x_n)$. If we restrict $\tilde{\Gamma}_n$ to be at most length $\pi$, there is a one-to-one mapping between this choice of the tangent vector $\tilde{\Gamma}_n$ and the resulting transformation $\Gamma(x_n)$. Indeed, such a choice of a tangent vector corresponds to an exponential map of $S^2$ [39]. The resulting expression for $\tilde{\Gamma}_n = (2C_n^2 \Gamma(x_n))[\Gamma\mathfrak{g}^2(\Gamma(x_n))]\sin^{-1}((\Gamma(x_n) - x_n)/2)$ is feasible, but more complicated than (5). In this paper, for simplicity, we follow the definition in (5).

Given $N$ vertices $(x_n)_{n=1}^N$, the set of transformed points $(\Gamma(x_n))_{n=1}^N$—or equivalently the tangent vectors $(\tilde{\Gamma}_n)_{n=1}^N$—together with a choice of an interpolation function define the transformation $\Gamma$ everywhere on $S^2$. Similarly, we can define the transformation $\bar{\Gamma}$ or the equivalent tangent vector field $\bar{\tilde{\Gamma}}$ through a set of $N$ tangent vectors $\{\tilde{\Gamma}_n\}_{n=1}^N$. We emphasize that these tangent vector fields are simply a convenient representation of the transformations $\gamma$ and $\Gamma$ and should not be confused with the stationary velocity field $v$ that will be used later on. We now set

$$\text{dist}(\gamma, \Gamma) = \sum_{n=1}^N \|\tilde{\Gamma}_n - \tilde{\Gamma}_n\|^2$$

which is well-defined since both $\tilde{\Gamma}_n$ and $\bar{\tilde{\Gamma}}_n$ belong to $T_{x_n} S^2$ for each $n = 1, \ldots, N$.

B. Spherical Demons Step 1

In this section, we show that the update for Step 1 of the Spherical Demons algorithm can be computed independently for each vertex. With our choice of $\text{dist}(\gamma, \Gamma)$, step 1 of the algorithm becomes a minimization with respect to the velocity field $v \triangleq \{v_n \in T_{x_n} S^2\}_{n=1}^N$. By substituting $u = \exp(\bar{v})$ and $\text{dist}(\gamma, \Gamma) = \sum_{n=1}^N \|\tilde{\Gamma}_n - \Gamma_n\|^2$ into (1), we obtain

$$\bar{v}^\top = \arg\min_v f(\bar{v})$$

$$= \arg\min_v \left\|\sum_{n=1}^N \left(F(\gamma(x_n) - \Gamma_n) - \Gamma_n \right) + \frac{1}{\sigma_x^2} \text{dist}(\gamma, \Gamma) \right\|^2$$

$$= \arg\min_v \sum_{n=1}^N \frac{1}{\sigma_x^2} \left(F(\gamma(x_n) - \Gamma_n) - \Gamma_n \right)^2$$

where $F(\cdot)$ is the mean operator.
The chart allows a reparameterization of the constrained optimization problem \( f \) in step 1 of the Spherical Demons algorithm into an unconstrained one:

\[
\begin{align*}
\psi_n(0) &= x_n \\
\vec{v}_n &= [\vec{e}_{n1} \ \vec{e}_{n2} \ \vec{z}_n] \\
\end{align*}
\]

\[
\begin{align*}
\nabla_n &= \Psi_n'(0) E_n z_n \\
\end{align*}
\]

where \( \sigma_n^2 \) is the \( n \)th diagonal entry of \( \Sigma \) and \( \circ \) denotes warp composition.

**Defining Coordinate Charts on the Sphere:** The cost function in (9) is a mapping from the tangent bundle \( TS^2 \) to the real numbers \( \mathbb{R} \). We can think of each tangent vector \( \vec{v}_n \) as a \( 3 \times 1 \) vector in \( \mathbb{R}^3 \) tangent to the sphere at \( x_n \). Therefore \( \vec{v}_n \) has 2 degrees of freedom and (9) represents a constrained optimization problem. Instead of dealing with the constraints, we employ coordinate charts that are diffeomorphisms (smooth and invertible mappings) between open sets in \( \mathbb{R}^2 \) and open sets on \( S^2 \). The differential of the coordinate chart establishes correspondences between the tangent bundles \( \mathbb{R}^2 \) and \( TS^2 \) [39], [44], so we can reparameterize the constrained optimization problem into an unconstrained one in terms of \( \mathbb{R}^2 \) (see Fig. 2).

It is a well-known fact in differential geometry that covering \( S^2 \) requires at least two coordinate charts. Since the tools of differential geometry are coordinate-free [39], [44], our results are independent of the choice of the coordinate charts. Let \( \vec{e}_{n1}, \vec{e}_{n2} \) be any two orthonormal \( 3 \times 1 \) vectors tangent to the sphere at \( x_n \), where orthonormality is defined via the usual Euclidean inner product in 3-D. In this work, for each mesh vertex \( x_n \), we define a local coordinate chart \( \Psi_n : \mathbb{R}^2 \rightarrow S^2 \)

\[
\Psi_n(x') = \frac{x_n + E_n x'}{||x_n + E_n x'||}, \quad \text{where } E_n = [\vec{e}_{n1} \vec{e}_{n2}].
\]

As illustrated in Fig. 2, \( \Psi_n(0) = x_n \). Let \( z_n \) be a \( 2 \times 1 \) tangent vector at the origin of \( \mathbb{R}^2 \). With the choice of the coordinate chart above, the corresponding tangent vector at \( x_n \) is given by the differential of the mapping \( \Psi_n(\cdot) \) evaluated at \( x' = 0 \)

\[
\vec{v}_n = D\Psi_n(0) z_n
\]

\[
= I_{3 \times 3} - \Psi_n(0)\Psi_n'(0) E_n z_n
\]

\[
= I_{3 \times 3} - x_n e_{n1}^T E_n z_n
\]

\[
= E_n z_n = [\vec{e}_{n1} \vec{e}_{n2}] z_n.
\]

The above equation defines the mapping of a tangent vector \( z_n \) at the origin of \( \mathbb{R}^2 \) to the tangent vector \( \vec{v}_n \) at \( x_n \) via the differential of the coordinate chart \( D\Psi_n \) at \( x' = 0 \). We note that for a tangent vector at an arbitrary point in \( \mathbb{R}^3 \), the expression for the corresponding tangent vector on the sphere is more complicated. This motivates our definition of a separate chart for each mesh vertex, to simplify the derivations.

**Gauss–Newton Step of Spherical Demons:** From (14), we obtain \( \exp(\vec{v}) = \exp(\{\vec{v}_n\}) = \exp(\{E_n z_n\}) \) and rewrite (9) as an unconstrained optimization problem

\[
\begin{align*}
\{\xi_n(t)\} &= \arg\min_{\{z_n\}} \sum_{n=1}^N \frac{1}{\sigma_n^2} \left( F_n(x_n) - M \circ \left( \gamma_n(t) \circ \exp(\{E_n z_n\}) \right)(x_n) \right)^2 \\
&\quad+ \frac{1}{\sigma_n^2} \sum_{n=1}^N \left( \psi_n(t) + G_n^2 \left( \gamma_n(t) \circ \exp(\{E_n z_n\}) \right)(x_n) \right)^2 \\
&= \arg\min_{\{z_n\}} \sum_{n=1}^N \frac{1}{\sigma_n^2} \left( F_n(z_n) + 1 \right) \sum_{n=1}^N ||g_n||^2(z).
\end{align*}
\]

This nonlinear least-squares form can be optimized efficiently with the Gauss–Newton method, which requires finding the gradient of both terms with respect to \( \{z_n\} \) at \( z_n = 0 \) and solving a linearized least-squares problem.

We let \( \tilde{m}_n^T \) be the \( 1 \times 3 \) spatial gradient of the warped moving image \( M \circ \gamma_n(t) \) at \( x_n \) and note that \( \tilde{m}_n^T \) is tangent to the sphere at \( x_n \). The computation of \( \tilde{m}_n^T \) is discussed in Appendix A-A. Defining \( u_n = \exp(\{E_n z_n\})(x_n) \), we differentiate the first term of the cost function \( f_n(z) \) in (15) using the chain rule, resulting in the \( 1 \times 2 \) vector

\[
\frac{\partial}{\partial \tilde{z}_k} \left[ F_n(x_n) - M \circ \left( \gamma_n(t) \circ \exp(\{E_n z_n\}) \right)(x_n) \right]_{z=0} = \frac{\partial}{\partial \tilde{z}_k} \left. M \circ \left( \gamma_n(t) \circ \exp(\{E_n z_n\}) \right)(x_n) \right|_{z=0} = \frac{\partial}{\partial \tilde{z}_k} \left. M \circ \left( \gamma_n(t) \circ \exp(\{E_n z_n\}) \right)(x_n) \right|_{z=0} = \frac{\partial}{\partial \tilde{z}_k} \left. \exp(\{E_n z_n\})(x_n) \right|_{z=0} = \frac{\partial}{\partial \tilde{z}_k} \delta(k, n)
\]

where \( \delta(k, n) = 1 \) if \( k = n \) and 0 otherwise. (20) uses the fact that the differential of \( \exp(\vec{v}) \) at \( \vec{v} = 0 \) is the identity [47], i.e.,

\[
[D\exp(0)]_{\vec{v}} = I. \quad \text{In other words, a change in velocity } \vec{v} \text{ at vertex } x_k \text{ does not affect } \exp(\vec{v})(x_n) \text{ for } n \neq k \text{ up to the first order derivatives.}
\]

Similarly, we define \( S_n^T \) to be the \( 3 \times 3 \) Jacobian of \( \gamma_n(\cdot) \) at \( x_n \). The computation of \( S_n^T \) is discussed in Appendix A-B. Differentiating the second term of the cost function \( g_n(z) \) in (15) using the chain rule, we get the \( 3 \times 2 \) matrix

\[
\frac{\partial}{\partial \tilde{z}_k} \left[ \psi_n(t) + G_n^2 \left( \gamma_n(t) \circ \exp(\{E_n z_n\}) \right)(x_n) \right]_{z=0} = G_n^2 \tilde{m}_n^T E_n \delta(k, n)
\]

\[
\text{Fig. 2. Coordinate chart of the sphere } S^2. \text{ The chart allows a reparameterization of the constrained optimization problem } f \text{ in step 1 of the Spherical Demons algorithm into an unconstrained one.}
\]
where $G_n$ is the skew-symmetric matrix defined in (4).

Once the derivatives are known, we can compute the corresponding gradients based on our choice of the metric of vector fields on $S^2$. In this work, we assume an $I_2$ inner product, so that the inner product of vector fields is equal to the sum of the inner products of the individual vectors. The inner product of individual vectors is in turn specified by the choice of the Riemannian metric on $S^2$. Assuming the canonical metric, so that the inner product of two tangent vectors is the usual inner product in the Euclidean space [39], the gradients are equal to the transpose of the derivatives (20) and (21) (see Appendix A-C). We can then rewrite (15) as a linearized least-squares objective function

$$
\hat{\zeta}^{(t)} = \arg \min_{\{\zeta_n\}} \sum_{n=1}^{N} \frac{1}{\sigma_n^2} \left( f_n(\zeta = 0) + \nabla_{\zeta_n} f_n(\zeta) \right)^2 \\
+ \frac{1}{\sigma_n^2} \sum_{n=1}^{N} \|g_n(\zeta = 0) + \nabla_{\zeta_n} g_n(\zeta)\|^2
$$

(22)

$$
= \arg \min_{\{\zeta_n\}} \sum_{n=1}^{N} \frac{1}{\sigma_n^2} \left( \begin{bmatrix} F(x_n) - M \circ \Upsilon^{(t)}(x_n) \\ -\frac{1}{\sigma_n} E_n \end{bmatrix}, \zeta_n \right)^2 \\
+ \frac{1}{\sigma_n^2} \sum_{n=1}^{N} \left\| G_n^2 S_n^T E_n \zeta_n \right\|^2
$$

(23)

$$
= \arg \min_{\{\zeta_n\}} \sum_{n=1}^{N} \left\| \begin{bmatrix} \frac{1}{\sigma_n} \left( F(x_n) - M \circ \Upsilon^{(t)}(x_n) \right) \\ 0 \\ -\frac{1}{\sigma_n} E_n \zeta_n \end{bmatrix} \right\|^2.
$$

(24)

Because of the delta function $\delta(k, n)$ in the derivatives in (20) and (21), $\zeta_n$ only appears in the $n$th term of the cost function (24). The solution of (24) can therefore be computed independently for each $\zeta_n$. Solving this linear least-squares equation yields an update rule for $\zeta_n$

$$
\hat{\zeta}_n^{(t+1)} = \frac{F(x_n) - M \circ \Upsilon^{(t)}(x_n)}{\sigma_n^2} E_n - \frac{1}{\sigma_n^2} \sum_{n=1}^{N} \left\| G_n^2 S_n^T E_n \zeta_n \right\|^2
$$

(25)

For each vertex, we only need to perform matrix-vector multiplication of up to $3 \times 3$ matrices and matrix inversion of $2 \times 2$ matrices. This implies the update rule for $\tilde{\eta}_n$

$$
\hat{\tilde{\eta}}^{(t+1)}_n = \frac{F(x_n) - M \circ \Upsilon^{(t)}(x_n)}{\sigma_n^2} E_n
$$

(26)

$$
\times \left( E_n^T \left[ \frac{1}{\sigma_n^2} \tilde{m}_n \tilde{m}_n^T + \frac{1}{\sigma_n^2} S_n \left( G_n^2 S_n^T E_n \right) \right] E_n^{-1}
$$

$$
\times E_n^T \tilde{m}_n.
$$

(27)

In practice, we use the Levenberg–Marquardt modification of Gauss–Newton optimization [49] to ensure matrix invertibility.

$$
\hat{\tilde{\eta}}^{(t+1)}_n = \frac{F(x_n) - M \circ \Upsilon^{(t)}(x_n)}{\sigma_n^2} E_n
$$

(26)

$$
\times \left( E_n^T \left[ \frac{1}{\sigma_n^2} \tilde{m}_n \tilde{m}_n^T + \frac{1}{\sigma_n^2} S_n \left( G_n^2 S_n^T E_n \right) \right] E_n^{-1}
$$

$$
\times E_n^T \tilde{m}_n.
$$

(27)

where $\hat{\tilde{\eta}}^{(t+1)}_n$ yields an update rule for $\tilde{\eta}_n$.
We can express the tangent vectors \( \tilde{\Gamma}_n \) and \( \tilde{\Upsilon}_n \) as \( E_n \Gamma_B \) and \( E_n \Upsilon_n \), respectively. Essentially, this represents \( \tilde{\Gamma}_n \) and \( \tilde{\Upsilon}_n \) in terms of the tangent space basis \( E_n \) at \( \Gamma_n \), where \( \Gamma_B \) and \( \Upsilon_n \) are the components of the tangent vectors with respect to this basis. Let \( \tilde{\Gamma} \) and \( \tilde{\Upsilon} \) be \( 2N \times 1 \) vectors corresponding to stacking \( \Gamma_n \) and \( \Upsilon_n \), respectively. The particular optimization formulated in (31) has a unique optimum [31], given by

\[
\hat{\tilde{\Gamma}} = K \left( \frac{\sigma^T}{\sigma^T J_{2N \times 2N} + K} \right)^{-1} \tilde{\Gamma}
\]

(32)

where \( K \) is a \( 2N \times 2N \) matrix consisting of \( N \times N \) blocks of \( 2 \times 2 \) matrices: the \((i, j)\) block corresponds to \( k(x_i, x_j)T_{x_i} x_{j} \). The \( 2 \times 2 \) linear transformation \( T_{x_i} x_{j} \) (\( \ast \)) parallel transports a tangent vector along the great circle from \( T_{x_i} S^2 \) to \( T_{x_j} S^2 \). \( k(x_i, x_j) \) is a nonnegative scalar function uniquely determined by the choice of the energetic norm. Typically, \( k(x_i, x_j) \) monotonically decreases as a function of the distance between \( x_i \) and \( x_j \). The proof of the uniqueness of the global optimum and the form of solution in (32) follow from the fact that the Hilbert space \( V \) is a reproducing kernel Hilbert space (RKHS), allowing the exploitation of the Riesz representation theorem [31]. This offers a wide range of choices of regularization depending on the choice of the energetic norm and the corresponding RKHS. In [31], the spherical vector spline interpolation problem was applied to landmark matching on \( S^2 \), resulting in a reasonable sized linear system of equations. Solving the matrix inversion shown in (32) is unfortunately prohibitive for cortical surfaces with more than 100,000 vertices. If one chooses a relatively wide kernel \( k(x_i, x_j) \), the system is not even sparse. Inspired by the convolution method of optimizing Step 2 in the Demons algorithm [14], [57], [66] and the convolution-based fast fluid registration in the Euclidean space [12], we propose an iterative smoothing approximation to the solution of the spherical vector spline interpolation problem.

In each smoothing iteration, for each vertex \( x_i \), tangent vectors of neighboring vertices \( x_j \) are parallel transported to \( x_i \) and linearly combined with the tangent vector at \( x_i \). The weights for the linear combination are set to \( \lambda(x_i, x_j) = 1/(1 + |N| \exp(-1/(2\gamma))) \) and \( \lambda(x_i, x_j) = \exp(-1/(2\gamma))/(1 + |N| \exp(-1/(2\gamma))) \) for \( i \neq j \), where \( |N| \) is the number of neighboring vertices of \( x_i \). Therefore, larger number of iterations \( m \) and values of \( \gamma \) results in greater smoothing of amount.

We note that the iterative smoothing approximation to spline interpolation is not exact because parallel transport is not transitive on \( S^2 \) due to the nonflat curvature of \( S^2 \) (unlike in Euclidean space), i.e., parallel transporting a tangent vector from point \( a \) to \( b \) to \( c \) results in a vector different from the result of parallel transporting a tangent vector from \( a \) to \( c \). Furthermore, the approximation accuracy degrades as the distribution of points becomes less uniform. In Appendix B, we provide a theoretical bound on the approximation error and demonstrate empirically that iterative smoothing provides a good approximation of spherical vector spline interpolation for a relatively uniform distribution of points corresponding to those of the subdivided icosahedron meshes used in this work.

E. Remarks

Algorithm 2. Spherical Demons Algorithm

Data: A fixed spherical image \( F \) and moving spherical image \( M \).

Result: Diffeomorphism \( \Gamma \) so that \( M \circ \Gamma \) is “close” to \( F \).

Set \( \Upsilon^0 = \text{identity transformation} \) (or some \textit{a priori} transformation, e.g., from a previous registration)

repeat

Step 1. Given \( \Upsilon(t) \),

\[
\text{foreach } \text{vertex } n \text{ do}
\]

Compute \( \tilde{a}_m(t) \) using (28).

\[
\text{end}
\]

Compute \( \Gamma(t) = \exp(i\tilde{\Upsilon}) \) using “scaling and squaring”.

Step 2. Given \( \Gamma(t) \),

\[
\text{foreach } \text{vertex } n \text{ do}
\]

Compute \( \tilde{a}_m(t) \) using (48) implemented via iterative smoothing.

\[
\text{end}
\]

until convergence;

The Spherical Demons algorithm is summarized in Algorithm 2.

We run the Spherical Demons algorithm in a multi-scale fashion on a subdivided icosahedral mesh. We begin from a subdivided icosahedral mesh (ic4) that contains 2562 vertices and work up to a subdivided icosahedral mesh (ic7) that contains 163 842 vertices, which is roughly equal to the number of vertices in the cortical meshes we work with. We perform 15 iterations of Step 1 and Step 2 at each level. Because of the fast convergence rate of the Gauss–Newton method, we find that 15 iterations are more than sufficient for our purposes. We also perform a rotational registration at the beginning of each multiscale level via a sectioned search of the three Euler angles.

Empirically, we find the computation time of the Spherical Demons algorithm is roughly divided equally among the four components: registration by rotation, computing the Gauss–Newton update, performing “scaling and squaring” and smoothing the vector field.

In practice, we work with spheres that are normalized to be of radius 100, because we find that at ic7, the average edge length of 1 mm corresponds to that of the original cortical surface meshes. This allows for meaningful interpretation of distances on the sphere. This requires slight modification of the equations presented previously to keep track of the radius of the sphere.

The Spherical Demons algorithm presented here registers pairs of spherical images. To incorporate a probabilistic atlas
defined by a mean image and a standard deviation image, we modify the Demons objective function in (3), as explained in Appendix C. This requires a choice of warping the subject or warping the atlas. We find that interpolating the atlas gives slightly better results, compared with interpolating the subject. However, interpolating the subject results in a runtime of under 3 min, while the runtime for interpolating the atlas is less than 5 min. In the next section, we report results for interpolating the atlas.

IV. EXPERIMENTS

We use two sets of experiments to evaluate the performance of the Spherical Demons algorithm by comparing it to the widely used and freely available FreeSurfer [27] software. The FreeSurfer registration algorithm uses the same similarity measure as Demons, but explicitly penalizes for metric and areal distortion. As we will show, even though the Spherical Demons algorithm does not specifically take into account the original metric properties of the cortical surface, we still achieve comparable if not better registration accuracy than FreeSurfer. Furthermore, FreeSurfer runtime is more than an hour while Spherical Demons runtime is less than 5 min.

There are four parameters in the algorithm. $1/\sigma^2_r$ and $\epsilon$ appear in (28). Larger values of $1/\sigma^2_r$ and $\epsilon$ decrease the size of the update taken in Step 1 of the Spherical Demons algorithm. In the experiments that follow, we set $1/\sigma^2_r = \epsilon$ and set their values such that the largest vector of the update velocity field is roughly two times the edge lengths of the mesh. The number of iterations $m$ and the weight $\exp(-1/2\gamma)$ determine the degree of smoothing. We set $\gamma = 1$ and explore a range of smoothing iterations $m$ in the following experiments.

A. Parcellation of In Vivo Cortical Surfaces

We validate Spherical Demons in the context of automatic cortical parcellation. Automatic labeling of cortical brain surfaces is important for identifying regions of interests for clinical, functional, and structural studies [20], [52]. Recent efforts have ranged from the identification of sulcal/gyral ridge lines [56], [62] to the segmentation of sulcal/gyral basins [20], [28], [38], [41], [42], [51], [52], [67]. Similar to these prior studies, we are interested in parcellation of the entire cortical surface meshes, where each vertex is assigned a label.

We consider a set of 39 left and right cortical surface models extracted from in vivo MRI [19]. Each surface is spherically parameterized and represented as a spherical image with geometric features at each vertex: mean curvature of the cortical surfaces, mean curvature of the inflated cortical surfaces and average convexity of the cortical surfaces, which roughly corresponds to sulcal depth [26]. These features are intrinsic and thus independent of the parameterization of the surface. The tools used for segmentation [19] and spherical parameterization [26] are freely available [29]. Both hemispheres of each subject were manually parcellated by a neuroanatomist into 35 labels, corresponding to the main sulci and gyri, enumerated in Table II.

We coregister all 39 spherical images of cortical geometry with Spherical Demons by iteratively building an atlas and registering the surfaces to the atlas. The atlas consists of the mean and variance of cortical geometry represented by the surface features described above. We then perform four-fold cross-validation of the parcellation of the coregistered cortical surfaces. In each iteration of cross-validation, we leave out ten subjects and use the remainder of the subjects to train a classifier [20], [28] that predicts the labels based on location and geometric features. We then apply the classifier to the hold-out set of ten subjects. We perform each iteration with a different hold-out set, i.e., subjects 1–10, 11–20, 21–30, and 31–39.

As mentioned previously, increasing the number of iterations of smoothing results in smoother warps. As discussed in [67], the choice of the tradeoff between the similarity measure and regularization is important for segmentation accuracy. Estimating the optimal registration regularization tradeoff is an active area of research [1], [18], [48], [65], [67], [68] that we do not deal with in this paper. Here, we simply repeat the above experiments using 6, 8, 10, 12, 14 iterations of smoothing. For brevity, we will focus the discussion on using 10 iterations of smoothing and comment on results obtained with the other levels of smoothing.

We repeat the above procedure of performing co-registration and cross-validation with the FreeSurfer registration algorithm [27] using the default FreeSurfer settings. Once again, we use the same features and parcellation algorithm [20], [28]. As before, the atlas consists of the mean and variance of cortical geometry.

To compare the cortical parcellation results, we compute the average Dice measure, defined as the ratio of cortical surface area with correct labels to the total surface area averaged over the test set. Because the average Dice can be misleading by suppressing small structures, we also compute the Dice measure for each structure.

<table>
<thead>
<tr>
<th>Table II</th>
<th>List of Parcellation Structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sylvian Fissure / Unknown</td>
<td>2. Bank of the Superior Temporal Sulcus</td>
</tr>
<tr>
<td>4. Caudal Middle Frontal Gyrus</td>
<td>5. Corpus Callosum</td>
</tr>
<tr>
<td>7. Entorhinal</td>
<td>8. Fusiform Gyrus</td>
</tr>
<tr>
<td>10. Inferior Temporal Gyrus</td>
<td>11. Isthmus Cingulate</td>
</tr>
<tr>
<td>16. Middle Temporal Gyrus</td>
<td>17. Parahippocampal</td>
</tr>
<tr>
<td>19. Parsopercularis</td>
<td>20. Paraventralis</td>
</tr>
<tr>
<td>22. Peri-calcarine</td>
<td>23. Post-central Gyrus</td>
</tr>
<tr>
<td>25. Pre-central Gyrus</td>
<td>26. Pre-cuneus</td>
</tr>
<tr>
<td>28. Rostral Middle Frontal</td>
<td>29. Superior Frontal Gyrus</td>
</tr>
<tr>
<td>31. Superior Temporal Gyrus</td>
<td>32. Supramarginal</td>
</tr>
<tr>
<td>34. Temporal Pole</td>
<td>35. Transverse Temporal</td>
</tr>
<tr>
<td>3. Caudal Anterior Cingulate</td>
<td>6. Cuneus</td>
</tr>
<tr>
<td>9. Inferior Parietal Complex</td>
<td>12. Lateral Occipital</td>
</tr>
<tr>
<td>15. Medial Orbito Frontal</td>
<td>18. Paracentral</td>
</tr>
<tr>
<td>27. Rostral Anterior Cingulate</td>
<td>30. Superior Parietal Complex</td>
</tr>
<tr>
<td>33. Frontal Pole</td>
<td>36. Transverse Temporal</td>
</tr>
</tbody>
</table>

YEO et al.: FAST DIFFEOMORPHIC LANDMARK-FREE SURFACE REGISTRATION 657
Furthermore, the overall difference is statistically significant for a one-sided t-test with the Dice measure of each subject treated as an independent sample ($p = 2 \times 10^{-6}$). Furthermore, the overall Dice is statistically significantly better than FreeSurfer for all levels of smoothing we considered, with the best overall dice achieved with 12 iterations of smoothing.

On the right hemisphere, FreeSurfer obtains a Dice of 88.8 and Spherical Demons achieves 89.1 with 10 iterations of smoothing. Here, the improvement is smaller, but still statistically significant ($p = 0.01$). Furthermore, the overall dice is statistically significantly better than FreeSurfer for all levels of smoothing we considered, except when 6 iterations of smoothing is used ($p = 0.06$). All results we report in the remainder of this section use 10 iterations of smoothing.

We analyze the segmentation accuracy separately for each structure. To compare Spherical Demons with FreeSurfer, we perform a one-sided paired-sampled t-test treating each subject as an independent sample and correct for multiple comparisons using a False Discovery Rate (FDR) of 0.05 [10]. On the left (right) hemisphere, the segmentations of 16 (8) structures are statistically significantly improved by Spherical Demons with respect to FreeSurfer, while no structure is significantly worse.

Fig. 3 shows the percentage improvement of individual structures over FreeSurfer. Fig. 4 displays the average Dice per structure for FreeSurfer and Spherical Demons (10 iterations of smoothing) for the left and right hemispheres. Standard errors of the mean are displayed as red bars. The numbering of the structures correspond to Table II. Parcellation improvements suggest that our registration is at least as accurate as FreeSurfer.

The structures with the worst Dice are the frontal pole 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. The frontal pole is defined by the surrounding structures, rather than by the underlying cortical geometry.

### B. Brodmann Area Localization on Ex Vivo Cortical Surfaces

Brodmann areas are cyto-architectonically defined parcellations of the cerebral cortex [13]. They can be observed through histology and more recently, through ex vivo high-resolution MRI [6]. Unfortunately, much of the cytoarchitecture cannot be observed with current in vivo imaging. Nevertheless, most studies today report their functional findings with respect to Brodmann areas, usually estimated by visual comparison of cortical folds with Brodmann’s original drawings without quantitative analysis of local accuracy. By combining histology and MRI, recent methods for creating probabilistic Brodmann area maps in the Talairach and Colin27 normalized space promise a more principled approach [2], [24], [54], [55], [71].

In this experiment, we consider a data set that contains Brodmann labels mapped to the corresponding MRI volume. Specifically, we work with postmortem histological images of ten brains created using the techniques described in [54], [71]. The histological sections were aligned to postmortem MR with nonlinear warps to build a 3-D histological volume. These volumes were segmented to separate white matter from other tissue classes, and the segmentation was used to generate topologically correct and geometrically accurate surface representations of the cerebral cortex using FreeSurfer [19]. The eight manually labeled Brodmann area maps (areas 2, 4a, 4p, 6, 44, 45, 17, and 18) were sampled onto the surface representations of each hemisphere, and errors in this sampling were manually corrected (e.g., when a label was erroneously assigned to both banks of a sulcus). A morphological close was then performed on each label to remove small holes. We note that Brodmann areas 4a, 4p, and 6 were mapped in only eight of the ten subjects. Fig. 5 shows these eight Brodmann areas on the resulting cortical representations for two subjects. Finally, we map the folding patterns and the Brodmann area labels onto a spherical coordinate system [27].

It has been shown that nonlinear surface registration of cortical folds can significantly improve Brodmann area overlap across different subjects [25], [68] compared with volumetric registration. Registering the ex vivo surfaces is more difficult than in vivo surfaces because the reconstructed volumes are extremely noisy due to the distortions introduced by the histology, resulting in noisy geometric features, as shown in Fig. 6.

We consider two strategies for aligning Brodmann areas. For both strategies, we will use 10 iterations of smoothing for Spherical Demons as it proved reasonable in the previous set of experiments. The first strategy involves coregistering the 10 ex vivo surfaces using cortical geometry by repeatedly building an atlas and registering the surfaces to the atlas, similar to the previous experiment on cortical parcellation. We use either Spherical Demons or FreeSurfer for registration. We refer to the coregistration using Spherical Demons and FreeSurfer as SD10 and FS10, respectively, (10 refers to the number of subjects in the study, not the number of smoothing iterations).

The second strategy involves registering the 10 ex vivo surfaces to the in vivo “Buckner40” atlas, constructed from 40 in...
**vivo** subjects, that is distributed with the FreeSurfer software. Once again, we use either Spherical Demons or FreeSurfer for the registration. We refer to the co-registration using Spherical Demons and FreeSurfer as SD40 and FS40, respectively.

To measure the quality of alignment of the Brodmann areas after cortical registration, we use an adaptation of the modified Hausdorff distance [21]. For each pair of registered subjects, we project each Brodmann area from the first subject onto the second subject and compute the mean distance between the boundaries, measured on the original cortical surface of the second subject. We obtain a second measurement by projecting each Brodmann area from the second subject onto the first subject. Since we have 10 surfaces, we get 90 ordered pairs and 90 alignment errors for each Brodmann area.

Table III reports the mean alignment errors for each Brodmann area and for each method. The lowest errors for each Brodmann area are shown in **bold**. We see that for almost all Brodmann areas, the best alignment come from SD10 or SD40. Similarly, Fig. 7 shows the median alignment error for each Brodmann area. The error bars indicate the lower and upper quartile alignment errors.

We use permutation testing to evaluate statistical significance of the results. We cannot use the t-test because the 90 alignment errors are correlated—since the subjects are coregistered together, good alignment between subjects 1 and 2 and between subjects 2 and 3 necessarily implies a higher likelihood of good alignment between subjects 1 and 3.

The tests show that SD10 is better than FS10 and SD40 is slightly better than FS40. SD10 and SD40 are comparable. Compared with FS10, SD10 improves the median alignment errors of 5 (4) Brodmann areas on the left (right) hemisphere (FDR = 0.05) and no structure gets worse. Compared with FS40, SD40 statistically improves the alignment of 2 (1) Brodmann areas on the left (right) hemisphere (FDR = 0.05) with no structure getting worse. Permutation tests on the mean alignment errors show similar results, except that FS40 per-
forms better than SD40 for BA4p on the right hemisphere when using the mean statistic. These results suggest that the Spherical Demons algorithm is at least as accurate as FreeSurfer in aligning cortical folds and Brodmann areas.

V. DISCUSSION

The Demons algorithms [57], [66] discussed in Section II and the Spherical Demons algorithm proposed in this paper use a regularization term that modulates the final deformation. Motivated by [12], [14], the Diffeomorphic Demons algorithm [66] admits a fluid prior on the velocity fields. This involves smoothing the velocity field updates before computing the exponential map to obtain the displacement updates exp(\(\tilde{v}(t)\)).

An alternative interpretation of the smoothing implementation of Christensen’s algorithm comes from choosing a different metric for computing the gradient from the derivatives [9]. The choice of the metric also arises in our problem when computing the gradient as discussed in Appendix A-C. This suggests that the Spherical Demons algorithm can incorporate a fluid prior by modifying the Gauss–Newton update step (28). Unfortunately, this process introduces coupling among the vertices resulting in the loss of the speed-up advantage of Spherical Demons (see for example the derivations of [34]). The exploration of the performance of the different fluid prior implementations is outside the scope of this paper.

Because the tools of differential geometry are general, the Spherical Demons algorithm can be in principle extended to arbitrary manifolds, besides the sphere. One challenge is that the definition of coordinate charts for an arbitrary manifold is more difficult than that for the sphere. Approaches of utilizing the embedding space [15] or the intrinsic properties of manifolds [40] are promising avenues of future work.

VI. CONCLUSION

In this paper, we presented the fast Spherical Demons algorithm for registering spherical images. We showed that the two-step optimization of the Demons algorithm can also be applied on the sphere. By utilizing the one parameter subgroups of diffeomorphisms, the resulting deformation is invertible. We tested the algorithm extensively in two different applications and showed that the accuracy of the algorithm compares favorably with the widely used FreeSurfer registration algorithm [27] while offering more than one order of magnitude speedup. Both
and the normal vector to the sphere is the barycentric
intersection point and we denote the 3 image is made continuous by barycentric
interpolation. We also compute the gradient at the point
Let \( v_i \) denote the vertices of the triangle
containing \( p(x) \) and \( \vec{n} \) denote the \( 3 \times 1 \) normal vector to the triangle. Since \( p(x) = \alpha x \) for some \( \alpha \) and \( (p(x) - p_1, \vec{n}) = 0 \), we can write

\[
M_{s}(x) = I(p(x))
\]  

where \( p(x) \) is the intersection point and \( I(\cdot) \) is the barycentric interpolation. Let \( p_1, p_2, p_3 \) denote the vertices of the triangle containing \( p(x) \) and \( \vec{n} \) denote the \( 3 \times 1 \) normal vector to the triangle. Since \( p(x) = \alpha x \) for some \( \alpha \) and \( (p(x) - p_1, \vec{n}) = 0 \), we can write

\[
p(x) = \frac{\langle p_1, \vec{n} \rangle}{\langle x, \vec{n} \rangle} x
\]  

and

\[
I(p) = \frac{A_1(p)M_s(p_1) + A_2(p)M_s(p_2) + A_3(p)M_s(p_3)}{A}
\]  

A. Computing Spatial Gradient of \( M \circ \gamma^T(\cdot) \)

In this appendix, we discuss the computation of \( M \circ \gamma^T(\cdot) \), the spatial gradient of the warped moving image \( M \circ \gamma^T(\cdot) \) at the point \( x \). We can think of \( M \circ \gamma^T(\cdot) \) as an image \( M_{s} \triangleq M \circ \gamma^T \) defined on the mesh vertices \( \{x_n\} \). This image is made continuous by the choice of an interpolation method. In this work, we assume that we are working with a triangular mesh. To evaluate \( M_{s} \) at a point \( x \in S^2 \), we first find the triangle that contains the intersection between the vector representing the point \( x \) (i.e., the vector between the center and the point \( x \) of the sphere) and the mesh. The image value at \( x \) is then given by the barycentric interpolation of the image values at the intersection point. Mathematically, we can write

\[
M_{s}(x) = I(p(x))
\]  

where \( p(x) \) is the intersection point and \( I(\cdot) \) is the barycentric interpolation. Let \( p_1, p_2, p_3 \) denote the vertices of the triangle containing \( p(x) \) and \( \vec{n} \) denote the \( 3 \times 1 \) normal vector to the triangle. Since \( p(x) = \alpha x \) for some \( \alpha \) and \( (p(x) - p_1, \vec{n}) = 0 \), we can write

\[
p(x) = \frac{\langle p_1, \vec{n} \rangle}{\langle x, \vec{n} \rangle} x
\]  

and

\[
I(p) = \frac{A_1(p)M_s(p_1) + A_2(p)M_s(p_2) + A_3(p)M_s(p_3)}{A}
\]  

Matlab and ITK versions of the Spherical Demons algorithm are publicly available.2

A clear future challenge is to take into account the original metric properties of the cortical surface in the registration process, as demonstrated in previously proposed registration methods [27], [59].

We note that while fast algorithms are useful for deploying the developed tool on large datasets, they can further allow for complex applications that were previously computationally intractable. For example, we have incorporated the ideas behind Spherical Demons into a meta-registration framework that learns registration cost functions which are optimal for specific applications [68].

APPENDIX A

STEP 1 GRADIENT DERIVATION

In this appendix, we provide details on the computation of the spatial gradient of the warped moving image \( M \circ \gamma^T(\cdot) \) and the Jacobian of the deformation \( \gamma^T(\cdot) \). We also compute the gradients of the demons cost function using the derivatives computed in (20) and (21), assuming the \( L^2 \) inner product space for vector fields and the canonical metric.

2The Matlab code was used for this paper. The ITK code is still preliminary. Please check website http://www.sites.google.com/site/yeoyeo02/software/sphericaldemonsrelease for updates.

<table>
<thead>
<tr>
<th>Left Hemisphere</th>
<th>Right Hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>BA4a</td>
</tr>
<tr>
<td>FS10</td>
<td>3.8</td>
</tr>
<tr>
<td>FS40</td>
<td>2.9</td>
</tr>
<tr>
<td>SD10</td>
<td>3.1</td>
</tr>
<tr>
<td>SD40</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Table III

MEAN ALIGNMENT ERRORS OF BRODMANN AREAS IN mm FOR THE FOUR REGISTRATION METHODS. LOWEST ERRORS ARE SHOWN IN bold. SD REFERS TO SPHERICAL DEMONS; FS REFERS TO FREESURFER

Fig. 7. Median alignment errors of Brodmann areas in mm for the four registration methods. The error bars indicate the upper and lower quartile alignment errors.  "#" indicates that the median errors of SD10 are statistically lower than those of FS10 (FDR = 0.05). "a" indicates SD40 outperforms FS40. For no Brodmann area does FreeSurfer outperform Spherical Demons. (a) Left hemisphere and (b) right hemisphere.
where $A_1(p)$, $A_2(p)$ and $A_3(p)$ are the areas of the triangles $\triangle pp_1p_3$, $\triangle pp_1p_2$, and $\triangle pp_2p_3$, respectively. Note that $A = A_1(p) + A_2(p) + A_3(p)$. $M_1(p_1)$, $M_2(p_2)$, and $M_3(p_3)$ are the image values at the mesh vertices $p_1$, $p_2$, and $p_3$, respectively.

Computing the derivative of the image value at $x$ follows easily from the chain rule

$$\frac{\partial p(x)}{\partial x} = \frac{\langle y_1, \hat{n} \rangle}{\langle x, \hat{n} \rangle} I_{3 \times 3} - \frac{\langle y_1, \hat{n} \rangle}{\langle x, \hat{n} \rangle^2} x x^T$$

(36)

$$\frac{\partial p}{\partial y} = \nabla_p A_1(p) M_1(p_1) + \nabla_p A_2(p) M_2(p_2) + \nabla_p A_3(p) M_3(p_3)$$

(37)

where $\nabla_p A_i(p)$ is the derivative of the triangle area $A_i$ for example, $\nabla_p A_1(p)$ is a $1 \times 3$ vector in the plane of the triangle $pp_1p_3$, perpendicular and pointing to the edge $p_2p_3$, with magnitude half the length of $p_2p_3$. Combining (36) and (37) gives the spatial gradient of the warped moving image.

A complication arises when $x$ corresponds to one of the mesh vertices, since the spatial gradient is not defined in this case. The same problem arises in Euclidean space with linear interpolation and the spatial gradient is typically defined via finite central difference. It is unclear what the equivalent definition on a mesh is. Here, for a mesh vertex $x$, we compute the spatial gradient for each of the surrounding triangles and linearly combine the spatial gradients using weights corresponding to the areas of the triangles.

B. Computing the Jacobian of Deformation $\Psi^{(t)}$

In this Appendix, we discuss the computation of $S_{\Psi}^T$, the Jacobian of the deformation $\Psi^{(t)}$ at $x_n$. We can think of $\Psi^{(t)}$ as a vector function on $S^2$ that maps each mesh vertex $x_n$ to a new point on the sphere. This vector image is made continuous by the choice of an interpolation method. We use the same interpolation as in Appendix A-A, except we need to normalize the barycentric interpolation so that the interpolated point is constrained to be on the sphere

$$\Psi^{(t)}(x) = I(p(x))$$

(38)

where $p(x)$ is the same as in the previous section and

$$I(p) = \frac{A_1(p)\Psi^{(t)}(p_1) + A_2(p)\Psi^{(t)}(p_2) + A_3(p)\Psi^{(t)}(p_3)}{[A_1(p)\Psi^{(t)}(p_1) + A_2(p)\Psi^{(t)}(p_2) + A_3(p)\Psi^{(t)}(p_3)]}.$$

The Jacobian is computed via chain rule, just like in the previous section.

C. Computing the Gradients From the Derivatives

In this Appendix, we seek to compute the gradients of $f_n(\hat{z}) \triangleq F(x_n) - M \circ \{\Psi^{(t)} \circ \exp(\{E_n\hat{z}_n\})\}(x_n)$ and $g_n(\hat{z}) \triangleq \Psi^{(t)} + G_n \circ \{\Psi^{(t)} \circ \exp(\{E_n\hat{z}_n\})\}(x_n)$, assuming a $l_2$ inner product for vector fields and the canonical metric $R$ for $S^2$. These assumptions imply that the inner product of two vector fields $\hat{z}_1 \{E_k\hat{z}_k\}$ and $\hat{z}_2 \{E_k\hat{z}_k\}$ are given by

$$\langle \hat{z}_1, \hat{z}_2 \rangle_{l_2} = \langle \{E_k\hat{z}_k\}, \{E_k\hat{z}_k\} \rangle_{l_2}$$

(40)

where

- (40) follows from the equivalence of the tangent bundles $T^\mathbb{R}^2$ and $T^\mathbb{S}^2$ induced by the coordinate charts $\{\Psi_n\}$.
- (41) is the result of the $l_2$ assumption that the inner product of vector fields is given by the sum of the inner products of individual vectors.
- Because we assume the canonical metric, each term in the inner product in (41) is simply the usual inner product between $3 \times 1$ vectors $E_k\hat{z}_k$ and $E_l\hat{z}_l$. Since the columns of $E_k$ are orthonormal with respect to the usual inner product and using linearity of the inner product, (41) implies (42), i.e., the inner product $\langle \hat{z}_1, \hat{z}_2 \rangle_{l_2}$ can be computed by the sum of the usual inner product between $2 \times 1$ tangent vectors $\hat{z}_1$ and $\hat{z}_2$.

Let $d_n(\hat{z})$ be the directional derivative of $f_n$ for any $\hat{z} \{\hat{z}_k\}$. The directional derivative is independent of the choice of metric. Since the derivative of $d_n(\hat{z})$ with respect to $\hat{z}_k$ is a $2 \times 1$ vector $-R_{n,k} E_n \delta(k, n)$ (20), we get

$$d_n(\hat{z}) = -R_{n,k} E_n \delta(k, n).$$

(43)

Recall that the gradient $\nabla_{l_2} f_n(\hat{z})$ is defined to be a tangent vector field such that $d_n(\hat{z}) = \langle \nabla_{l_2} f_n, \hat{z} \rangle_{l_2}$ for any $\hat{z} \{\hat{z}_k\}$. The gradient is therefore dependent on the choice of the inner product. From (42) and (43), we can write

$$-R_{n,k} E_n \delta(k, n) = d_n(\hat{z})$$

(44)

$$= \langle \nabla_{l_2} f_n, \hat{z} \rangle_{l_2}$$

(45)

$$= \sum_{k=1}^N \langle \nabla_{l_2} f_n(x_n), z_k \rangle_{l_2}$$

(46)

Therefore, the gradient $\nabla_{l_2} f_n$ can be written as a $2N \times 1$ vector consisting of $N$ blocks of $2 \times 1$ vectors, where all the blocks are zeros, except the $n$th block is equal to $-R_{n,k} E_n \delta(k, n)$.

Similarly, we denote the gradient of $g_n(\hat{z})$ as $\nabla_{l_2} g_n(j)$ for $j = 1, 2, 3$ corresponding to the three output components of $g_n(\hat{z})$. The derivative of $g_n$ with respect to $\hat{z}_k$ is a $3 \times 2$ matrix $G_n^T S_{\Psi}^T E_n \delta(k, n) \triangleq [\partial_{g_{n1}}, \partial_{g_{n2}}, \partial_{g_{n3}}] \delta(k, n)$ (21), where $G_n^T E_n \delta(k, n)$ is a $1 \times 2$ vector corresponding to the derivative of the $j$th component of $g_n$ with respect to $\hat{z}_k$. Using the same derivation as before, we can show that $\nabla_{l_2} g_n(j)$ can be written as a $2N \times 1$ vector consisting of $N$ blocks of $2 \times 1$ vectors, where all the blocks are zeros, except the $n$th block is equal to $\partial_{g_{nj}}$.

APPENDIX B

APPROXIMATING SPLINE INTERPOLATION WITH ITERATIVE SMOOTHING

In this Appendix, we demonstrate empirically that iterative smoothing provides a good approximation of spherical vector spline interpolation for a relatively uniform distribution of points corresponding to those of the subdivided icosahedron...
and the relatively uniform distribution of points and a subdivided icosahedron mesh with blocks of is the Frobenius norm. and each block corresponds to a matrix consisting of , such that we seek .

\[ K \left( \frac{\sigma^2}{\sigma_T^2} I_{2N \times 2N} + K \right)^{-1} (K')^m \approx 0. \]

Rearranging the terms, we get

\[ (I_{2N \times 2N} - (K')^m)^{-1} (K')^m \frac{\sigma^2}{\sigma_T^2} \approx K. \]

To make the “≈” concrete, we optimize for

\[ k^* = \arg \min_k \left\| K - (I_{2N \times 2N} - (K')^m)^{-1} (K')^m \frac{\sigma^2}{\sigma_T^2} \right\|_F. \]

where \( \| \cdot \|_F \) is the Frobenius norm. The cost function (51) can be optimized component-wise, i.e., we can solve for \( k^*(x_i, x_j) \) for each pair \( x_i, x_j \). For \( \gamma = 1, m = 10 \) and a subdivided icosahedron mesh with 642 vertices, we plot the resulting \( k^*(x_i, x_j) \) as a function of the geodesic distance between \( x_i \) and \( x_j \) in Fig. 8.

2) In the second stage, we perform a least-squares fit of a b-spline function to the estimated \( k^*(x_i, x_j) \) to obtain the final estimate of \( \tilde{k}(x_i, x_j) \). Fig. 8 illustrates an example kernel \( \tilde{k}(x_i, x_j) \) we obtain (c.f., the kernel illustrated in [31]). We note that an alternative to b-spline interpolation is to fit the coefficients of the general kernel function suggested in Appendix A of [31]. This will guarantee that the estimated kernel corresponds to an energetic norm. We leave exploring this direction to future work.

B. Evaluating Approximation

We now investigate the quality of the estimate \( \tilde{k}(x_i, x_j) \) by computing

\[ \left\| K - (I_{2N \times 2N} - (K')^m)^{-1} (K')^m \frac{\sigma^2}{\sigma_T^2} \right\|_{l_2}^2 \]

where \( \| \cdot \|_{l_2} \) is the \( l_2 \) matrix operator norm. The difference metric (52) measures the maximum \( l_2 \) difference between smoothed vector fields obtained from iterative smoothing and spherical vector spline interpolation for any possible input vector field \( \{ \hat{\Gamma}_n \} \) of unit \( l_2 \) norm, i.e., \( \sum_n \| \hat{\Gamma}_n \|_{l_2}^2 = 1 \). We note that \( \tilde{k}(x_i, x_j) \) can be in principle estimated by minimizing (52) instead of the proposed two-stage process. However, the optimization is difficult since evaluating the cost function itself requires finding the largest singular value of a large, nonsparse matrix.

Fig. 9 displays the difference metric we obtained with different values of \( \gamma \) and iterations \( m \) for meshes ic2, ic3, ic4, and ic5. Each of the meshes is obtained from recursively subdividing a lower resolution mesh: ic2 indicates that the mesh was obtained from subdividing an icosahedron mesh twice. The number of vertices quadruples with each subdivision, so that ic5 corresponds to 10,242 vertices.

We conclude from the figure that the differences between the two smoothing methods are relatively small and increase with mesh resolution. As discussed in the next section, we run...
Spherical Demons on different mesh resolutions, including ic7. Unfortunately, because of the large non-sparse matrices we are dealing with, we were only able to compute the differences up to ic5. Computing the difference metric for ic5 took an entire week on a machine with 128 GB of RAM. However, the plots in Fig. 9 indicate that the differences appear to have converged by ic5.

To better understand the incurred differences, Fig. 10 illustrates the outputs and differences of the two smoothing methods for different inputs on ic4. The first row illustrates an input vector field which is zero everywhere except for a single tangent vector of unit norm. The results of spline interpolation and iterative smoothing correspond to our intuition that smoothing a single tangent vector propagates tangent vectors of smaller magnitudes to the surrounding areas. The two methods also produce almost identical results as shown by the clean difference image in the fourth column.

The second row of Fig. 10 demonstrates the worst unit norm input vector field as measured by the difference metric (52). This worst unit norm input vector field corresponds to the largest eigenvector in (52). The pattern of large differences correspond to the original 12 vertices of the uniform icosahedron mesh. These original 12 vertices are the only vertices in the subdivided icosahedron meshes with five, instead of six neighbors, as shown by the pentagon pattern. The fact that these 12 ver-
tics are local maxima of differences suggest that these vertices are treated differently by the two smoothing techniques.

The last row of Fig. 10 demonstrates an input vector field that represents the deformation of an actual registration performed in Section IV. The norm of the input vector field is 700 times that in the first two rows, but the discrepancies between spline interpolation and iterative smoothing are less than expected. The differences of 90% of the vectors are less than 0.2 mm, with larger differences in the neighborhoods of the 12 vertices identified previously. Since we conclude previously that the differences are local maxima of differences suggest that these vertices are treated differently by the two smoothing techniques.

We should emphasize that the discrepancies between spline interpolation and iterative smoothing do not necessarily imply registration errors. The differences only indicate the deviations of the deformations from true local optima of the Demons registration cost function (3) assuming the estimated kernel. Approximating smoothing kernels by iterative smoothing is an active area of research in medical imaging [17], [32]. Future work would involve understanding the interaction between the number of smoothing iterations \( n \) and the choice of the weights \( \exp(-1/2\gamma) \) on the quality of the spherical registration.

**APPENDIX C**

**ATLAS-BASED SPHERICAL DEMONS**

In this Appendix, we demonstrate how an atlas consisting of a mean image and a standard deviation image can be incorporated into the Spherical Demons algorithm. The standard deviation image replaces \( \Sigma \) in (3). We first discuss a probabilistic interpretation of the Demons objective function and its relationship to atlases. We then discuss the optimization of the resulting probabilistic objective function.

**A. Probabilistic Demons Objective Function**

The Demons objective function reviewed in Section II is defined for the pairwise registration of images. To incorporate a probabilistic atlas, we now reformulate the objective function. Consider the following maximum a posteriori objective function:

\[
(\hat{\Gamma}, \hat{\Gamma}^*) = \arg \max_{\hat{\Gamma}} \log p(\Gamma, \mathcal{Y} | F, M) \tag{53}
\]

\[
\quad = \arg \max_{\hat{\Gamma}} \log p(F, M | \Gamma, \mathcal{Y}) p(\Gamma | \mathcal{Y}) p(\mathcal{Y}) \tag{54}
\]

\[
\quad = \arg \max_{\hat{\Gamma}} \log p(F, M \circ \Gamma | \mathcal{Y}) + \log p(\Gamma | \mathcal{Y}) + \log p(\mathcal{Y}). \tag{55}
\]

Assuming a Gaussian noise model, we define

\[
p(F, M \circ \Gamma) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi} \sigma^2(\Gamma, x_n)} \exp \left( -\frac{(F(x_n) - M \circ \Gamma(x_n))^2}{2\sigma^2(\Gamma, x_n)} \right) \tag{56}
\]

\[
\log p(\Gamma | \mathcal{Y}) = \frac{1}{Z(\sigma^2_\Gamma)} \exp \left( -\frac{1}{\sigma^2_\Gamma} \sum_{n=1}^{N} \| \hat{\mathcal{Y}}_n - \mathcal{Y}_n \|^2 \right) \tag{57}
\]

The probabilistic objective function in (55) becomes

\[
p(\mathcal{Y}) = \frac{1}{Z(\sigma^2_\Gamma)} \exp \left( -\frac{1}{\sigma^2_\Gamma} \text{Reg}(\mathcal{Y}) \right) \tag{58}
\]

where \( \text{Reg}(\mathcal{Y}) \) is defined via the energetic norm as discussed in Section III-C and for reasons that will soon be clear, we are being purposefully agnostic about the form of \( \sigma(\Gamma, x_n) \). The objective function in (55) becomes

\[
(\hat{\Gamma}, \hat{\Gamma}^*) = \arg \min_{\Gamma, \mathcal{Y}} \sum_{n=1}^{N} \frac{(F(x_n) - M \circ \Gamma(x_n))^2}{2\sigma^2(\Gamma, x_n)} \tag{59}
\]

\[
+ \frac{1}{\sigma^2_\Gamma} \sum_{n=1}^{N} \| \hat{\mathcal{Y}}_n - \mathcal{Y}_n \|^2 + \frac{1}{\sigma^2_\Gamma} \text{Reg}(\mathcal{Y}) + \sum_{n=1}^{N} \log \sigma(\Gamma, x_n) \tag{59}
\]

which is the instantiation of the Demons objective function (3), except for the extra term \( \sum_{n=1}^{N} \log \sigma(\Gamma, x_n) \). Note that we have omitted the partition functions \( Z(\sigma^2_\Gamma) \) and \( Z(\sigma^2_\Gamma) \) because \( \sigma_\Gamma \) and \( \sigma_\Gamma \) are constant with respect to the deformations \( \Gamma \) and \( \mathcal{Y} \).

In this probabilistic interpretation, the two regularization terms \( p(\Gamma | \mathcal{Y}) \) and \( p(\mathcal{Y}) \) act as a hierarchical prior on \( \Gamma \), with the hidden transformation \( \mathcal{Y} \) as a hyperparameter.

As before, \( \sigma(\Gamma, x_n) \) is the standard deviation of the intensity at vertex \( n \). Given a set of co-registered images, we can create an atlas by computing the mean intensity and standard deviation at each vertex. To incorporate the atlas, we need to make the choice of treating the atlas as the fixed or moving image. If we treat the atlas as the fixed image, then we set \( F \) to be the mean image and \( \sigma \) to be the standard deviation. In this case, we do not need to interpolate the mean or standard deviation image. Consequently, \( \sigma(\Gamma, x_n) = \sigma(x_n) \) and \( \log \sigma(\Gamma, x_n) \) can be omitted from the optimization. The registration becomes identical to the Spherical Demons algorithm for two images.

However, recent work [1], [53] suggests that treating the atlas as a moving image might be more correct theoretically. This involves setting the moving image \( M \) to be the mean image. In this case, \( \sigma(\Gamma, x_n) = \sigma(\Gamma(x_n)) \) is a function of \( \Gamma \) and we must include \( \log(\sigma(\Gamma(x_n))) \) in the optimization. We performed experiments for both choices and found the results from interpolating the atlas, i.e., treating it as a moving image, to be only slightly better than interpolating the subject. However, interpolating the subject results in a faster algorithm, whose computational time is less than 3 min. We report the results of interpolating the atlas in the experimental section.

**B. Optimization of Atlas-Based Spherical Demons**

We now discuss the optimization in (59). Note that the introduction of the new term \( \sum_{n=1}^{N} \log \sigma(\Gamma, x_n) \) only affects Step 1 of the Spherical Demons algorithm. By parameterizing \( \Gamma(t) = \mathcal{Y}(t) \circ \exp(\{ E_n x_n \}) \), we get

\[
\left\{ x_n \right\} = \arg \min_{x_n} \sum_{n=1}^{N} \frac{(F(x_n) - M \circ \mathcal{Y}(t) \circ \exp(\{ E_n x_n \}))^2}{2\sigma^2(t)} \tag{60}
\]

\[
+ \frac{1}{\sigma^2_\Gamma} \sum_{n=1}^{N} \| \mathcal{Y}(t) + G_n \left\{ \gamma(t) \circ \exp(\{ E_n x_n \}) \right\}(x_n) \|^2 \tag{61}
\]
Let the current estimate of \( x(t) \) is stationary in the case of the one and second terms and estimate their second derivatives using the Gauss–Newton method. The difficulty lies in the third term, which is not quadratic and is even strictly concave, so we have to make further approximations.

Consider the problem of optimizing a one-dimensional function \( f(x) \). Let the current estimate of \( x \) be \( x_0 \). Newton’s optimization [49] involves the following update:

\[
\Delta x = -(f''(x_0))^{-1} f'(x_0)
\]

where \( f'(x_0) \) and \( f''(x_0) \) are the gradient and the Hessian of \( f \) evaluated at \( x_0 \) respectively. When \( f''(x) \) is negative (positive), the update \( \Delta x \) increases (decreases) the objective function, regardless of whether one is attempting to increase or decrease the objective function! The Gauss–Newton approximation of the Hessian for minimizing nonlinear quadratic functions actually stabilizes the Newton’s method by ensuring the estimated Hessian is positive.

To optimize \( f(x) \) with Newton’s method, we need to compute the gradient and the Hessian. Because we are using the \( L_2 \) inner product and the canonical metric (see Appendix A–C), the gradient and the Hessian correspond to the first and second derivatives. The first derivative or gradient corresponds to

\[
\frac{\partial f}{\partial x_k} = 2f_n1(x) \frac{\partial f_n1}{\partial x_k} + 2f_n2(x) \frac{\partial f_n2}{\partial x_k} + \frac{1}{f_n3(x)} \frac{\partial f_n3}{\partial x_k}
\]

and the second derivative corresponds to

\[
\frac{\partial^2 f}{\partial x_k^2} = 2 \left( \frac{\partial f_n1}{\partial x_k} \right)^2 + 2f_n1(x) \frac{\partial^2 f_n1}{\partial x_k^2} + 2 \left( \frac{\partial f_n2}{\partial x_k} \right)^2 + 2f_n2(x) \frac{\partial^2 f_n2}{\partial x_k^2} - \frac{\partial f_n3}{\partial x_k} \frac{\partial^2 f_n3}{\partial x_k^2} + \frac{1}{f_n3(x)} \frac{\partial^2 f_n3}{\partial x_k^2}
\]

\[
\approx 2 \left( \frac{\partial f_n1}{\partial x_k} \right)^2 + 2 \left( \frac{\partial f_n2}{\partial x_k} \right)^2 - \left( \frac{\partial f_n3}{\partial x_k} \right)^2
\]

where the last approximation was made using the Gauss–Newton method. Not surprisingly, the third term corresponding to \( \log \) is negative, which can introduce instability in the Gauss–Newton update. Consequently, we drop the last term, resulting in

\[
\frac{\partial^2 f}{\partial x_k^2} \approx 2 \left( \frac{\partial f_n1}{\partial x_k} \right)^2 + 2 \left( \frac{\partial f_n2}{\partial x_k} \right)^2.
\]

Note that the resulting update (62) is always in the direction of descent since the estimated second derivative is always positive. Theoretically, it is necessary to do a line search along the Gauss–Newton update direction to ensure convergence. In practice, we find that the objective function decreases reliably for each full Newton’s step.

**APPENDIX D**

**NUMERICS OF DIFFEOMORPHISM**

While \( v \) and \( \Phi_v(x) = \exp(v(x)) \) are technically defined on the entire continuous image domain, in practice, \( v \) and \( u \) are represented by vector fields defined on a discrete set of points in the image, such as at each pixel [57], [66] or control points [4], [9] or in our case, vertices of a spherical mesh. From the theories of ODEs [11], we know that the integral curves or trajectories \( u(t) = \Phi_u(x(t)) \) of a velocity field \( v(x, t) \) exist and are unique if \( v(x, t) \) is Lipschitz continuous in \( x \) and continuous in \( t \). This is true in both Euclidean spaces and on manifolds. Uniqueness means that the trajectories do not cross, implying that the deformation is invertible. Furthermore, we know from the theories of ODEs that a \( C^r \) continuous velocity field \( v \) produces a \( C^r \) continuous deformation field. Therefore, a sufficiently smooth velocity field results in a diffeomorphic transformation.

Since the velocity field \( v \) is stationary in the case of the one parameter subgroup of diffeomorphisms, \( u \) is clearly continuous (and in fact \( C^\infty \)) in \( t \). A smooth interpolation of \( v \) is continuous in the spatial domain and is Lipschitz continuous if we consider a compact domain, which holds since we only consider images that are closed and bounded.

To compute the final deformation of an image, we have to estimate \( \exp(v) \) at least at the set of image grid points. We can compute \( \exp(v) \) by numerically integrating the smoothly interpolated velocity field \( v \) with Euler integration. In this case, the estimate becomes arbitrarily close to the true \( \exp(v) \) as the number of integration time steps increases. With a sufficiently large number of integration steps, we expect the estimate to be invertible and the resulting transformation to be diffeomorphic.

The parameterization of diffeomorphisms by a stationary velocity field is popular due to the “scaling and squaring” approach [3] for computing \( \exp(v) \). Instead of Euler integration, the “scaling and squaring” approach iteratively composes displacement fields. Because we are working on the sphere \( S^2 \), the “scaling and squaring” procedure discussed in [3] has to be slightly modified

\[
\Phi^n_{\frac{1}{2^K}} v(x) = \left\{ \Psi_n \left( \Phi^n_{\frac{1}{2^K}} v(x_n) \right) \right\}_{n=1,...,N}
\]

\[
\Phi^n_{\frac{1}{2^K}} v(x) = \Phi^n_{\frac{1}{2^K}} v \left( \Phi^n_{\frac{1}{2^K}} v(x) \right)
\]

\[
\Phi^n_{\frac{1}{2^K}} v(x) = \Phi^n_{\frac{1}{2^K}} v \left( \Phi^n_{\frac{1}{2^K}} v(x) \right)
\]

where \( \Psi_n \) is the local coordinate chart defined in (10), such that \( \Psi_n(0) = x_n \). Equation (67) differs from “scaling and squaring” in Euclidean space. \( E_n(1/2^K)v(x_n) \) is the velocity vector at the origin of \( R^2 \) corresponding to the velocity vector \( (1/2^K)v(x_n) \) at \( x_n \). For large enough \( K \), we can approximate a particle at the origin to move to position \( E_n(1/2^K)v(x_n) \) via the flow of \( E_n(1/2^K)v(x_n) \). Finally, the coordinate chart \( \Psi_n \) maps the point \( E_n(1/2^K)v(x_n) \) back to the sphere. The correctness of this process follows from the fact that the solution trajectories
of the ODEs of a vector field can be consistently transformed via the coordinate charts.

While “scaling and squaring” converges to the true answer as $K$ approaches $\infty$ in the continuous case, in the discrete case, composition of the displacement fields requires interpolation of displacement fields, introducing errors in the process. In particular, suppose $\Phi_{t_0}(x)$ and $\Phi_{2t_0}(x)$ are the true trajectories found by performing an accurate Euler integration up to time $t_0$ and $2t_0$, respectively. Then, there does not exist a trivial interpolation scheme, so that $\Phi_{2t_0}(x) = \Phi_{t_0}(\Phi_{t_0}(x))$. In practice however, it is widely reported that in $\mathbb{R}^2$ and $\mathbb{R}^3$, “scaling and squaring” tends to preserve invertibility even with rather large deformations [4], [66].

As discussed in Appendix A-B, we employ barycentric interpolation, followed by normalization to ensure the warp stays on the unit sphere. In practice, we find that the resulting transformation is indeed diffeomorphic. Technically speaking, since we use linear interpolation for the displacement field, the transformation is indeed diffeomorphic. Technically speaking, since we use linear interpolation for the displacement field, the transformation is indeed diffeomorphic. This is because the transformation is continuous, but not differentiable across mesh edges. However, we follow the convention of [3], [4], [66] who call their transformation diffeomorphic even though they are homeomorphic.

ACKNOWLEDGMENT

The authors would like to thank H. Mohlberg, K. Amunts, and K. Zilles for providing the histological dataset. They also thank X. Pennec, W. Ou, and S. Balci for helpful discussions and T. Riklin Raviv for reading the conference draft of this paper. They also thank L. Ibanez and M. Audette for reading the journal draft of this paper and for their help in implementing the ITK version of the Spherical Demons algorithm. Finally, the authors thank the reviewers for their many helpful suggestions on improving and clarifying the paper.

REFERENCES

[29] [Online]. Available: http://surfer.nmr.mgh.harvard.edu/fswiki