MICCAI 2006 Workshop Proceedings

MFCA'06 workshop

Mathematical Foundations of Computational Anatomy

Geometrical and Statistical Methods for Modelling Biological Shape Variability

October 1st, 2006, Copenhagen, Denmark.

http://www-sop.inria.fr/asclepios/events/MFCA06

Editors:

Xavier Pennec (Asclepios, INRIA Sophia-Antipolis, France)
Sarang Joshi (SCI, University of Utah, USA)

INRIA

MICCAI

Mathematical Foundations of Computational Anatomy (MFCA'06)

Forward

Non-linear registration and shape analysis are well developed research topic in the medical image analysis community. There is nowadays a growing number of methods that can faithfully deal with the underlying biomechanical behaviour of intra-subject shape deformations. However, it is more difficult to relate the anatomical shape of different subjects. The goal of computational anatomy is to analyse and to statistically model this specific type of geometrical information. In the absence of any justified physical model, a natural attitude is to explore very general mathematical methods, for instance diffeomorphisms. However, working with such infinite dimensional space raises some deep computational and mathematical problems. In particular, one of the key problem is to do statistics. Likewise, modelling the variability of surfaces leads to rely on shape spaces that are much more complex than for curves. To cope with these, different methodological and computational frameworks have been proposed. The goal of the workshop was to foster interactions between researchers investigating the combination of geometry and statistics for modelling biological shape variability from image and surfaces. A special emphasis was put on theoretical developments, applications and results being welcomed as illustrations.

Contributions were solicited in the following areas:

- Riemannian and group theoretical methods on non-linear transformation spaces
- Advanced statistics on deformations and shapes
- Metrics for computational anatomy
- Geometry and statistics of surfaces

26 submissions of very high quality were recieved and were reviewed by two members of the programm committee. 12 papers were finally selected for oral presentations and 8 for poster presentations. 16 of these papers are published in these proceedings, and 4 papers are published in the proceedings of MICCAI'06 (for copyright reasons, only extended abstracts are provided here).

Chairs

Xavier Pennec (Asclepios, INRIA Sophia-Antipolis, France)
Sarang Joshi (SCI, Univ. Of Utah, Salt Lake City, USA).

Program committee

Ian L. Dryden (University of Nottingham, UK),
Olivier Faugeras (INRIA, France),
Tom Fletcher (University of Utah, USA),
James Gee (Univ. of Pennsylvania, USA),
Guido Gerig (Univ. North Carolina Chapel Hill, USA),
Stephen Marsland (Massey University, New-Zeland),
Michael I. Miller (John Hopkins University, USA),
Mads Nielsen (IT Univ of Copenhagen, Danmark),
Jerry Prince (John Hopkins University, USA),
Paul Thompson (Univ. of Calif. Los-Angeles, USA),
Alain Trouvé (ENS-Cachan, France),
Carole Twinning (University of Manchester, UK),
Guilermo Sapiro (University of Minnesota, USA),
Hemant D. Tagare (Yale School of Medicine, USA) ,
Baba Vemuri (University of Florida, USA),
Ross T. Whitaker (University of Utah, USA),
Laurent Younes (John Hopkins University, USA).

Organization

MFCA-2006 is a satellite workshop of MICCAI 2006, held in Copenhagen on October 1st.

Table of Content

Session 1 - Statistics on Diffeomorphisms

Session 2 - Metrics on curves and surfaces

Session 3 - New methods for warping, statistics and shape description

Session 4 - Point set Methods

Poster Session

Left-Invariant Riemannian Elasticity: a distance on shape diffeomorphisms?

X. Pennec

INRIA Sophia - Projet Epidaure, 2004 Route des Lucioles BP 93
06902 Sophia Antipolis Cedex, France
Xavier.Pennec@sophia.inria.fr

Abstract. In inter-subject registration, one often lacks a good model of the transformation variability to choose the optimal regularization. Some works attempt to model the variability in a statistical way, but the re-introduction in a registration algorithm is not easy. In [1], we interpreted the elastic energy as the distance of the Green-St Venant strain tensor to the identity. By changing the Euclidean metric for a more suitable Riemannian one, we defined a consistent statistical framework to quantify the amount of deformation. In particular, the mean and the covariance matrix of the strain tensor could be efficiently computed from a population of non-linear transformations and introduced as parameters in a Mahalanobis distance to measure the statistical deviation from the observed variability. This statistical Riemannian elasticity was able to handle anisotropic deformations but its isotropic stationary version was locally inverse-consistent. In this paper, we investigate how to modify the Riemannian elasticity to make it globally inverse consistent. This allows to define a left-invariant "distance" between shape diffeomorphisms that we call the left-invariant Riemannian elasticity. Such a closed form energy on diffeomorphisms can optimize it directly without relying on a time and memory consuming numerical optimization of the geodesic path.

1 Introduction

Most non-linear image registration algorithms optimize a criterion including an image intensity similarity and a regularization term. Many image similarity criteria are now available, ranging from the simple sum of squared intensity differences to robust information theory based measures. In inter-subject registration, the main problem is not really the intensity similarity measure but rather the regularization criterion. Some authors used physical models like elasticity or fluid models [2, 3]. For efficiency reasons, other authors proposed to use non-physical but efficient regularization methods like Gaussian filtering [4–6]. This type of regularization was then extended to more general isotropic vectorial filters [7], and to non-stationary regularization criteria in order to take into account some anatomical information about the tissue types [8, 9].

However, since we do not have in general a model of the deformation of organs across subjects, no regularization criterion is obviously more justified than

the others. We could think of of relating the anatomy of two different subjects by building a model of the organ growth: inverting the model from the first subject to a sufficiently early stage and growing toward the second subject image would allow to relate the two anatomies. However, such a computational model is out of reach now, and most of the existing work in the literature rather try to capture the organ variability from a statistical point of view on a representative population of subjects (see e.g. [10–12]). Although the image databases are now large enough to be representative of the organ variability, the problem remains of how to use this information to better guide inter-subject registration.

Ashburner at al observed in [13] that, *as the structural variability is often greater in certain directions [14], some form of a tensor field describing normal variability in each direction may be appropriate. A data representation of this form, together with a canonical brain template and associated error variance image, would allow anatomical comparisons to be made against the normal population.* This is in essence what we proposed with the Statistical Riemannian Elasticity [1]: an integrated framework to compute statistics on deformations and reintroduce them in the registration procedure, based on the field of strain tensors. The basic idea is to interpret the elastic energy as a distance in the space of positive definite symmetric matrices (tensors). By changing the classical Euclidean metric for a more suitable one, namely a log-Euclidean one in [1], we defined a natural framework for computing statistics on the strain tensor. A related idea was already present in [13] with a regularization prior based on a log-Gaussian distribution of the singular values of the Jacobian matrix of the transformation. Our key contribution in [1] was to consider the strain tensor instead of the Jacobian of the transformation. This allows to easily extend such an isotropic and stationary prior to anisotropic and non stationary ones.

In this paper, the goal is to better understand the link between Riemannian elasticity and invariant metrics on groups of diffeomorphisms, as used for instance in [15, 16]. We also reformulate the derivation of the whole theory to better stress the link with classical mechanics. We first detail how the standard elastic regularization can be optimized in a gradient descent based registration algorithm. Then, we introduce in Section 3 the Riemannian elasticity energy by changing the Euclidean distance on the strain Tensor to the identity by a log-Euclidean Riemannian distance. The simplest distances are the isotropic ones: the energy expression turns out to be very similar to the classical elastic energy while being locally inverse-consistent. One can also include non-stationary and anisotropic statistics on the strain tensors observed in a population by taking the Mahalanobis distance on the logarithmic strain tensor (statistical Riemannian elasticity). The gradients of these Riemannian elastic criteria needed to implement a practical registration algorithm are detailed in Section 4. In Section 5, we modify the spatial integration of the isotropic Riemannian elasticity in order to make it globally inverse-consistent. This leads to a left- (or right-) invariant energy on shape diffeomorphisms that can be optimized directly without having to find the geodesics through an optimization process as in standard diffeomorphic matching algorithm.

2 Standard elastic regularization

Let $I(x)$ and $J(x)$ the intensity functions of two images and $\Phi(x)$ be a non-linear space transformation assumed to be diffeomorphic with a positive Jacobian everywhere. We denote by $\{e_\alpha\}$ a set of orthonormal vectors (a basis) of the three-dimensional space, and by $\partial_\alpha \Phi$ the directional derivatives of the transformation along the spaces axis α. The general registration method is to optimize an energy of the type: $C(\Phi) = Sim(\text{Images}, \Phi) + Reg(\Phi)$. Starting from an initial transformation Φ_0, a first order gradient descent methods computes the gradient of the energy $\nabla C(\Phi)$, and update the transformation using: $\Phi_{t+1} = \Phi_t - \eta \nabla C(\Phi_t)$. From a computational points of view, this Lagrangian framework can be advantageously changed into a Eulerian framework to better conserve the diffeomorphic nature of the mappings [9]. In the following, we do not focus on the optimization of the similarity criterion (see e.g. [5, 6]), but rather on the computation of the gradient of the regularization. We assume Neumann boundary conditions on transformations and an invariant integration domain ($\Phi(\Omega) = \Omega$), so that we can drop the integration domain to simplify notations.

2.1 Elastic deformations

In continuum mechanics [17], one characterizes the deformation of an infinitesimal volume element in the Lagrangian framework using the Cauchy-Green tensor $\Sigma = \nabla\Phi^{\mathrm{T}} \nabla\Phi = \sum_\alpha \partial_\alpha\Phi \, \partial_\alpha\Phi^{\mathrm{T}}$. This symmetric matrix is positive definite if the transformation is diffeomorphic, and measures the local amount of non-rigidity. Let $\nabla\Phi = V \, S \, R^{\mathrm{T}}$ be a singular value decomposition of the transformation Jacobian (R and V are two rotation matrices and S is the diagonal matrix of the positive singular values). The Cauchy-Green tensor $\Sigma = R \, S^2 \, R^{\mathrm{T}}$ is equal to the identity if and only if the transformation is locally a rigid transformation. Eigenvalues between 0 and 1 indicate a local compression of the material along the associated eigenvector, while a value above 1 indicates an expansion.

To quantify the deformation, one usually prefers the related Green-St Venant strain tensor $E = \frac{1}{2}(\Sigma - \text{Id})$, whose eigenvalues are null for no deformation. This tensor is often expressed using the displacement field: $E = \frac{1}{2}(\nabla U + \nabla U^{\mathrm{T}} + \nabla U^{\mathrm{T}} \nabla U)$ (dropping the quadratic term leads to the linear elasticity). Assuming an isotropic material and a linear Hooks law to relate strain and stress tensors, one can show that the motion equations derive from the *St Venant-Kirchoff elasticity* energy [17]:

$$Reg_{SVKE}(\Phi) = \int \mu \text{Tr}(E^2) + \frac{\lambda}{2}\text{Tr}(E)^2 = \int \frac{\mu}{4}\text{Tr}\left((\Sigma - \text{Id})^2\right) + \frac{\lambda}{8}\text{Tr}(\Sigma - \text{Id})^2$$

2.2 Optimizing the elasticity

To minimize this energy in a registration algorithm, we need its gradient. Since $\partial_u \Sigma = \sum_\alpha \left(\partial_\alpha\Phi \, \partial_\alpha u^{\mathrm{T}} + \partial_\alpha u \, \partial_\alpha\Phi^{\mathrm{T}}\right)$, the derivative of the elastic energy in the

direction (i.e. displacement field) u is:

$$\begin{aligned}\partial_u Reg_{SVKE}(\Phi) &= \int \tfrac{\mu}{2}\mathrm{Tr}((\Sigma - \mathrm{Id})\,\partial_u\Sigma) + \tfrac{\lambda}{4}\mathrm{Tr}(\Sigma - \mathrm{Id})\,\mathrm{Tr}(\partial_u\Sigma)\\ &= \textstyle\sum_\alpha \int \langle\, \mu\,(\Sigma - \mathrm{Id})\,\partial_\alpha\Phi \mid \partial_\alpha u\,\rangle + \tfrac{\lambda}{2}\mathrm{Tr}(\Sigma - \mathrm{Id})\ \langle\, \partial_\alpha\Phi \mid \partial_\alpha u\,\rangle\end{aligned}$$

Using an integration by part with homogeneous Neumann boundary conditions [6], we have $\int \langle\, v \mid \partial_\alpha u\,\rangle = -\int \langle\, \partial_\alpha v \mid u\,\rangle$, so that the gradient is finally:

$$\nabla Reg_{SVKE}(\Phi) = -\textstyle\sum_\alpha \partial_\alpha\big(Z\,\partial_\alpha\Phi\big) \quad \text{with} \quad Z = \mu(\Sigma - \mathrm{Id}) + \tfrac{\lambda}{2}\mathrm{Tr}(\Sigma - \mathrm{Id})\ \mathrm{Id}$$

Here, Z is the derivative of the density of energy at each point with respect to the strain tensor Σ and is known as the 2nd Piola-Kirchoff tensor. The 3rd order tensor $Z\partial_\alpha\Phi$ is the first Piola-Kirchoff tensor and corresponds to the derivative of the density of energy with respect to the Jacobian of the transformation.

3 Log-Euclidean Riemannian elasticity

In the standard elasticity theory, the deviation of the positive definite symmetric matrix Σ (the strain tensor) from the identity (the rigidity) is measured using the Euclidean matrix distance $\mathrm{dist}^2_{Eucl}(\Sigma, \mathrm{Id}) = \mathrm{Tr}((\Sigma - \mathrm{Id})^2)$. However, it has been argued in recent works that the Euclidean metric is not a good metric for the tensor space because positive definite symmetric matrices only constitute a cone in the Euclidean matrix space. Thus, the tensor space is not complete (null or negative eigenvalues are at a finite distance). For instance, an expansion of a factor $\sqrt{2}$ in each direction (leading to $\Sigma = 2\,\mathrm{Id}$) is at the same Euclidean distance from the identity than the "black hole" transformation $\Phi(x) = 0$ (which has a non physical null strain tensor). In non-linear registration, this asymmetry of the regularization leads to different results if we look for the forward or the backward transformation: this is the inverse-consistency problem [18].

3.1 A Log-Euclidean metric on the strain tensor

To solve the problems of the Euclidean tensor computing, affine-invariant Riemannian metrics were recently proposed [19–22]. Using these metrics, symmetric matrices with null eigenvalues are basically at an infinite distance from any tensor, and the notion of mean value corresponds to a geometric mean, even if it has to be computed iteratively. More recently, [23] proposed Log-Euclidean metrics, which exhibit the same properties while being much easier to compute. As these metrics simply consist in taking a standard Euclidean metric after a (matrix) logarithm, and since they correspond to the previous ones as long as the reference point is the identity, we relied on the later in the original definition of the Riemannian elasticity [1]. However, the Riemannian Elasticity principle can be generalized to any Riemannian metric on the tensor space without any restriction. We will see that the full linear invariance properties of affine-invariant metrics will prove to be necessary in Section 5 to properly define a left-invariant energy on shape diffeomorphisms.

In the log-Euclidean Riemannian framework, the deviation between the tensor Σ and the identity is the tangent vector $\log(\Sigma) - \log(\text{ Id}) = \log(\Sigma)$. Interestingly, this tensor is known in continuum mechanics as the logarithmic or Hencky strain tensor [24], and is used for modeling very large deformations. [25]. It is considered as the natural strain tensor for many materials, but its use was hampered for a long time because of its computational complexity [26].

For registration, the basic idea is to replace the elastic energy with a regularization that measures the amount of logarithmic strain by taking a Riemannian distance between Σ and Id. With a log-Euclidean metric, this give the *log-Euclidean Riemannian elasticity*:

$$Reg_{LERE}(\Phi) = \tfrac{1}{4}\int \text{dist}^2_{Log}(\Sigma,\ \text{Id}) = \tfrac{1}{4}\int \text{dist}^2_{Eucl}(\log(\Sigma), \log(\text{ Id}))^2 = \tfrac{1}{4}\int \|\log(\Sigma)\|^2$$

3.2 Isotropic Log-Euclidean Riemannian elasticity

The simplest metric on a logarithmic strain tensor $W = \log(\Sigma)$ is $\|W\|^2 = \text{Tr}(W^2)$. More generally, any metric is given by a bilinear form $G(W_1, W_2)$ on the space of symmetric matrices, and is uniquely specified by the quadratic form $\|W\|^2 = G(W, W)$. A metric is isotropic if $\|W\|^2 = \|R\,W\,R^{\text{T}}\|^2$ for any rotation R. This means that it only depends on the eigenvalues of W, or equivalently on the matrix invariants $\text{Tr}(W)$, $\text{Tr}(W^2)$ and $\text{Tr}(W^3)$. However, as the form is quadratic in W, we are left only with $\text{Tr}(W)^2$ and $\text{Tr}(W^2)$ that can be weighted arbitrarily, e.g. by μ and $\lambda/2$ (with $n.\lambda > -2\mu$ where n is the dimension of the space to ensure the positive definiteness of the metric). Finally, the *isotropic log-Euclidean Riemannian elasticity (ILERE)* energy has the form:

$$Reg_{ILERE}(\Phi) = \int \tfrac{\mu}{4}\text{Tr}\left((\log(\Sigma))^2\right) + \tfrac{\lambda}{8}\text{Tr}(\log(\Sigma))^2$$

We retrieve the classical form of the isotropic elastic energy with Lamé coefficients, but with the logarithmic strain tensor. This form was expected as the St Venant-Kirchoff energy was also derived for isotropic materials.

3.3 Incorporating deformation statistics

In the context of inter-subject or atlas-to-image registration, we do not know a priori the deformability of the material. Moreover, we don't expect it to be isotropic nor stationary. An interesting idea is to learn the local deformability characteristics from a population of typical transformations $\Phi_i(x)$.

Based on the statistical framework presented in [22], we considered in [1] the strain tensor as a random variable in the Riemannian space of tensors. We defined the *a priori* deformability $\bar{\Sigma}(x)$ as the Riemannian mean of deformation tensors $\Sigma_i(x) = \nabla\Phi_i^{\text{T}}\,\nabla\Phi_i$. A related idea was suggested directly on the Jacobian matrix of the transformation $\nabla\Phi$ in [27], but using a general matrix instead of a symmetric one raises important computational and theoretical problems. With the Log-Euclidean metric on strain tensors, the statistics are quite simple since we have a closed form for the mean value:

$$\bar{\Sigma}(x) = \exp(\bar{W}(x)) \quad \text{with} \quad \bar{W}(x) = \tfrac{1}{N}\textstyle\sum_i \log(\Sigma_i(x))$$

This mean deformability $\bar{\Sigma}$ is not so easy to understand. If the reference image is optimally centered with respect to the data, one could expect the mean deformation to be null ($\bar{W} = 0$). However, this equation specifies $n(n+1)/2$ scalar components while there are only n free scalar components at each point of a displacement field. Thus, it seems at the first glance that $n(n-1)/2$ scalar components (e.g. off diagonal terms of $\bar{W}$) could not be prescribed to zero. More powerful tools from the singularity theory are probably necessary to definitely conclude on that point.

Going one step further, we can compute the covariance matrix of the random process $\mathrm{Cov}(\Sigma_i(x))$ at each point. Let us decompose the symmetric tensor $W = \log(\Sigma)$ into a vector $\mathrm{Vect}(W)^{\mathrm{T}} = (w_{11}, w_{22}, w_{33}, \sqrt{2}w_{12}, \sqrt{2}w_{13}, \sqrt{2}w_{23})$ that gathers all the tensor components in an orthonormal basis. In this coordinate system, we define the covariance matrix $\mathrm{Cov} = \frac{1}{N}\sum \mathrm{Vect}(W_i - \bar{W})\,\mathrm{Vect}(W_i - \bar{W})^{\mathrm{T}}$.

To adapt the metric on strain tensors to these first and second order moments of the random deformation process, a well known and simple tool is the Mahalanobis distance, so that we finally define the *statistical Log-Euclidean Riemannian elasticity (SLERE)* energy as:

$$Reg_{SLERE}(\Phi) = \tfrac{1}{4}\int \mu^2_{(\bar{W},Cov)}(\log(\Sigma(x))) = \tfrac{1}{4}\int \mathrm{Vect}(W-\bar{W})\mathrm{Cov}^{(-1)}\mathrm{Vect}(W-\bar{W})^{\mathrm{T}}$$

As we are using a Mahalanobis distance, this least-squares criterion can be seen as the log-likelihood of a Gaussian process on strain tensor fields: we are implicitly modeling the a-priori probability of the deformation. In a registration framework, this point of view is particularly interesting as it opens the way to use Bayesian estimation methods for non-linear registration.

4 Optimizing the Riemannian elasticity

To use the logarithmic elasticity energies as regularization criteria in the registration framework, we have to compute their gradient. Let us consider the isotropic Riemannian elasticity first. Thanks to the properties of the differential of the log (see appendix A), we have $\mathrm{Tr}(\partial_V \log(\Sigma)) = \mathrm{Tr}(\Sigma^{(-1)}\, V)$ and $\langle\, \partial_V \log(\Sigma) \mid W \,\rangle = \langle\, \partial_W \log(\Sigma) \mid V \,\rangle$. Thus, using $V = \partial_u \Sigma = \sum_\alpha (\partial_\alpha u\, \partial_\alpha \Phi^{\mathrm{T}} + \partial_\alpha \Phi\, \partial_\alpha u^{\mathrm{T}})$ and $W = \log(\Sigma)$, we can write the directional derivative of the criterion:

$$\begin{aligned}
\partial_u Reg_{ILERE}(\Phi) &= \int \tfrac{\mu}{2} \langle\, W \mid \partial_V \log(\Sigma) \,\rangle + \tfrac{\lambda}{4}\mathrm{Tr}(W)\,\mathrm{Tr}(\partial_V \log(\Sigma)) \\
&= \int \tfrac{\mu}{2} \langle\, \partial_W \log(\Sigma) \mid V \,\rangle + \tfrac{\lambda}{4}\mathrm{Tr}(W)\,\mathrm{Tr}(\Sigma^{(-1)}\, V) \\
&= \textstyle\sum_\alpha \int \mu \langle\, \partial_W \log(\Sigma)\, \partial_\alpha \Phi \mid \partial_\alpha u \,\rangle + \tfrac{\lambda}{2}\mathrm{Tr}(W)\ \langle\, \Sigma^{(-1)}\, \partial_\alpha \Phi \mid \partial_\alpha u \,\rangle
\end{aligned}$$

Integrating by part with homogeneous Neumann boundary conditions, we end up with the gradient:

$$\nabla Reg_{IRE}(\Phi) = -\textstyle\sum_\alpha \partial_\alpha (Z\, \partial_\alpha \Phi) \quad \text{with} \quad Z = \mu\, \partial_W \log(\Sigma) + \tfrac{\lambda}{2}\mathrm{Tr}(W)\, \Sigma^{(-1)} \quad (1)$$

The same formula still holds for the general statistical Riemannian elasticity with $Z = \partial_X \log(\Sigma)$ where X is the symmetric matrix defined by $\mathrm{Vect}(X) =$

$\mathrm{Cov}^{(-1)}\ \mathrm{Vect}(\log(\Sigma) - \bar{W})$. Thus, we may write the gradient of all (St-Venant-Kirchoff, Isotropic Riemannian and Statistical Riemannian) elastic energies as:

$$\nabla Reg(\Phi) = -\sum_\alpha \partial_\alpha (Z\, \partial_\alpha \Phi) \tag{2}$$

and only the 2nd Piola-Kirchoff tensor Z differs:

$$Z_{SVKE} = \mu(\Sigma - \mathrm{Id}) + \tfrac{\lambda}{2}\mathrm{Tr}(\Sigma - \mathrm{Id})\ \mathrm{Id} \tag{3}$$

$$Z_{ILERE} = \mu\, \partial_W \log(\Sigma) + \tfrac{\lambda}{2}\mathrm{Tr}(\log(\Sigma))\ \Sigma^{(-1)} \tag{4}$$

$$Z_{SLERE} = \partial_X \log(\Sigma) \quad \text{with} \quad \mathrm{Vect}(X) = \mathrm{Cov}^{(-1)}\ \mathrm{Vect}(\log(\Sigma) - \bar{W}) \tag{5}$$

4.1 Practical implementation

A simple and easily parallelisable implementation is the following. First, one computes the image of the gradient of the transformation, or more particularly the directional derivatives, for instance using finite differences. $\partial_\alpha \Phi(x) = (\Phi(x + \tau_\alpha\, e_\alpha) - \Phi(x - \tau_\alpha\, e_\alpha))/2\tau_\alpha$, where τ_α is the voxel size in the direction α. This operation is not computationally expensive, but requires to access the value of the transformation field at neighboring points, which can be time consuming due to systematic memory page faults in large images.

Then, we process these 3 vectors completely locally to compute 3 new vectors $v_\alpha = Z(\partial_\alpha \Phi)$. This operation is computationally more expensive but is memory efficient as the resulting vectors can replace the old directional derivatives. Finally, the gradient of the criterion $\nabla E = \sum_\alpha \partial_\alpha v_\alpha$ may be computed using finite differences on the resulting image. $\nabla E(x) = \sum_\alpha (v_\alpha(x + \tau_\alpha\, e_\alpha) - v_\alpha(x - \tau_\alpha\, e_\alpha))/2\tau_\alpha$. Once again, this is not computationally expensive, but it requires intensive memory accesses.

The only additional cost for the Riemannian Elasticity is the computation of the logarithm $W = \log(\Sigma)$ and its directional derivative $\partial_W \log(\Sigma)$. This would probably be prohibitive if we had to rely on numerical approximation methods. Fortunately, we were able to compute an explicit and very simple and efficient closed-form expression that only requires the diagonalization of Σ (see appendix A). Experiments performed in [1] showed that optimizing the isotropic Riemannian elasticity was only 3 time longer than optimizing the standard elasticity.

5 Left Invariant Riemannian Elasticity

Let us now investigate the invariance properties in view of relating the Riemannian elasticity to metrics on diffeomorphisms. Since $\nabla(\Phi^{(-1)}) \circ \Phi = (\nabla \Phi)^{(-1)}$, the isotropic logarithmic distance of a strain tensor to the identity is locally inverse-consistent. We have indeed $\mathrm{Tr}\left(\log(\Sigma_\Phi)^2\right) = \mathrm{Tr}\left(\log(\Sigma_{\Phi^{(-1)}} \circ \Phi)^2\right)$ and $\mathrm{Tr}\left(\log(\Sigma_\Phi)\right) = \mathrm{Tr}\left(\log(\Sigma_{\Phi^{(-1)}} \circ \Phi)\right)$. This means that, locally, a scaling of a factor 2 at the same distance from the identity than a scaling of 0.5. However, this property does not hold globally due to the change of the volume element during

the change of variable $y = \Phi(x)$:

$$\begin{aligned} Reg_{IRE}(\Phi^{(-1)}) &= \int \tfrac{\mu}{4}\mathrm{Tr}\left((\log(\Sigma_{\Phi^{(-1)}}(y)))^2\right) + \tfrac{\lambda}{8}\mathrm{Tr}(\log(\Sigma_{\Phi^{(-1)}}(y)))^2.dy \\ &= \int \tfrac{\mu}{4}\mathrm{Tr}\left((\log(\Sigma_{\Phi}(x)))^2\right) + \tfrac{\lambda}{8}\mathrm{Tr}(\log(\Sigma_{\Phi}(x)))^2.\sqrt{|\Sigma_{\Phi}(x)|}.dx \end{aligned}$$

5.1 Inverse Consistent Riemannian Elasticity

Following an idea suggested in [28], we can integrate with a volume element which is the geometric mean between the one in the original space and the one in the arrival space, i.e.: $\sqrt{|\nabla\Phi(x)|}.dx = |\Sigma(x)|^{1/4}.dx$. If f is a locally inverse consistent functional (i.e. such that $f(\Phi^{(-1)}) \circ \Phi = f(\Phi)$), then the integral value $F(\Phi) = \int f(\Phi).\sqrt{|\nabla\Phi|}$ is also inverse consistent. Indeed, the change of variable $y = \Phi(x)$ induces $dy = |\nabla\Phi(x)|.dx$, but since $|\nabla(\Phi^{(-1)}) \circ \Phi| = |\nabla\Phi|^{(-1)}$, we have:

$$F(\Phi^{(-1)}) = \int f(\Phi^{(-1)})(y).\sqrt{|\nabla(\Phi^{(-1)})(y)|}.dy = \int f(\Phi)(x).\sqrt{|\nabla(\Phi)(x)|}.dx = F(\Phi)$$

As the log-Euclidean distance of a strain tensor to the identity is locally inverse consistent, we thus obtain a globally inverse consistent (isotropic) Riemannian elasticity with:

$$Reg_{ICRE}(\Phi) = \int \left\{ \frac{\mu}{4}\mathrm{Tr}\left(\log(\Sigma)^2\right) + \frac{\lambda}{8}\mathrm{Tr}(\log(\Sigma))^2 \right\}.|\Sigma|^{1/4}.dx$$

Another formulation may be obtained using the change of variable $y = \Phi(x)$ and will turn out to be generalizable to a left-invariant energy:

$$Reg_{ICRE}(\Phi) = \int \left\|\log(\Sigma \circ \Phi^{(-1)})\right\|^2.|\Sigma \circ \Phi^{(-1)}|^{-1/4} \tag{6}$$

In this formula, the norm $\|.\|$ refers to an isotropic norm on symmetric matrices.

The derivative of this new criterion can be deduced from $\partial_u Reg_{IRE}$ using:

$$\partial_u \det(\Sigma)^{1/4} = \tfrac{1}{4}\mathrm{Tr}(\Sigma^{(-1)}.\partial_u\Sigma).\det(\Sigma)^{1/4} = \tfrac{1}{2}\textstyle\sum_\alpha \langle\, \Sigma^{(-1)}.\partial_\alpha\Phi \mid \partial_\alpha u \,\rangle \det(\Sigma)^{1/4}$$

We have one again $\partial_u Reg_{ICRE}(\Phi) = \sum_\alpha \int \langle\, Z.\partial_\alpha\Phi \mid \partial_\alpha u \,\rangle$, with

$$Z_{ICRE} = \left(Z_{IRE} + \frac{1}{2}\|W\|^2.\Sigma^{(-1)}\right)\det(\Sigma)^{1/4}$$

Thus, we have obtained an inverse invariant energy on diffeomorphisms which allows us to optimize directly their regularity in registration processes without having to integrate numerically along the transformation trajectory for computing the length of geodesics, as for the invariant metrics on diffeomorphisms proposed in [15, 16].

5.2 Left-invariant Riemannian elasticity

This energy is positive and null only is the transformation is locally rigid everywhere. It can be turned into a left- (or right-) invariant "distance" by left- (resp. right) translation. Let us investigate the left-invariant "distance" (The right-invariant distance is automatically given by $\mathrm{dist}_R(\Phi,\Psi) = \mathrm{dist}_L(\Phi^{(-1)},\Psi^{(-1)})$):

$$\mathrm{dist}_L^2(\Phi,\Psi) = Reg_{ICRE}(\Phi^{(-1)}\circ\Psi) = \int \|\log(\Sigma_{\Phi^{(-1)}\circ\Psi})\|^2 .|\Phi^{(-1)}\circ\Psi|^{1/4}$$

Thanks to the inverse invariance, the "distance" is symmetric. It is null if and only if the two diffeomorphisms differ by a local rotation everywhere. However, to show that this is really a left-invariant distance on diffeomorphisms of rigid shapes, the triangular inequality remains to be established. Moreover, we suspect that we obtain an extrinsic distance and not a Riemannian one.

The expression of the left-invariant distance can be worked out to see how much it differs from the previously proposed statistical Riemannian elasticity. We first notice that $\nabla(\Phi^{(-1)}\circ\Psi) = \nabla\Psi.\nabla(\Phi^{(-1)})\circ\Psi = \nabla\Psi.(\nabla\Phi)^{(-1)}\circ(\Phi^{(-1)}\circ\Psi)$. Using the singular value decomposition $\nabla\Phi = U.S.V^{\mathrm{T}}$, there exists a rotation $R = V.U^{\mathrm{T}}$ at each point such that $R.\nabla\Phi = \Sigma_\Phi^{1/2}$. Thus, we have:

$$\Sigma_{\Phi^{(-1)}\circ\Psi} = R^{\mathrm{T}}.\left(\Sigma_\Phi^{-1/2}\circ(\Phi^{(-1)}\circ\Psi)\right).\Sigma_\Psi.\left(\Sigma_\Phi^{-1/2}\circ(\Phi^{(-1)}\circ\Psi)\right).R$$

But thanks to $\log(R^{\mathrm{T}}.\Sigma.R) = R^{\mathrm{T}}.\log(\Sigma).R$ and to the isotropy of the norm on symmetric matrices, the rotation R disappear in the distance. Finally, using the change of variable $y = \Psi(x)$, we end up with

$$\mathrm{dist}_L^2(\Phi,\Psi) = \int \frac{\left\|\log\left((\Sigma_\Phi^{-1/2}\circ\Phi^{(-1)}).(\Sigma_\Psi\circ\Psi^{(-1)}).(\Sigma_\Phi^{-1/2}\circ\Phi^{(-1)})\right)\right\|^2}{\det(\Sigma_\Psi\circ\Psi^{(-1)})^{-1/4}.\det(\Sigma_\Phi\circ\Phi^{(-1)})^{-1/4}}.$$

Besides symmetric corrections for the volume element, one recognize here the affine-invariant distance on symmetric matrices instead of the log-Euclidean one as we originally proposed for the statistical Riemannian elasticity. Using the resampled tensor fields $\hat{\Sigma}_\phi = \Sigma_\Phi\circ\Phi^{(-1)}$ and $\hat{\Sigma}_\psi = \Sigma_\Psi\circ\Psi^{(-1)}$, we finally obtain:

$$\mathrm{dist}_L^2(\Phi,\Psi) = \int \mathrm{dist}_{Aff}^2\left(\hat{\Sigma}_\Phi\ ,\ \hat{\Sigma}_\Psi\right).\det(\hat{\Sigma}_\Psi)^{-1/4}.\det(\hat{\Sigma}_\Phi)^{-1/4} \qquad (7)$$

Other simple formulations of the left (and of the right) invariant "distance" are possible, and we are currently analyzing them to find out the more intuitive ones. Following the statistical framework of [29], computing the derivatives will allow determining the barycentric equation of the Fréchet "mean diffeomorphisms" according to these "metrics", and a gradient descent algorithm to obtain them. Then, we hope to be able to compute second order moment and to define a kind of Mahalanobis distance (including local anisotropy and non-stationarity) on shape diffeomorphisms.

6 Discussion

Riemannian elasticity is an integrated framework to compute the statistics on deformations and re-introduce them as constraints in non-linear registration algorithms. This framework is based on the interpretation of the elastic energy as a Euclidean distance between the Cauchy-Green strain tensor and the identity (i.e. the local rigidity). By providing the space of tensors with a more suitable Riemannian metric, for instance a Log-Euclidean one, we can define proper statistics on deformations, like the mean and the covariance matrix. Taking these measurements into account in a statistical (i.e. a Mahalanobis) distance, we end-up with the statistical Riemannian elasticity regularization criterion. This criterion can also be viewed as the log-likelihood of the deformation probability, which opens the way to Bayesian deformable image registration algorithms.

We investigated in this paper the theoretical properties of the isotropic and stationary version and we showed that it was possible to obtain an inverse-consistent criterion by modifying the spatial integration measure. It is remarkable that this allows to define a left or right invariant energy between two diffeomorphisms without having to optimize for the geodesic path between them. However, many questions are left open. For instance, it remains to be established that our energy is a distance, and if it is Riemannian or extrinsic. Determining the geodesics (if they exist) would also be very interesting to better understand the properties of these energies. This would probably help also in generalizing the statistical Riemannian elasticity in a consistent way, in order to measure and take into account anisotropic and non-stationary behavior of the deformations. On a more theoretical point of view, it would be interesting to make the link between our approach and the Brownian warps of [28, 30].

References

1. X. Pennec, R. Stefanescu, V. Arsigny, P. Fillard, and N. Ayache. Riemannian elasticity: A statistical regularization framework for non-linear registration. In *Proc. of MICCAI 2005, Part II*, volume 3750 of *LNCS*, pages 943–950, Palm Springs, CA, USA, October 26-29, 2005. Springer Verlag.
2. R. Bajcsy and S. Kovačič. Multiresolution elastic matching. *Computer Vision, Graphics and Image Processing*, 46:1–21, 1989.
3. G. E. Christensen, S. C. Joshi, and M. I. Miller. Volumetric transformation of brain anatomy. *IEEE Trans. on Medical Imaging*, 16(6):864–877, December 1997.
4. J.-P. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3), 1998.
5. X. Pennec, P. Cachier, and N. Ayache. Understanding the "demon's algorithm": 3D non-rigid registration by gradient descent. In *Proc. of MICCAI'99*, volume 1679 of *LNCS*, pages 597–605, Cambridge, UK, September 1999. Springer Verlag.
6. J. Modersitzki. *Numerical Methods for Image Registration*. Numerical Mathematics and Scientific Computations. Oxford University Press, 2004.
7. P. Cachier and N. Ayache. Isotropic energies, filters and splines for vectorial regularization. *J. of Math. Imaging and Vision*, 20(3):251–265, May 2004.

8. H. Lester, S.R. Arridge, K.M. Jansons, L. Lemieux, J.V. Hajnal, and A. Oatridge. Non-linear registration with the variable viscosity fluid algorithm. In *Proc of IPMI;99*, pages 238–251, 1999.
9. R. Stefanescu, X. Pennec, and N. Ayache. Grid powered nonlinear image registration with locally adaptive regularization. *Medical Image Analysis*, 8(3):325–342, September 2004.
10. P.M. Thompson, M.S. Mega, K.L. Narr, E.R. Sowell, R.E. Blanton, and A.W. Toga. Brain image analysis and atlas construction. In M. Fitzpatrick and M. Sonka, editors, *Handbook of Medical Image Proc. and Analysis*, chapter 17. SPIE, 2000.
11. D. Rueckert, A.F. Frangi, and J.A. Schnabel. Automatic construction of 3D statistical deformation models of the brain using non-rigid registration. *IEEE TMI*, 22:1014–1025, 2003.
12. P. Fillard, V. Arsigny, X. Pennec, P. Thompson, and N. Ayache. Extrapolation of sparse tensor fields: Application to the modeling of brain variability. In *Proc. of IPMI'05*, volume 3565 of *LNCS*, pages 27–38, Glenwood springs, Colorado, USA, July 2005. Springer.
13. J. Ashburner, J.L.R. Andersson, and K.J. Friston. High-dimensional image registration using symmetric priors. *NeuroImage*, 9:619–628, 1999.
14. P.M. Thompson, C. Schwartz, R. T. Lin, A. A. Khan, and A.W. Toga. 3D statistical analysis of sulcal variability in the human brain. *J. Neurosci*, 16(13):4261–4274, 1996.
15. S.C. Joshi and M.I Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370, 2000.
16. M.F. Beg, M.I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. Journal of Computer Vision*, 61(2):139–157, 2005.
17. P.G. Ciarlet. *Mathematical elasticity Vol. 1: Three-dimensionnal elasticity*. Elsevier Science B.V., 1988.
18. GE Christensen and HJ. Johnson. Consistent image registration. *IEEE Trans Med Imaging*, 20(7):568–82, July 2001.
19. P. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A rigorous framework for diffusion tensor calculus. *Mag. Res. in Med.*, 53:221–225, 2005.
20. P.T. Fletcher and S.C. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Proc. of CVAMIA and MMBIA Workshops, Prague, Czech Republic, May 15, 2004*, LNCS 3117, pages 87–98. Springer, 2004.
21. Ch. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processingy. *Journal of Mathematical Imaging and Vision*, 2006. To appear. A preliminary version is available as INRIA Research Report RR-5242, 2004.
22. X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, January 2006.
23. V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the log-Euclidean framework. In *Proc. of MICCAI 2005, Part I*, volume 3749 of *LNCS*, pages 115–122, Palm Springs, CA, USA, October 26-29, 2005. Springer Verlag.
24. H. Hencky. Uber die form des elastizitatsgesetzes bei ideal elastischen stoffen. *Zeit. Techn. Phys.*, 9:241–247, 1928.
25. P. Roué. *Mécanique des grandes transformations*. Number 25 in Mathématiques & Applications. Springer, 1997.
26. A.D. Freed. Natural strain. *Journal of Engineering Materials & Technology*, 117:379–385, 1995.

27. R.P. Woods. Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation. *NeuroImage*, 18(3):769–788, 2003.
28. M. Nielsen, P. Johansen, A.D. Jackson, and B. Lautrup. Brownian warps: A least committed prior for non-rigid registration. In *Proc of MICCAI'02, Part II*, volume 2489 of *LNCS*, pages 557 – 564, 2002.
29. Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 2006. To appear (avalailable at dx.doi.org/10.1007/s10851-006-6228-4).
30. B. Markussen. A statistical approach to large deformation diffeomorphisms. *2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, 12:181, 2004.

A Appendix: tensor derivatives

A.1 Exponential of a tensor

Let $W = R\,S\,R^{\mathrm{T}}$ be a diagonalization of a symmetric matrix. We can write any power of W in the same basis: $W^k = R\,S^k\,R^{\mathrm{T}}$. Thus, the rotation matrices can be factored out in the series defining the matrix exponential, so that the exponential is applied directly to the eigenvalues:

$$\exp(W) = \sum_{k=0}^{+\infty} \frac{W^k}{k!} = R\,\mathrm{DIAG}(\exp(s_i))\,R^{\mathrm{T}}$$

This series converges for any symmetric matrix argument, and it is easy to see that its inverse is well defined for any positive definite symmetric matrix $\Sigma = R\,\mathrm{DIAG}(a_i)\,R^{\mathrm{T}}$. This is the function: $\log(\Sigma) = R\,(\mathrm{DIAG}(\log(a_i)))\,R^{\mathrm{T}}$. It is important to notice that there is no series expansion which is converging for all arguments, like for the exponential.

A.2 Differential of the exponential

The matrix exponential and logarithm realize a one-to-one mapping between the space of symmetric matrices to the the space of tensors. Moreover, one can show that this mapping is diffeomorphic, since the differential has no singularities. Using the Taylor expansion $(W+\varepsilon V)^k = W^k + \varepsilon \sum_{i=0}^{k-1} W^i\,V\,W^{k-i-1} + O(\varepsilon^2)$ for $k \geq 1$, we obtain by identification the directional derivative $\partial_V \exp(W)$ by gathering the first order terms in ε in the series $\exp(W+\varepsilon V) = \sum_{k=0}^{+\infty} (W+\varepsilon V)^k/k!$:

$$\partial_V \exp(W) = (d\exp(W))(V) = \sum_{k=1}^{+\infty} \frac{1}{k!} \sum_{i=0}^{k-1} W^i\,V\,W^{k-i-1} \tag{8}$$

For simplifying the differential, we can see that using the diagonalization $W = RS\,R^{\mathrm{T}}$ in the series gives:

$$\partial_V \exp(W) = R\,\partial_{(R^{\mathrm{T}}\,V\,R)} \exp(S)\,R^{\mathrm{T}}$$

Thus, we are left with the computation of $\partial_V \exp(S)$ for S diagonal. As $[S^l V S^{k-l-1}]_{ij} = s_i^l \, v_{ij} \, s_j^{k-l-1}$, we have $[\partial_V \exp(S)]_{ij} = \left\{ \sum_{k=1}^{+\infty} \frac{1}{k!} \sum_{l=0}^{k-1} s_i^l \, s_j^{k-l-1} \right\} v_{ij} = q_{ij} \, v_{ij}$ with

$$\begin{aligned} q_{ij} &= \sum_{k=1}^{+\infty} \frac{1}{k!} \sum_{l=0}^{k-1} s_i^l \, s_j^{k-l-1} = \sum_{k=1}^{+\infty} \frac{1}{k!} \frac{s_i^k - s_j^k}{s_i - s_j} = \frac{\exp(s_i) - \exp(s_j)}{s_i - s_j} \\ &= \exp(s_j) \left(1 + \frac{(s_i - s_j)}{2} + \frac{(s_i - s_j)^2}{6} + O\left((s_i - s_j)^3\right) \right) \end{aligned}$$

The last Taylor expansion shows that this formula is computationally well posed. Moreover, we have $q_{ij} \geq 1 > 0$, so that we can conclude that $d\exp(S)$ is a diagonal linear form that is always invertible: the exponential is a diffeomorphism.

A.3 Differential of the logarithm

To compute the differential of the logarithm function, we do not have a series that we could perturb like for the exponential, but we can simply inverse the differential of the exponential as a linear form: as $\exp(\log(\Sigma)) = \Sigma$, we have $(d\log(\Sigma)(V) = (d\exp(\log(\Sigma))^{(\text{-}1)} V$. Using $D = \exp(S)$, the inverse is easily expressed for a diagonal matrix: $[(d\exp(S))^{(\text{-}1)} \, V]_{ij} = v_{ij}/q_{ij}$. Thus we have:

$$[\partial_V \log(D)]_{ij} = v_{ij} \frac{\log(d_i) - \log(d_j)}{d_i - d_j}$$

Notice that

$$q_{ij}^{(\text{-}1)} = \frac{\log(d_i) - \log(d_j)}{d_i - d_j} = \frac{1}{d_j} \left(1 - \frac{d_i - d_j}{2 \, d_j} + \frac{(d_i - d_j)^2}{3 \, d_j^2} + O\left((d_i - d_j)^3\right) \right)$$

so that the formula is numerically stable. Finally, using the identity $\log(\Sigma) = R^{\mathrm{T}} \, \log(R \, \Sigma \, R^{\mathrm{T}}) \, R$ for any rotation R, we have:

$$\partial_V \log(R \, D \, R^{\mathrm{T}}) = R \, \left(\partial_{R^{\mathrm{T}} \, V \, R} \log(D) \right) R^{\mathrm{T}}$$

That way, we may compute the differential at any point $\Sigma = R \, D \, R^{\mathrm{T}}$.

A.4 Remarkable identities

$$\partial_{\log(\Sigma)} \log(\Sigma) = \Sigma^{(\text{-}1)} \, \log(\Sigma) = \log(\Sigma) \, \Sigma^{(\text{-}1)} \tag{9}$$

$$\langle \, \partial_V \log(\Sigma) \mid W \, \rangle = \langle \, \partial_W \log(\Sigma) \mid V \, \rangle \tag{10}$$

$$\partial_{\log(\Sigma)} \log(\Sigma) = R \, \left(\partial_{\log(D)} \log(D) \right) R^{\mathrm{T}} = R \, \mathrm{Diag}(\log(d_i)/d_i) \, R^{\mathrm{T}} = \Sigma^{(\text{-}1)} \, \log(\Sigma)$$

$$\begin{aligned} \langle \, \partial_V \log(\Sigma) \mid R \, D \, R^{\mathrm{T}} \, \rangle &= \mathrm{Tr}\left((\partial_{R^{\mathrm{T}} \, V \, R} \log(D) \,) \, R^{\mathrm{T}} \, W \, R \right) \\ &= [R^{\mathrm{T}} \, V \, R]_{ij} \, \frac{\log(d_i) - \log(d_j)}{d_i - d_j} \, [R^{\mathrm{T}} \, W \, R]_{ij} \\ &= \mathrm{Tr}\left((\partial_{R^{\mathrm{T}} \, W \, R} \log(D) \,) \, R^{\mathrm{T}} \, V \, R \right) = \langle \, \partial_W \log(\Sigma) \mid V \, \rangle \end{aligned}$$

Statistics on Diffeomorphisms in a Log-Euclidean Framework

Vincent Arsigny[1], Olivier Commowick[1,2], Xavier Pennec[1], and Nicholas Ayache[1]

[1] INRIA Sophia - Epidaure Project, 2004 Route des Lucioles BP 93
06902 Sophia Antipolis Cedex, France
Vincent.Arsigny@Polytechnique.org
[2] DOSISoft S.A., 45 Avenue Carnot, 94 230 Cachan, France

Abstract. In this article, we focus on the computation of statistics of invertible geometrical deformations (i.e., diffeomorphisms), based on the generalization to this type of data of the notion of *principal logarithm*. Remarkably, this logarithm is a simple 3D vector field, and can be used for diffeomorphisms close enough to the identity. This allows to perform *vectorial* statistics on diffeomorphisms, while preserving the invertibility constraint, contrary to Euclidean statistics on displacement fields.

Overview

In this article, which is an extended abstract of [1], we focus on the computation of statistics of general *diffeomorphisms*, i.e. of geometrical deformations (non-linear in general) which are both one-to-one and regular (as well as their inverse). To quantitatively compare non-linear registration algorithms, or in order to constrain them, computing statistics on *global* deformations would be very useful as was done in [6] with *local* statistics.

The computation of statistics is closely linked to the issue of the *parameterization* of diffeomorphisms. Many algorithms, as in [5], provide transformations are always diffeomorphic, and parameterize them via their displacement field. However, Euclidean means of displacement fields do not necessary yield invertible deformations, which makes Euclidean statistics on these parameters problematic for diffeomorphisms. In [7], it was proposed to parameterize arbitrary diffeomorphisms with Geodesic Interpolating Splines control points [4], and then to perform Euclidean operations on these low-dimensional parameters. However, although this guarantees the invertibility of the results, this may not be adequate for the whole variety of invertible transformations used in medical imaging.

To fully take into account the group structure of diffeomorphisms, it has been proposed to parameterize dense deformations with Hilbert spaces of *time-varying* speed vector fields, which yield geometrical deformations via the integration of an Ordinary Differential Equation (ODE) during one unit of time [8, 3]. In [9], it is suggested that the linear space of intial momenta of the geodesics of these spaces could provide an appropriate setting for statistics on diffeomorphisms.

However, this is illustrated in [9] only in the case of landmark matching. To our knowledge, this statistical framework has not been used yet in the *general case*, certainly because of the *iterative nature* of the computation of the mean in this setting, which requires very stable numerical algorithms to converge.

In this work, we introduce a novel parameterization of diffeomorphisms, based on the generalization of the *principal logarithm* to non-linear geometrical deformations. Interestingly, this corresponds to parameterizing diffeomorphisms with *stationnary* speed vectors fields. As for matrices, this logarithm can be used only for transformations close enough to the identity. However, our preliminary numerical experiments on 3D non-rigid registration suggest that this limitation affects only very large deformations, and may not be problematic for image registration results. This novel setting is the infinite-dimensional analogous of the Log-Euclidean framework proposed in [2] for tensors. In this framework, usual Euclidean statistics can be performed on diffeomorphisms via their logarithms, whith excellent mathematical properties like inversion-invariance.

In [1], our contributions are presented as follows. We first present the Log-Euclidean framework for diffeomorphisms, which is closely linked to the notion of *one-parameter subgroups*. Then, we present two efficient algorithms to compute the exponential of a vector field and the logarithm of a diffeomorphism, which are exemplified on synthetic data. Finally, we apply our framework to non-linear registration results to compute a Log-Euclidean mean deformation between a 3D atlas and a dataset of 9 T_1 MR images of human brains.

References

1. V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A Log-Euclidean framework for statistics on diffeomorphisms. In *Proc. of MICCAI'06*, 2006. To appear.
2. V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the Log-Euclidean framework. In *MICCAI (1)*, pages 115–122, 2005.
3. M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. Jour. Comp. Vis.*, 61(2):139–157, 2005.
4. V. Camion and L. Younes. Geodesic interpolating splines. In M. Figuiredo, J. Zerubia, and A. Jain, editors, *Proc. of Energy Minimization Methods in Comp. Vis. and Pat. Rec. (EMMCVPR;01)*, LNCS 2134, pages 513–527, 2001.
5. C. Chefd'hotel, G. Hermosillo, and O. Faugeras. Flows of diffeomorphisms for multimodal image registration. In *Proc. of ISBI*, 2002.
6. O. Commowick, R. Stefanescu, P. Fillard, V. Arsigny, N. Ayache, X. Pennec, and G. Malandain. Incorporating statistical measures of anatomical variability in atlas-to-subject registration for conformal brain radiotherapy. In *Proc. of MICCAI'2005 (II)*, LNCS, pages 927–934, 2005.
7. S. Marsland and C. J. Twining. Constructing diffeomorphic representations for the groupwise analysis of nonrigid registrations of medical images. *IEEE Trans. Med. Imaging*, 23(8):1006–1020, 2004.
8. A. Trouvé. Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision*, 28(3):213–221, 1998.
9. M. Vaillant, M. Miller, L. Younes, and A. Trouvé. Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 23:S161–S169, 2004.

Multivariate Statistics of the Jacobian Matrices in Tensor Based Morphometry and their application to HIV/AIDS

Natasha Lepore[1], Caroline A. Brun[1], Ming-Chang Chiang[1], Yi-Yu Chou[1], Rebecca A. Dutton[1], Kiralee M. Hayashi[1], Oscar L. Lopez[2], Howard J. Aizenstein[3], Arthur W. Toga[1], James T. Becker[2], and Paul M. Thompson[1]

[1] Laboratory of Neuro Imaging, Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095 USA
[2] Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15213 USA
[3] Department of Neurology, University of Pittsburgh, Pittsburgh, PA 15213 USA †

ABSTRACT

Tensor-based morphometry (TBM) is widely used in computational anatomy as a means to understand shape variation between structural brain images. A 3D nonlinear registration technique is typically used to align all brain images to a common neuroanatomical template, and the deformation fields are analyzed statistically to identify group differences in anatomy. However, the differences are usually computed solely from the determinants of the Jacobian matrices J that are associated with the deformation fields computed by the registration procedure. Thus, much of the information contained within those matrices gets thrown out in the process. Only the magnitude of the expansions or contractions is examined, while the anisotropy and directional components of the changes are ignored. Here we remedy this problem by computing multivariate shape change statistics using the strain matrices, defined as $(J^T J)^{1/2}$. As the strain matrices belong to the space of positive-definite matrices, we first transform them into a vector space using the 'Log-Euclidean metric' [1]. We study the brain morphology of 26 HIV/AIDS patients and 14 matched healthy control subjects using our method. The goal of the work was to find out whether multivariate statistics on the strain tensor afforded additional power in detecting anatomical differences between patients and controls.

The images are registered using a high-dimensional 3D fluid registration algorithm, which optimizes the Jensen-Rényi divergence [2], an information-theoretic measure of image correspondence. A pixelwise Hotelling T^2 test is used as a measure of variation between patients and controls. To assess the difference between

† This research was supported by the National Institute on Aging (AG021431 and AG016570), the National Library of Medicine, the National Institute for Biomedical Imaging and Bioengineering, the National Center for Research Resources (LM05639, EB01651, RR019771), and a Research Scientist Development Award - Level II (MH01077)

our results and the ones found from the determinant of the Jacobian, these results are compared to the one-dimensional Student's t test on the determinant of the Jacobian matrices. Patterns of detected white matter atrophy were of greater spatial extent and the corresponding effect size was greater in the case of Hotelling's T^2 test, indicating its better sensitivity. Group differences in brain structure between AIDS patients and healthy subjects are visible throughout the brain, with the greatest effect sizes in the corpus callosum and the basal ganglia. The cortical region is noisier, perhaps because the registration method is intensity-based and does not perform as well in that area. The anatomical profile of group differences is in line with studies using traditional volumetric methods, as the HIV virus is known to cause widespread neuronal loss and corresponding atrophy of the gray and white matter, especially in subcortical regions. Multivariate statistics on matrix-valued measures derived from deformation fields may therefore provide greater power to detect structural differences in the brain than more conventional methods that use the Jacobian determinant alone.

References

1. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic Resonance in Medicine (2006). To appear. Preprint: INRIA RR 5584, 2005.
2. Chiang, M.C., Dutton, R.A., Hayashi, K.M., Lopez, O.L., Aizenstein, H.J., Toga, A.W., Becker, J.T., Thompson, P.M.: 3D pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry. NeuroImage submitted (2006)

Singular solutions, momentum maps and computational anatomy

Colin J. Cotter[1] and Darryl D. Holm[1,2]

[1] Department of Mathematics, Imperial College London, SW7 2AZ, UK
colin.cotter@imperial.ac.uk
[2] Los Alamos National Laboratory, Los Alamos, NM 87545 USA
d.holm@imperial.ac.uk, dholm@lanl.gov

Abstract. This paper describes the variational formulation of template matching problems of computational anatomy (CA); introduces the EPDiff evolution equation in the context of an analogy between CA and fluid dynamics; discusses the singular solutions for the EPDiff equation and explains why these singular solutions exist (singular momentum map). Then it draws the consequences of EPDiff for outline matching problem in CA and gives numerical examples.

"I shall speak of things . . . so singular in their oddity as in some manner to instruct, or at least entertain, without wearying." – Lorenzo da Ponte

1 Introduction

Computational Anatomy (CA) must measure and analyze a range of variations in shape, or appearance, of highly deformable structures. The problem statement for CA was formulated long ago [1]

> In a very large part of morphology, our essential task lies in the ***comparison of related forms*** rather than in the precise definition of each. . . . This process of comparison, of recognizing in one form a definite permutation or deformation of another, . . . lies within the immediate province of mathematics and finds its solution in . . . the Theory of Transformations. – D'Arcy Thompson, *On Growth and Form* (1917)

The pioneering work of Bookstein, Grenander and Bajscy [2–4] first took up this challenge by introducing a method called template matching. The past several years have seen an explosion in the use of template matching methods in computer vision and medical imaging that is fulfilling D'Arcy Thompson's expectation [5–19]. These methods enable the systematic measurement and comparison of anatomical shapes and structures in medical imagery. The mathematical theory of Grenander's deformable template models, when applied to these problems, involves smooth invertible maps (diffeomorphisms), as presented in this context in [9, 10, 18–21]. In particular, the template matching approach involves Riemannian metrics on the diffeomorphism group and employs their projections onto specific landmark shapes, or image spaces.

The problem for CA then becomes to minimize the distance between two images as specified in a certain representation space, V. Metrics are written so that the optimal path in V satisfies an evolution equation, which was first discovered in abstract form [22] and later called EPDiff when it arose in the Euler-Poincaré theory of optimal motion on smooth invertible mappings called diffeomorphisms, [23].

The EPDiff equation coincides with the Euler equation for ideal fluids in the case that the Riemannian metric for the distance between two images is the L^2 norm. Another type of norm on V (called the H^1 norm) arises in the theory of the fascinating nonlinear coherent solutions of shallow water waves called ***solitons***. Solitons interact with each other elastically, so they re-emerge unscathed from fully nonlinear collisions. EPDiff with the H^1 norm on V describes the peaked soliton solutions of the Camassa-Holm shallow water wave equation. As we shall see, the Camassa-Holm peakons arise from a general property of Hamiltonian systems called their ***momentum map***. A discussion of EPDiff and peakons in the particular case of template matching appears in [24].

In this paper, we shall draw parallels between the two endeavors of fluid dynamics and template matching for computational anatomy, by showing how the Euler-Poincaré theory of ideal fluids can be used to develop new perspectives in CA. In particular, we discover that CA may be informed by the concept of weak solutions, solitons and momentum maps for geodesic flows [24–26].

1.1 Problem & Approach for Computational Anatomy

Computational Anatomy (CA) compares shapes (graphical structures) by making a ***geodesic deformation*** from one shape to the another. Among these graphical structures, landmarks and image outlines in CA are found to be singular solutions of the geodesic ***EPDiff*** equation. A ***momentum map*** for singular solutions of EPDiff yields their canonical Hamiltonian formulation, which provides a ***complete parameterization*** of the landmarks and image outlines by their ***canonical positions and momenta***. The momentum map provides an ***isomorphism*** between landmarks (and outlines) for images and ***singular*** (weak) solutions of EPDiff. (These solutions are solitons in 1D.) This isomorphism provides for CA: (1) a complete and non-redundant data representation; (2) a dynamical paradigm in which image outlines interact by exchange of momentum; (3) methods for numerical simulation & data assimilation. Euler-Poincaré theory also provides a framework for unifying and extending the various approaches in CA.

Thus, the concept of momentum becomes important for CA, because momentum:
– Completes the representation of images (momentum of cartoons);
– Informs template matching of the possibility of soliton-like collisions and ***momentum exchange in image outline interactions***;
– Encodes the subsequent deformation into the ***initial locus and momentum*** of an image outline;
– Provides numerical simulation methods using the ***momentum map for***

right action as a data structure; and
– Accomplishes matching and data assimilation via the ***adjoint linear problem*** for template matching, using the ***initial momentum as a control variable***.
All of these momentum properties flow from the ***EPDiff*** equation.

Outline of the paper. Section 2 describes the template matching variational problems of computational anatomy, explains the analogy with fluid dynamics and introduces the fundamental EPDiff evolution equation. The singular solutions for the EPDiff equation (2.1) with diffeomorphism group G are discussed in section 3. They are, in particular, related to the outline matching problem in computer vision, examples of which are given in section 4.

2 Mathematical formulation of template matching for CA

2.1 Cost

Most problems in CA can be formulated as: ***Find the deformation path (flow) with minimal cost, under the constraint that it carries the template to the target.*** Such problems have a remarkable analogy with fluid dynamics. The ***cost*** assigned in template matching for comparing images $\mathcal{I}_0$ & $\mathcal{I}_1$ is specified as a functional

$$\text{Cost}(t \mapsto \varphi_t) = \int_0^1 \ell(\mathbf{u}_t)\, dt$$

defined on curves φ_t in a Lie group with tangents $\frac{d\varphi_t}{dt} = \mathbf{u}_t \circ \varphi_t$ and $\mathcal{I}_t = \varphi_t \cdot \mathcal{I}_0$. In what follows, the function $\mathbf{u}_t \mapsto \ell(\mathbf{u}_t) = \|\mathbf{u}_t\|^2$ will be taken as a squared functional norm on the space of velocity vectors along the flow. The Lie group property specifies the representation space for template matching as a manifold of smooth mappings, which may be differentiated, composed and inverted. The vector space of ***right invariant*** instantaneous velocities, $\mathbf{u}_t = (d\varphi_t/dt) \circ \varphi_t^{-1}$ forms the tangent space at the identity of the considered Lie group, and may be identified as the group's ***Lie algebra***, denoted $\mathfrak{g}$.

2.2 Mathematical analogy between template matching and fluid dynamics

(I) The frameworks of CA and fluid dynamics both involve a ***right-invariant*** stationary principle with action, or cost function. The main differences are that template matching is formulated as an optimal control problem whose cost function is designed for the application, while fluid dynamics is formulated as an initial value problem whose cost function is the fluid's kinetic energy.
(II) The geodesic evolution for both template matching and fluid dynamics is governed by the ***EPDiff equation*** [27, 21],

$$\Big(\frac{\partial}{\partial t} + \mathbf{u}\cdot\nabla\Big)\mathbf{m} + (\nabla\mathbf{u})^T\cdot\mathbf{m} + \mathbf{m}(\text{div}\,\mathbf{u}) = 0\,. \tag{2.1}$$

Here $\mathbf{u} = G * \mathbf{m}$, where $G*$ denotes convolution with the Green's kernel G for the operator Q_{op}, where

$$\mathbf{m} = \frac{\delta \ell}{\delta \mathbf{u}} =: Q_{op}\mathbf{u}$$

The operator Q_{op} is symmetric and positive definite for the cost defined by

$$\text{Cost}(t \mapsto \varphi_t) = \int_0^1 \ell(\mathbf{u}_t)\, dt = \frac{1}{2}\int_0^1 \|\mathbf{u}_t\|^2\, dt = \frac{1}{2}\int_0^1 \langle\, \mathbf{u}_t\,,\, Q_{op}\mathbf{u}_t\,\rangle\, dt$$

with L^2 pairing $\langle\, \cdot\,,\, \cdot\,\rangle$ whenever $\|\mathbf{u}_t\|^2$ is a norm.
(III) The flows in CA and fluid dynamics both evolve under a left group action on a linear representation space, $\mathcal{I}_t = \varphi_t \cdot \mathcal{I}_0$. They differ in the roles of their advected quantities, $a_t = a_0 \circ \varphi_t^{-1}$. The main difference is that image properties are passive and affect the template matching as a constraint in the cost function, while advected quantities may affect fluid flows directly, for example through the pressure, so as to produce waves.

2.3 How EPDiff emerges in CA

Choose the cost function for continuously morphing $\mathcal{I}_0$ into $\mathcal{I}_1$ as

$$\text{Cost}(t \mapsto \varphi_t) = \int_0^1 \ell(\mathbf{u}_t)\, dt = \int_0^1 \|\mathbf{u}_t\|^2\, dt\,,$$

where u_t is the velocity of the fluid deformation at time t and

$$\|\mathbf{u}_t\|^2 = \langle\, \mathbf{u}_t\,,\, Q_{op}\mathbf{u}_t\,\rangle\,,$$

and Q_{op} is our positive symmetric linear operator. Then, the momentum governing the process, $\mathbf{m}_t = Q_{op}\mathbf{u}_t$ with Green's function $G:\ \mathbf{u}_t = G * \mathbf{m}_t$ satisfies the EPDiff equation, (2.1). This equation arises in both template matching and fluid dynamics, and it informs both fields of endeavor.

2.4 Deriving EPDiff from Euler-Poincaré Reduction of Hamilton's principle

.

Euler-Poincaré Reduction starts with a right (or left) $G-$invariant Lagrangian $L : TG \to \mathbb{R}$ on the tangent bundle of a Lie group G. ***Right invariance*** of the Lagrangian may be written as

$$L(g(t), \dot{g}(t)) = L(g(t)h, \dot{g}(t)h)\,,\ \text{for all } h \in G$$

A $G-$invariant Lagrangian defined on TG possesses a symmetry-reduced Hamilton's principle defined on the Lie algebra $TG/G \simeq \mathfrak{g}$. Stationarity of the symmetry-reduced Hamilton's principle yields the ***Euler-Poincaré equations*** on the dual Lie algebra $\mathfrak{g}^*$. For $G = \text{Diff}$, this equation is ***EPDiff*** (2.1).

3 Outline matching & momentum measures

Problem statement for outline matching:
Given two collections of curves $c_1, \ldots, c_N$ and $C_1, \ldots, C_N$ in Ω, find a time-dependent diffeomorphic process $(t \mapsto \varphi_t)$ of minimal action (or cost) such that $\varphi_0 = \mathrm{id}$ and $\varphi_1(c_i) = C_i$ for $i = 1, \ldots, N$. The matching problem for the image outlines seeks ***singular momentum solutions*** which naturally emerge in the computation of geodesics.

3.1 Image outlines as Singular Momentum Solutions of EPDiff

For example, in the 2D plane, EPDiff has weak ***singular momentum solutions*** that are expressed as [25, 23, 26]

$$\mathbf{m}(\mathbf{x}, t) = \sum_{a=1}^{N} \int_s \mathbf{P}_a(t, s)\delta\big(\mathbf{x} - \mathbf{Q}_a(t, s)\big)\, ds\,, \tag{3.1}$$

where s is a ***Lagrangian coordinate*** defined along a set of N curves in the plane *moving with the flow* by the equations $\mathbf{x} = \mathbf{Q}_a(t, s)$ and supported on the delta functions in the EPDiff solution (3.1). Thus, the singular momentum solutions of EPDiff represent evolving "wavefronts" supported on delta functions defined along curves $\mathbf{Q}_a(t, s)$ with arclength coordinate s and carrying momentum $\mathbf{P}_a(t, s)$ at each point along the curve as specified by (3.1). These solutions exist in any dimension and they provide a means of performing CA matching for points (landmarks), curves and surfaces, in any combination.

3.2 Here is the Geometry – Leading to the Numerics

The basic observation that ties everything together in $n-$dimensions is the following:
Theorem (Holm and Marsden, [23]): EPDiff singular momentum solutions $T^*\mathrm{Emb}(S, \mathbb{R}^n) \to \mathfrak{g}^* :\ (\mathbf{P}, \mathbf{Q}) \to \mathbf{m}$ ***define a momentum map.***

It is beyond our scope here to explain either the proof of this theorem or the mathematics underlying momentum maps for diffeomorphisms. However, we summarize the main results for template matching, as follows:
– The embedded manifold S is the support set of the P's and Q's.
– The momentum map is for left action of the diffeomorphisms on S.
– The whole system is right invariant.
– Consequently, its momentum map for right action is conserved.
– These constructions persist for a certain class of numerical schemes.
– They apply in template matching for every choice of norm.

3.3 A familiar example of a momentum map

A ***momentum map*** $\mathbf{J} : T^*Q \mapsto \mathfrak{g}^*$ is a Hamiltonian for the canonical action of a Lie group G on phase space T^*Q. It is expressed in terms of the pairing

$\langle \cdot\,, \cdot \rangle : \mathfrak{g}^* \times \mathfrak{g} \mapsto \mathbb{R}$ as

$$\langle \mathbf{J}\,, \xi \rangle = \langle p\,,\, \pounds_\xi q \rangle =: \langle q \diamond p\,, \xi \rangle\,,$$

where $(q, p) \in T^*_q Q$ and the Lie derivative $\pounds_\xi q$ is the infinitesimal generator of the action of the Lie algebra element $\xi \in \mathfrak{g}$ on q in the manifold Q.

The standard example is $\pounds_\xi q = \xi \times q$ for $\mathbb{R}^3 \times \mathbb{R}^3 \mapsto \mathbb{R}^3$, with pairing $\langle \cdot\,, \cdot \rangle$ given by scalar product of vectors. The momentum map is then

$$\mathbf{J} \cdot \xi = p \cdot \xi \times q = q \times p \cdot \xi \Rightarrow \mathbf{J} = q \times p$$

This is ***angular momentum***, the Hamiltonian for phase-space rotations. The outlines of images may be parameterized as curves whose dynamics must be invariant under reparameterizing the arclengths that label those curves. This symmetry leads to a conserved momentum map, called the ***circulation*** along the curves. The analog of this conservation law for fluids is the classical Kelvin circulation theorem.

3.4 EPDiff dynamics informs optimal control for CA

CA must compare two geometric objects, and thus it is concerned with an ***optimal control problem***. However, the ***initial value problem*** for EPDiff also has ***important consequences for CA applications***.

- When matching two geometric structures, the ***momentum at time*** *t=0* ***contains all required information for reconstructing the target from the template***. This is done via ***Hamiltonian geodesic flow***.
- Being canonically conjugate, the momentum has exactly the same dimension as the matched structures, so there is ***no redundancy***.
- Right invariance mods out the relabeling motions from the optimal solution. This symmetry also yields a ***conserved momentum map***.
- Besides being one-to-one, the momentum representation is defined on a ***linear space***, being dual to the velocity vectors.

  This means one may, for example,:
  - study linear instability of CA processes,
  - take averages and
  - apply statistics to the space of image contours.

  The ***advantage*** is the ease of building, sampling and estimating statistical models on a ***linear space***.

3.5 Summary

We have identified ***momentum as a key concept*** in the representation of image data for CA and discussed analogies with fluid dynamics. The ***fundamental idea*** transferring from fluid dynamics to CA is the idea of ***momentum maps*** corresponding to group actions.

4 Numerical examples of outline matching

In this section we describe our new technique applying particle-mesh methods to the problem of matching outlines. First we describe the approach to calculating geodesics in the space of outlines.

Let $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$ be two embeddings of S^1 in $\mathbb{R}^2$ which represent two shapes, each a closed planar curve. We seek a 1-parameter family of embeddings $\boldsymbol{Q}(t) : S^1 \times [0,1] \to \mathbb{R}^2$ so that $\boldsymbol{Q}(0) = \boldsymbol{Q}_0$ and $\boldsymbol{Q}(1)$ matches $\boldsymbol{Q}_1$ (up to relabeling). $\boldsymbol{Q}(t)$ is found by minimizing the constrained norm of its velocity. To find the equation for $\boldsymbol{Q}$ we require extremal values of the action

$$A = \int_0^1 \frac{1}{2} L(\boldsymbol{u}) \,\mathrm{d}\, t + \int_0^1 \int_{S^1} \boldsymbol{P}(s,t) \cdot (\dot{\boldsymbol{Q}}(s,t) - \boldsymbol{u}(\boldsymbol{Q}(s,t))) \,\mathrm{d}\, t, \qquad L = \|\boldsymbol{u}(t)\|^2_{\mathfrak{g}},$$

i.e. we seek time-series of vector fields $\boldsymbol{u}(t)$ which are minimized in some norm subject to the constraint that $\boldsymbol{Q}$ is advected by the flow using the Lagrange multipliers $\boldsymbol{P}$ (which we call momentum). The minimizing solutions are

$$\frac{\delta L}{\delta \boldsymbol{u}} = \int_{S^1} \boldsymbol{P}(s,t) \delta(\boldsymbol{x} - \boldsymbol{Q}(s,t)) \,\mathrm{d}\, s, \tag{4.1}$$

$$\dot{\boldsymbol{P}}(s,t) = -\boldsymbol{P}(s,t) \cdot \nabla \boldsymbol{u}(\boldsymbol{Q}(s,t),t), \tag{4.2}$$

$$\dot{\boldsymbol{Q}}(s,t) = \boldsymbol{u}(\boldsymbol{Q}(s,t),t), \tag{4.3}$$

subject to $\boldsymbol{Q}(s,0) = \boldsymbol{Q}_0(s)$.

We note that equation (4.1) is the momentum map corresponding to the cotangent-lift of the action of vector fields $\boldsymbol{u}$ on embedded curves given by

$$\boldsymbol{Q} \mapsto \boldsymbol{u}(\boldsymbol{Q}).$$

For a suitable test function $\boldsymbol{w}$, we obtain

$$\frac{d}{dt}\langle \boldsymbol{w}, \boldsymbol{m} \rangle - \langle \nabla \boldsymbol{w}, \boldsymbol{u}\boldsymbol{m} \rangle + \langle \boldsymbol{w}, (\nabla \boldsymbol{u})^T \cdot \boldsymbol{m} \rangle = 0, \qquad \boldsymbol{m} = \frac{\delta L}{\delta \boldsymbol{u}},$$

which is the weak form of the EPDiff equation.

Now one must seek initial momentum $\boldsymbol{P}(s,0)$ which takes shape $\boldsymbol{Q}_0(s)$ to shape $\boldsymbol{Q}_1(s)$. To do this, we choose some functional J of the advected shape $\boldsymbol{Q}(1,s)$ which is minimized when $\boldsymbol{Q}(1,s)$ matches $\boldsymbol{Q}_1(s)$. Following [28], we describe the curves by singular densities:

$$\mu = \int_{S_1} \hat{\mu}(s) \delta(\boldsymbol{x} - \boldsymbol{Q}(1,s) \,\mathrm{d}\, s \,\mathrm{d}\, V(\boldsymbol{x}), \tag{4.4}$$

$$\eta = \int_{S_1} \hat{\eta}(s) \delta(\boldsymbol{x} - \boldsymbol{Q}_1(s) \,\mathrm{d}\, s \,\mathrm{d}\, V(\boldsymbol{x}), \tag{4.5}$$

and write $J = \|\mu - \eta\|^2_G$ where $\|\cdot\|^2_G$ is a norm for a densities in a reproducing kernel Hilbert space with kernel G. This approach means that we do not need to force particular points to be matched to each other on the shapes. This last problem can be solved by using a gradient algorithm, where the gradient of the residual error with respect to $\boldsymbol{P}(s,0)$ is calculated using the adjoint equation [29].

4.1 Numerical discretization

We use the Variational Particle-Mesh (VPM) method [30, 31] to discretize the equations (4.1-4.3), as follows: discretize the velocity on an Eulerian grid with n_g points and approximate $\|\boldsymbol{u}\|$ there; replace S^1 by representing the shape by a finite set of n_p Lagrangian particles $\{\boldsymbol{Q}_\beta\}_{\beta=1}^{n_p}$, and interpolate from the grid to the particles using basis functions

$$\boldsymbol{u}(\boldsymbol{Q}_\beta) = \sum_{k=1}^{n_g} \boldsymbol{u}_k \psi_k(\boldsymbol{Q}_\beta)\,, \quad \text{with} \quad \sum_{k=1}^{n_g} \psi_k(\boldsymbol{x}) = 1\,,\ \forall\ \boldsymbol{x}.$$

The action for the continuous time motion on the grid then becomes

$$A = \int_0^1 \frac{1}{2}\|\boldsymbol{u}(t)\|_{grid}^2 + \sum_\beta \boldsymbol{P}_\beta \cdot \left(\dot{\boldsymbol{Q}}_\beta - \sum_k \boldsymbol{u}_k \psi_k(\boldsymbol{Q}_\beta)\right) \mathrm{d}\, t,$$

and one can obtain a fully discrete method by discretizing the action in time. For example, we can obtain a first-order method by extremizing

$$A = \Delta t \sum_{n=1}^{N} \left(\frac{1}{2}\|\boldsymbol{u}^n\|_{grid}^2 + \sum_\beta \boldsymbol{P}_\beta^n \cdot \left(\frac{\boldsymbol{Q}_\beta^n - \boldsymbol{Q}_\beta^{n-1}}{\Delta t} - \sum_k \boldsymbol{u}_k^n \psi_k(\boldsymbol{Q}_\beta^{n-1}) \right) \right).$$

The resulting time-stepping method is the (first-order) symplectic Euler-A method for the time-continuous Hamiltonian system for the Lagrangian particles. In general, the method will always be symplectic since it arises from a discrete variational principle (see [32] for a broad introduction to symplectic numerical methods and their conservation properties). The conservation properties of VPM are discussed in [31].

We approximate the densities μ and η on the grid using the standard particle-mesh approach (see [33]):

$$\mu_k = \sum_\beta \hat{\mu}_\beta \psi_k(\boldsymbol{Q}_\beta^N), \qquad \eta_k = \sum_\beta \hat{\eta}_\beta \psi_k(\boldsymbol{Q}_{1,\beta}),$$

where $\boldsymbol{Q}_{1,\beta}$ are the positions of particles on the target shape. This amounts to "pixellating" the singular densities (4.4-4.5) on the grid. For a given kernel G, we approximate J with

$$J = \sum_{kl} G(\boldsymbol{x}_k - \boldsymbol{x}_l)(\mu_k - \eta_k)(\mu_l - \eta_l).$$

The discrete adjoint is then applied in computing the inversion for the initial conditions for $\boldsymbol{P}_\beta$ which generate the flow. A numerical example calculated using this method is given in figure 1.

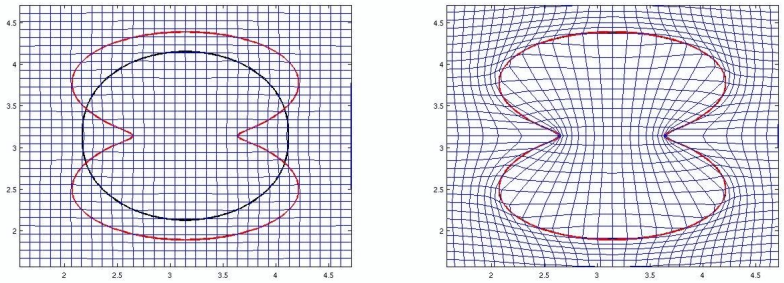

Fig. 1. Results from a VPM calculation to calculate the minimal path between a two simple shapes. On the left, the initial and final shapes are shown, and on the right, the deformation of the initial shape into the final shape is depicted together with a grid which shows how the flow map deforms the space around the shape. We used the H^1 norm for velocity on a $2\pi \times 2\pi$ periodic domain on a 128×128 grid, discretized using FFT, and the corresponding kernel was used to calculate J. Cubic B-splines were used as basis functions.

Conclusion

We have presented the formulation and numerical implementation of D'Arcy Thompson's vision of how transformation theory would provide a natural means of anatomical comparison, of recognizing shape and form. For this, we have identified momentum as a key concept in the representation of image data for CA and have discussed the analogies of CA with fluid dynamics. The fundamental idea that transfers from fluid dynamics to CA is the idea of momentum maps corresponding to group actions. The analogy of CA with fluid dynamics also suggests that much remains to be gained by understanding modern methods in fluid dynamics and seeking ways to transfer these ideas to CA.

In addition, we have illustrated the power of this method by solving a simple but challenging problem of exact matching of two closed contours in the plane. This example is deceptively simple looking, in fact the computation of exact matching had never been done before and is still a challenging problem for computational anatomy.

Acknowledgments

This work was partially supported by US DOE, under contract W-7405-ENG-36 for Los Alamos National Laboratory, and Office of Science ASCAR/AMS/MICS.

References

1. Thompson, D.: On Growth and Form. (1917)

2. Bajcsy, R., Lieberson, R., Reivich, M.: A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. J. Comput. Assisted Tomogr. **7** (1983) 618–625
3. Bookstein, F.L.: Morphometric tools for landmark data; geometry and biology. Cambridge University press (1991)
4. Grenander, U.: General Pattern Theory. Oxford Science Publications (1993)
5. Ashburner, J., Csernansky, J.G., Davatzikos, C., Fox, N.C., Frisoni, G.B., Thompson, P.M.: Computer-assisted imaging to assess brain structure in healthy and diseased brains. Lancet Neurology **2** (2003) 79–88
6. Grenander, U., Miller, M.I.: Computational anatomy: An emerging discipline. Quarterly of Applied Mathematics **LVI**(4) (1998) 617–694
7. Hallinan, P.: A low dimensional model for face recognition under arbitrary lighting conditions. In: Proceedings CVPR'94. (1994) 995–999
8. Jain, A.K., Zhong, Y., Dubuisson-Jolly, M.P.: Deformable template models: a review. Signal Processing **71** (1998) 109–129
9. Miller, M.I., Trouvé, A., Younes, L.: On metrics and Euler-Lagrange equations of computational anatomy. Ann. Rev. Biomed. Engng **4** (2002) 375–405
10. Miller, M.I., Younes, L.: Group action, diffeomorphism and matching: a general framework. Int. J. Comp. Vis. **41** (2001) 61–84
11. Montagnat, J., Delingette, H., Ayache, N.: A review of deformable surfaces: topology, geometry and deformation. Image and Vision Computing **19** (2001) 1023–1040
12. Mumford, D.: Mathematical theories of shape: Do they model perception ? In: Proc. SPIE Workshop on Geometric Methods in Computer Vision. Volume 1570. SPIE, Bellingham, WA (1991) 2–10
13. Mumford, D.: Elastica and computer vision. In: Algebraic Geometry and its Applications. Springer-Verlag, New York, NY (1994) 507–518
14. Mumford, D.: Pattern theory: a unifying perspective. In: Perception as Bayesian Inference. Cambridge University Press, Cambridge, UK (1996) 25–62
15. Thompson, P.M., Giedd, J.N., Woods, R.P., MacDonald, D., Evans, A.C., Toga, A.W.: Growth patterns in the developing brain detected by using continuum mechanical tensor maps. Nature **404**(6774) (2000) 190–3.
16. Toga, A.W., ed.: Brain warping. Academic Press (1999)
17. Toga, A.W., Thompson, P.M.: Temporal dynamics of brain anatomy. Ann. Rev. Biomedical Engng. **5** (2003) 119–145
18. Trouvé, A.: An infinite dimensional group approach for physics based model. Technical report (electronically available at http://www.cis.jhu.edu) (1995)
19. Trouvé, A.: Diffeomorphism groups and pattern matching in image analysis. Int. J. of Comp. Vis. **28**(3) (1998) 213–221
20. Dupuis, P., Grenander, U., Miller, M.I.: Variational problems on flows of diffeomorphisms for image matching. Quarterly of Applied Math. **56** (1998) 587–600
21. Mumford, D.: Pattern theory and vision. In: Questions Mathmatiques En Traitement Du Signal et de L'Image. Institut Henri Poincar (1998) 7–13
22. Arnold, V.I.: Mathematical methods of Classical Mechanics. Springer (1978) Second Edition: 1989.
23. Holm, D.D., Marsden, J.E.: Momentum maps & measure valued solutions of the Euler-Poincar equations for the diffeomorphism group. Progr. Math. **232** (2004) 203–235 http://arxiv.org/abs/nlin.CD/0312048.
24. Holm, D.D., Rananather, J.T., Trouvé, A., Younes, L.: Soliton dynamics in computational anatomy. NeuroImage **23** (2004) S170–178 http://arxiv.org/abs/nlin.SI/0411014.

25. Camassa, R., Holm, D.D.: An integrable shallow water equation with peaked solitons. Phys. Rev. Lett. **71** (1993) 1661–1664
26. Holm, D.D., Staley, M.F.: Interaction dynamics of singular wavefronts in nonlinear evolutionary fluid equations. Technical report, Los Alamos National Laboratory (2004) In preparation, see http://cnls.lanl.gov/ staley/.
27. Holm, D.D., Marsden, J.E., Ratiu, T.S.: The Euler–Poincar equations and semidirect products with applications to continuum theories. Adv. in Math. **137** (1998) 1–81 http://xxx.lanl.gov/abs/chao-dyn/9801015.
28. Glaunes, J., Trouve, A., Younes, L.: Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2. (2004) 712–718
29. Gunzburger, M.D.: Perspectives in Flow Control and Optimization. Advances in design and control. SIAM, Philadelphia, USA (2003)
30. Cotter, C.J.: A general approach for producing Hamiltonian numerical schemes for fluid equations. arXiv:math.NA/0501468 (2005)
31. Cotter, C.J., Holm, D.D.: Discrete momentum maps for lattice EPDiff. arXiv:math.NA/0602296 (2006)
32. Leimkuhler, B., Reich, S.: Simulating Hamiltonian Dynamics. CUP (2005)
33. Frank, J., Gottwald, G., Reich, S.: A Hamiltonian particle-mesh method for the rotating shallow-water equations. In: Lecture Notes in Computational Science and Engineering. Volume 26. Springer-Verlag (2002) 131–142

Template estimation form unlabeled point set data and surfaces for Computational Anatomy

Joan Glaunès[1] and Sarang Joshi[2]

[1] Center for Imaging Science, Johns Hopkins University, joan@cis.jhu.edu
[2] SCI, University of Utah, scjoshi@gmail.com

Abstract. A central notion in Computational Anatomy is the generation of registration maps,mapping a large set of anatomical data to a common coordinate system to study intra-population variability and inter-population differences. In previous work [1, 2] methods for estimating the common coordinate system or the template given a collection imaging data were presented based on the notion of Fréchet mean estimation using a metric on the space of diffeomorphisms. In this paper we extend the methodology to the estimation of a template given a collection of unlabeled point sets and surfaces. Using a representation of points and surfaces as currents a Reproducing Kernel Hilbert Space (RKHS) norm is induced on the space of Borel measures. Using this norm and a metric on the space of diffeomorphisms the template estimation problem is possed as a minimum mean squared error estimation problem. An efficient alternating conjugate gradient decent algorithm is derived and results exemplifying the methodology are presented.

1 Introduction

A major focus of computational anatomy has been the development of image mapping algorithms [3] that can map and transform a single brain atlas on to a population. Most digital brain atlases currently being used in computational anatomy are based on a single subjects anatomy. Although these atlases provide a standard coordinate system, and form the template in deformable template setting, they are limited because a single anatomy cannot faithfully represent the complex structural variability evident in a population.

Construction of atlases is a key procedure in population based medical image analysis. In the paradigm of computational anatomy the atlas serves as a deformable template[4]. The deformable template can project detailed atlas data such as structural, biochemical, functional as well as vascular information on to the individual or an entire population of brain images. The transformations encode the variability of the population under study. Statistical analysis of the transformations can be used to characterize different populations [5]. For a detailed review of deformable atlas mapping and the general framework for computational anatomy see [3].

1.1 Unbiased Large Deformation Atlas Estimation via minimum mean squared error estimation.

For representations in which the underlying geometry is parameterized as a Euclidean vector space, training data can be represented as a set of vectors $x_1, \cdots, x_N$ in a vector space V . In a vector space, with addition and scalar multiplication well defined, an average representation of the training set can be computed as the linear average $\mu = \frac{1}{N}\sum_i^N x_i$. When studying statistical geometric properties of anatomy, the transformations representing the variability of a population are parameterized via diffeomorphic transformations of the ambient space. In the group of diffeomorphisms, the addition of two diffeomorphisms is not generally a diffeomorphism and, hence, a template based on linear averaging of transformations is not well defined. Fréchet [6] extended the notation of averaging to general metric spaces via minimum mean squared error estimation. For a general metric space M, with a distance $d : M \times M \rightarrow R$, the intrinsic mean for a collection of data points $x_i \in M$ can be defined as the minimizer of the sum-of-squared distances to each of the data points. That is

$$\mu = \arg\min_x \frac{1}{N}\sum_i^N d(x, xi)^2 \ .$$

This approach, combined with the mathematical metric theory of diffeomorphisms developed in [7, 8] , represents the core of the atlas estimation methodology. This approach has been applied previously to a sets of images, and can be stated as the following estimation problem. Given a metric, $D : S \times S \rightarrow R$ on a group of transformations, along with an image dissimilarity metric $E(I1, I2)$, we wish to find the image $\hat{I}$ such that

$$\{\hat{\varphi}_i, \hat{I}\} = \arg\min_{\varphi_i, I}, \frac{1}{N}\sum_i^N E(I_i \circ \varphi_i, I)^2 + D(e, \varphi_i)^2$$

where e is the identity transformation. In this paper we follow the same basic paradigm for unlabeled point sets and surfaces.

Previous work on diffeomorphic construction of atlases from unlabeled point sets presented in [9, 10] take a clustering approach. In this paper rather than explicitly clustering point sets, geometric measure theory is used to directly define a dissimilarity metrics for unlabel point sets following [11].

The paper is organized as follows: In section 2, geometric measure theory is used to define dissimilarity metrics for unlabel point sets and surfaces. For completeness in section 4 the metric on the space of deffiomorphisims is reviewed. In section 5 the template estimation problem given a collection of unlabeled point sets and surfaces is formulated. Finally in sections **??** 7 implementation details and results are presented.

2 Hilbert norms for dissimilarity metric for point set and surface data.

2.1 Dissimilarity metric for unlabeled point sets.

We follow the novel framework proposed in [11] for measuring dissimilarities between unlabeled point sets. Given a set of unlabeled landmark points $\{x_p\}_{p=1}^n$ in $\mathbb{R}^d$, a vector space structure is induced on the space of all such unlabeled point sets by modeling them as a weighted sum of Dirac measures centered at each of the points that is: $\sum_{p=1}^n a_p \delta_{x_p}$. The weights $a_p \in \mathbb{R}$, such that $\sum_p a_p = 1$, are user-defined and capture the relative confidence associated with each landmark (without such information, one should simply set $a_p = \frac{1}{n}$ for each index p). The space of Dirac measures forms a vector space where point sets can be added and subtracted. A Dissimilarity metric is induced by a Reproducing Kernel Hilbert norm based on a user-defined kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. The scalar product of two Dirac measures δ_x, δ_y is defined by the rule $\langle \delta_x, \delta_y \rangle_K = K(x, y)$, which leads to the following expression for the norm of the sum of Diracs:

$$\left\| \sum_{p=1}^n a_p \delta_{x_p} \right\|_K^2 = \sum_{p=1}^n \sum_{p'=1}^n a_p a_{p'} K(x_p, x_{p'}). \tag{1}$$

Dissimilarity between two points sets is given by the norm of the difference of the corresponding sums of Diracs:

$$\begin{aligned} \left\| \sum_{p=1}^n a_p \delta_{x_p} - \sum_{q=1}^m b_q \delta_{y_q} \right\|_K^2 &= \sum_{p=1}^n \sum_{p'=1}^n a_p a_{p'} K(x_p, x_{p'}) \\ &\quad - 2 \sum_{p=1}^n \sum_{q=1}^m a_p b_q K(x_p, y_q) + \sum_{q=1}^m \sum_{q'=1}^m b_q b_{q'} K(y_q, y_{q'}). \end{aligned} \tag{2}$$

These formulas arise naturally as the expression of the dual norm of the functional Hilbert space corresponding to the reproducing kernel K. We refer to [11] for all the mathematical details. In this setting, this dual norm is also defined, by mathematical completion with respect to the norm defined by Eqn. 1, for a larger class which include all signed Borel measures of $\mathbb{R}^d$. The K-norm of a measure μ is given by

$$\|\mu\|_K^2 = \iint K(x, y) d\mu(x) d\mu(y). \tag{3}$$

The action of a deformation map $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ on a Borel measure μ satisfies $\varphi\mu(A) = \mu(\varphi^{-1}(A))$ for any subset $A \subset \mathbb{R}^d$. When μ is a sum of Dirac masses, it consists in moving all positions of points by φ, leaving the weights unchanged: $\varphi \left(\sum_{p=1}^n a_p \delta_{x_p} \right) = \sum_{p=1}^n a_p \delta_{\varphi(x_p)}$.

3 Hilbert norms for dissimilarity metric between surfaces

In [12], an extension of the above described method has been proposed to measure dissimilarities between oriented surfaces embedded in $\mathbb{R}^3$. We recall its main point here. A convenient mathematical model of oriented m-dimensional submanifolds in $\mathbb{R}^d$ with $m < d$, is the concept of "current", which comes from Geometric Measure Theory [13, 14]. Currents are generalizations of Schwartz distributions, where smooth test functions are replaced by smooth m-differential forms in $\mathbb{R}^d$, i.e. smooth maps ω that associate to each point $x \in \mathbb{R}^d$ a skew-symmetric m-multilinear form, or m-covector, $\omega(x)$. Hence currents can be viewed as generalized m-forms, or as linear functionals acting on m-forms. The current associated to an oriented m-submanifold S is the linear functional $[S]$ defined by $[S](\omega) = \int_S \omega$. When S is an oriented surface in $\mathbb{R}^3$, $[S]$ is nothing but the vector valued Borel measure corresponding to the collection of unit-normal vectors to S, distributed with density equal to the element of surface area ds and can be written as $\eta(y)ds(y)$, where $\eta(y)$ is the normal and $ds(y)$ is the surface measure at point $y \in S$.

In this setting, similarly to what was described above for scalar measures, one can introduce a matrix-valued kernel $K(x, y)$ and define the K-norm of $[S]$ by

$$\left\|[S]\right\|_K^2 = \int_S \int_S \eta(y)^* K(x,y)\eta(x)\; ds(x)ds(y). \tag{4}$$

In practice we use kernels of the type $k(x, y)I$ where I is a 3×3 identity metrix and $k(x, y)$ a scalar kernal of the Gaussian or the Cauchy type. Then the integrand in Eqn. 4 becomes $K(x, y)\,\langle \eta(x), \eta(y)\rangle$. When S is a triangular mesh, a good approximation of this formula can be computed by replacing $[S]$ by a sum of vector-valued Dirac masses:

$$\left\|\sum_{f=1}^{nf} \eta(f)\delta_{c(f)}\right\|_K^2 = \sum_{f=1}^{nf}\sum_{f'=1}^{nf} \eta(f')^*\; K(c(f), c(f'))\; \eta(f), \tag{5}$$

where nf is the number of faces of the triangulation, and for any face f with vertices x, y, z, $c(f)$ is its center and $\eta(f)$ its non-normalized normal vector with the length capturing the area of each triangular patch:

$$c(f) = \frac{1}{3}(x + y + z), \qquad \eta(f) = \frac{1}{2}(y - x) \times (z - x). \tag{6}$$

The sum $\sum_{f=1}^{nf} \eta(f)\delta_{c(f)}$ does not correspond to a surface, but it is close to $[S]$ in the space of currents.

One can mathematically define an action of deformation maps $\varphi : \mathbb{R}^3 \to \mathbb{R}^3$ on currents, called push-forward, which is consistent to the intuitive action on subset of $\mathbb{R}^3$. This means that for a surface S, the push-forward $\varphi[S]$ is exactly the current associated to the deformed surface: $\varphi[S] = [\varphi(S)]$. For a sum of

Diracs, the push-forward action becomes:

$$\varphi\left(\sum_{f=1}^{nf}\eta(f)\delta_{c(f)}\right)=\sum_{f=1}^{nf}|d\varphi|^{-1}(d\varphi^{*})^{-1}\eta(f)\delta_{\varphi(c(f))}, \tag{7}$$

where $d\varphi$ is the Jacobian matrix of φ and $|d\varphi|$ its determinant, evaluated at $c(f)$.

4 Diffeomorphic metric mapping

Having defined metrics between two unlabeled point sets and surfaces for completeness we briefly review metric on the space of Diffeomorphic transformations. To generate dense deformation maps in $\mathbb{R}^d$ we use the large deformation framework [8] which consist of integrating time-dependent velocity fields in $\mathbb{R}^d$. The corresponding flow equation is given by

$$\frac{\partial\varphi^{v}(t,x)}{\partial t}=v(t,\varphi^{v}(t,x)), \tag{8}$$

with $\varphi(0,x)=x$, and we define $\varphi(x):=\varphi^{v}(1,x)$, which is a one-to-one map in $\mathbb{R}^d$ (diffeomorphism). To ensure regularity of these maps, an energy functional is defined on velocity fields:

$$\|v(t,\cdot)\|_{V}^{2}=\int_{\mathbb{R}^d}\langle Lv(t,x),Lv(t,x\rangle\,dx, \tag{9}$$

where L is a differential operator acting on vector fields. Minimality constraints on this energy are included in the matching variational problems. Moreover, this energy also defines a distance in the group of diffeomorphisms:

$$D^{2}(e,\varphi)=\inf_{v,\varphi^{v}(1,\cdot)=\varphi}\int_{0}^{1}\|Lv(t)\|_{V}^{2}\,dt. \tag{10}$$

As was noticed in [15, 11, 12], the practical use of such models is simplified, in the context of point-based matching methods, by the fact that optimal vector fields take the form:

$$v(t,x)=\sum_{p=1}^{n}G(x_p(t),x)\alpha_p(t), \tag{11}$$

where $x_p(t)=\varphi^{v}(t,x_p)$ are the trajectories of the control points x_p, $G(x,y)$ is the Green kernel of operator $L^{*}L$, and $\alpha_p(t)\in\mathbb{R}^d$ are unknown variables called momentum vectors. From 8 and 11, one derives the equations which link trajectories and momentum vectors:

$$\frac{dx_p}{dt}(t)=\sum_{p_1=1}^{n}G(x_{p_1}(t),x_p(t))\alpha_{p_1}(t). \tag{12}$$

Hence computing momentum vectors from trajectories requires to solve T linear systems, where T is the number of time discretization steps, while computing trajectories from momentum vectors requires to solve a first order ODE, which can be preformed at a lower cost. Therefore, the choice of momentum vectors as variables of minimization is preferred in practice. The reformulation of the energy functional in terms of $x_p(t)$ and $\alpha_p(t)$ becomes:

$$\|v(t,\cdot)\|_V^2 = \sum_{p=1}^{n} \left\langle \frac{dx_p}{dt}(t), \alpha_p(t) \right\rangle . \tag{13}$$

5 Atlas construction for points sets and surfaces

Having defined the metrics on unlabeled point sets, surfaces and diffeomorphic transformations we are ready to apply the recipe of minimum mean squared estimation to the atlas construction problem.

5.1 Formulation for point sets

Let $\{x_{ip}\}$, $1 \leq i \leq N$, $1 \leq p \leq n_i$ be N unlabeled point sets in $\mathbb{R}^d$, and $a_{ip} \in \mathbb{R}$ the associated weights (e.g. $a_{ip} = \frac{1}{n_i}$). Let $\mu_i = \sum_{p=1}^{n_i} a_{ip}\delta_{x_{ip}}$ be the borel measure representation of each of the point sets. The template estimation is now defined as the following minimum mean squared estimation problem:

$$\{\hat{\varphi}_i, \hat{\mu}\} = \arg\min_{\varphi_i, \mu} \sum_{i=1}^{N} \left\{ \left\| \mu - \varphi_i \mu_i \right\|_K^2 + D^2(e, \varphi_i) \right\}, \tag{14}$$

where $D(e, \varphi_i)$ is the metric on the space of diffeomorphic mappings φ_i, described in section 4. This problem is simplified by the following remark: for fixed φ_i, the Borel measure $\hat{\mu}$ which minimizes 14 is the average of $\varphi_i \mu_i$:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \varphi_i \mu_i = \frac{1}{N} \sum_{i=1}^{N} \sum_{p=1}^{n_i} a_{ip} \delta_{\varphi_i(x_{ip})}. \tag{15}$$

It is a sum of Dirac masses associated to the union of all points $\varphi_i(x_{ip})$. Consequently, minimization of 14 can be done with respect to deformation maps $\varphi_i(t)$ only:

$$\{\hat{\varphi}_i\} = \arg\min_{\varphi_i} \sum_{i=1}^{N} \left\{ \left\| \left(\frac{1}{N} \sum_{i=1}^{N} \varphi_i \mu_i \right) - \varphi_i \mu_i \right\|_K^2 + D^2(e, \varphi_i) \right\}. \tag{16}$$

5.2 Formulation for surfaces

Let S_i be N surfaces in $\mathbb{R}^3$. Let $[S_i]$ denote either the current corresponding to S_i, or its approximation by a finite sum of vectorial Diracs. Note again that

this sum is itself a current, and therefore the following is well defined for the continuous problem or its discretization. Using the metric defined on currents and diffeomorphic mappings in sections 2 and 4, the minimum mean squared error template estimation problem can now be formulated as:

$$\{\hat{\varphi}_i, [\hat{S}]\} = \arg\min_{\varphi_i, [S]} \sum_{i=1}^{N} \left\{ \left\| [S] - \varphi_i[S_i] \right\|_K^2 + D^2(e, \varphi_i) \right\}. \tag{17}$$

For fixed φ_i, the minimizer $[S]$ of 17 is the average of $\varphi_i[S_i]$: $[\hat{S}] = \frac{1}{N} \sum_{i=1}^{N} \varphi_i[S_i]$. This average $[S]$ is a current corresponding not to a single surface, but to the union of the deformed surfaces $\varphi_i(S_i)$, each weighted by $\frac{1}{N}$. One can think about this as a sort of fuzzy surface and the exact interpretation of this is beyond the scope of this paper. However since it is a well defined current, the registration process of all the surfaces to the average current is well defined. The reformulation of 17, in terms of minimization with respect to the diffeomorphisms φ_i only, becomes:

$$\{\hat{\varphi}_i\} = \arg\min_{\varphi_i} \sum_{i=1}^{N} \left\{ \left\| \left(\frac{1}{N} \sum_{i=1}^{N} \varphi_i[S_i] \right) - \varphi_i[S_i] \right\|_K^2 + D^2(e, \varphi_i) \right\}. \tag{18}$$

6 Implementation

We now describe in detail an implementation for performing minimizations of 16 and 18 using an alternating algorithm which estimates, on a per iteration basis, each φ_i in turn, analogous to the method described in [2] for images. At each step, minimization is performed with respect to φ_i alone, the other deformation maps φ_j, $j \neq i$, being fixed, which transforms the whole averaging process into a sequence of source-to-target matchings, for which the algorithms described in [11] and [12] can be directly applied. More precisely, at each step of this sequence of matchings, we perform minimization of a functional involving an energy term and an end-point matching term:

$$J\left(\{\alpha_{ip}(t)\}_{1\leq p\leq n_i}^{t\in[0,1]}\right) = \int_0^1 \sum_{p=1}^{n_i} \left\langle \frac{dx_{ip}}{dt}(t), \alpha_{ip}(t) \right\rangle \, dt + A(\{x_{ip}(1)\}). \tag{19}$$

Assuming that G is a function of the squared distance: $G(x, y) = G(|x - y|^2)$, the gradient becomes

$$\nabla J_{ip}(t) = 2\alpha_{ip}(t) + \beta_{ip}(t), \tag{20}$$

where $\beta_{ip}(t)$ is solution to the following ODE:

$$\frac{d\beta_{ip}}{dt}(t) = -2 \sum_{p_1=1}^{n} G'(|x_{ip_1}(t) - x_{ip}(t)|^2) \Big\{ \langle \alpha_{ip}(t), \beta_{ip_1}(t) \rangle + \langle \alpha_{ip_1}(t), \beta_{ip}(t) \rangle + 2 \langle \alpha_{ip}(t), \alpha_{ip_1}(t) \rangle \Big\} (x_{ip}(t) - x_{ip_1}(t)), \tag{21}$$

with $\beta_{ip}(1) = \nabla_{x_{ip}(1)} A$.

The matching term A, in the case of point sets, is given by:

$$A(\{x_{ip}(1)\}) = \|\varphi_i \mu_i - \hat{\mu}\|_K^2 = \sum_{p=1}^{n_i} \sum_{p'=1}^{n_i} a_{ip} a_{ip'} K(x_{ip}(1), x_{ip'}(1)) \\ - 2 \sum_{p=1}^{n_i} \sum_{q=1}^{m} a_{ip} b_q K(x_{ip}(1), y_q) + \sum_{q=1}^{m} \sum_{q'=1}^{m} b_q b_{q'} K(y_q, y_{q'}), \quad (22)$$

where points y_q and weights b_q denote the average template $\hat{\mu}$ as follows:

$$\sum_{q=1}^{m} b_q \delta_{y_q} := \hat{\mu} = \frac{1}{N} \sum_{j=1}^{N} \sum_{p=1}^{n_j} a_{jp} \delta_{\varphi_j(x_{jp})}. \quad (23)$$

The gradient of this term, required for the computation of Eqn. 20, becomes:

$$\nabla_{x_{ip}(1)} A = 2 a_{ip} \left(\sum_{p'=1}^{n} a_{ip'} \nabla_{x_{ip}(1)} K(x_{ip}(1), x_{ip'}(1)) - \right. \\ \left. \sum_{q=1}^{m} b_q \nabla_{x_{ip}(1)} K(x_{ip}(1), y_q) \right). \quad (24)$$

We now follow the same recipe in the case of surfaces. Denote $\sum_{f_i=1}^{nf_i} \eta(f_i) \delta_{c(f_i)}$ the sum of vectorial Diracs approximating each of the triangular meshs given by moving all vertices x_{ip} of S_i under the deformation map φ_i, and

$$\sum_{g=1}^{m} \eta(g) \delta_{c(g)} := \frac{1}{N} \sum_{i=1}^{N} \sum_{f_i=1}^{nf_i} \eta(f_i) \delta_{c(f_i)} \quad (25)$$

the averaged template. The matching term becomes:

$$A(\{x_{ip}(1)\}) = \sum_{f_i=1}^{nf_i} \sum_{f_i'=1}^{nf_i} \eta(f_i)^* K(c(f_i), c(f_i')) \eta(f_i') \\ - 2 \sum_{f_i=1}^{nf_i} \sum_{g=1}^{m} \eta(f_i)^* K(c(f_i), c(g)) \eta(g) + \sum_{g=1}^{m} \sum_{g'=1}^{m} \eta(g)^* K(c(g), c(g')) \eta(g'). \quad (26)$$

We recall that $x_{ip}(1) = \varphi_i(x_{ip})$ where x_{ip} are the vertices of the triangular mesh S_i. Centers and normal vectors are computed from the vertices using formulas in Eqn. 6. Finally, we derive the gradient of term A in this case. If $x_{ip}(1)$ is a

vertex of face f_i, the contribution of f_i to the gradient at $x_{ip}(1)$ is given by

$$\sum_{f_i'=1}^{n_i} e(f_i) \times K(c(f_i'), c(f_i))\eta(f_i') - \sum_{g=1}^{m} e(f_i) \times K(c(g), c(f_i))\eta(g) + \frac{2}{3}\sum_{f_i'=1}^{n_i} \frac{\partial K(c(f_i), c(f_i'))}{\partial c(f_i)}\eta(f_i) - \frac{2}{3}\sum_{g=1}^{m} \frac{\partial K(c(f_i), c(g))}{\partial c(f_i)}\eta(f_i), \quad (27)$$

and gradient $\nabla_{x_{ip}(1)} A$ is obtained by summing all contributions of faces f_i which share $x_{ip}(1)$ as a vertex.

Having computed the gradients, a conjugate gradient method is used to perform minimization of functional J on variables $\alpha_p(t)$ evaluated on finite number of time steps regularly spaced between 0 and 1, and a centered corrector scheme is applied to solve ODE 21, required for each computation of gradient 20. To speed up computations when a large number of control points is involved, we use multipole[16] methods for convolutions with kernels G and K.

7 Results

We now present results form applying the algorithms described previously. Figure 1 shows a synthetic experiment of averaging two point sets in $\mathbb{R}^2$, which are drawn from a circle and another oblong closed curve(shown in the left collum). Shown in the middle collum are the results of applying the estimated deformation to the two data sets. Shown in the right collum is the average estimated template.

Figure 2 shows a example of averaging three segmented surfaces of hippocampus in $\mathbb{R}^3$. Shown in the top row are three surfaces of hippocampi from three different subjects. The middle row shows the estimated deformation applied to each of the surfaces. Shown at the bottom is the estimated template.

8 Discussion

References

1. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage **23** (2004) s151–s160
2. P. Lorenzen, B. Davis, S.J.: Unbiased atlas formation via large deformations metric mapping. In: MICCAI. (2005)
3. Grenander, U., Miller, M.I.: Computational anatomy: An emerging discipline. Quarterly of Applied Mathematics **56** (1998) 617–694
4. Grenander, U.: General Pattern Theory. Oxford Univ. Press (1994)
5. Joshi, S., Grenander, U., Miller, M.: On the geometry and shape of brain submanifolds. International Journal of Pattern Recognition and Artificial Intelligence: Special Issue on Processing of MR Images of the Human **11** (1997) 1317–1343

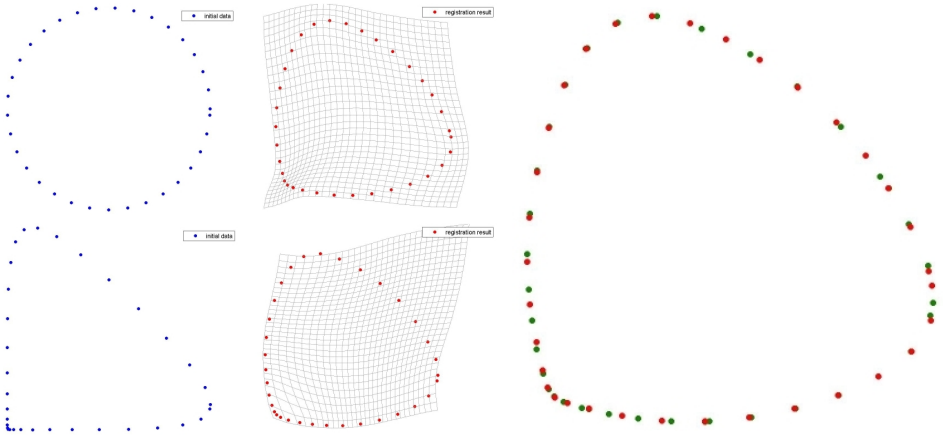

Fig. 1. Experiment with synthetic 2d data. Left column: the two unlabeled point sets; middle: deformed point sets and deformation of a grid; right: averaged template, composed of the two deformed point sets.

6. Frechet, M.: Les elements aleatoires de nature quelconque dans un espace distancie. Annales De L'Institut Henri Poincare **10** (1948) 215–310
7. Trouvé, A., Younes, L.: Metamorphoses through lie group action. Foundations of Computational Mathematics **5** (2005) 173–198
8. Miller, M, I., Younes, L.: Group action, diffeomorphism and matching: a general framework. Int. J. Comp. Vis **41** (2001) 61–84 (*Originally published in electronic form in: Proceeding of SCTV 99, http://www.cis.ohio-state.edu/ szhu/SCTV99.html*).
9. Fei Wang1, Baba C. Vemuri1, A.R.I.M.S.S.J.E.: Simultaneous nonrigid registration of multiple point sets and atlas construction. In: ECCV. (2006)
10. Hongyu Guo, A.R., Joshi, S.: 3d diffeomorphic shape registration using hippocampal datasets. In: MICCAI. (2005)
11. Glaunès, J., Trouvé, A., Younes, L.: Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. In: CVPR, IEEE Computer Society (2004) 712–718
12. Vaillant, M., Glaunes, J.: Surface matching via currents. In Christensen, G.E., Sonka, M., eds.: IPMI. Volume 3565 of Lecture Notes in Computer Science., Springer (2005) 381–392
13. Federer, H.: Geometric measure theory. Springer-Verlag (1987)
14. Morgan, F.: Geometric measure theory, 2nd ed. Acad. Press, INC. (1995)
15. Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. IEEE transactions in image processing **9** (2000) 1357–1370
16. Greengard, L., Strain, J.: The fast gauss transform. SIAM J. Sci. Stat. Comput. **12** (1991) 79–94

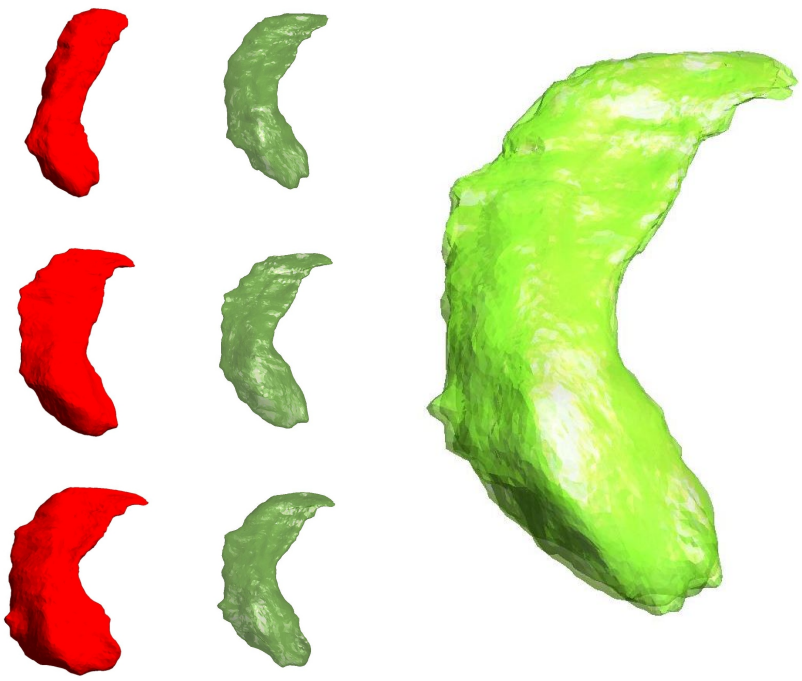

Fig. 2. Experiment with hippocampus surface data. Top row shows the three segmented surfaces of hippocampus, middle row shows the three deformations of the surfaces, and bottom is the averaged template which is composed by these three deformed surfaces.

An H^2 type Riemannian metric on the space of planar curves *

Jayant Shah
Mathematics Department, Northeastern University, Boston MA
email: shah@neu.edu

Abstract

An H^2 type metric on the space of planar curves is proposed and equation of the geodesic is derived. A numerical example is given to illustrate the differneces between H^1 and H^2 metrics.

1 Introduction

Riemannian geometry has by now become a well-established tool for studying variations in shapes. A general differential geometric framework for studying deformations of plane curves has been set up by Mumford and Michor in [3,4] and by Younes et al in [7,8]. The simplest of Riemannian metrics, namely the H^0 metrics, have been studied in [3,10,13,14]. H^1 metrics have been proposed in [2,9,11] and recently, Michor el al have derived a formula for computing the geodesic distances between curves in this metric [6]. The purpose of the present paper is to extend the H^2 metric proposed in [2] for inelastic curves to elastic curves and derive the equation of its geodesic (§3). It is compared with the H^1 metric by means of a numerical example in §4. A scale-invariant version of the H^1 metric given in [9] is described in §2. Extension of the techniques described in this paper to 3D shapes is nontrivial and beyond the scope of this paper; however, see Michor and Mumford [5] for a general framework.

2 H^1 metric

2.1 Definition

Let S denote the unit circle parametrized by the polar angle ω. Let $c(\omega) : S \to \mathbb{R}^2$ be a smooth immersion of degree 1. For a function f on S, let f' denote its derivative $du/d\omega$. Identify $\mathbb{R}^2$ with the complex plane so that $c(\omega) = x(\omega) + iy(\omega) by$. Then, c' may be written as $e^{z(\omega)}$ where $z(\omega) = \lambda(\omega) + i\theta(\omega)$.

*This work was supported by NIH Grant I-R01-NS34189-08

We will denote z also by the notation (λ, θ). Since θ is determined only modulo 2π, we normalize it a proper choice of coordinates in $\mathbb{R}^2$ such that

$$\frac{1}{2\pi}\int_S \theta e^{\lambda} d\omega = \pi \tag{1}$$

The infinitesimal arclength $ds = e^{\lambda} d\omega$. Since the image of c is a closed curve, z satisfies the closure condition

$$\int_S e^{z} d\omega = \int_S c' d\omega = c(2\pi) - c(0) = 0 \tag{2}$$

Conversely, given $z(\omega)$ satisfying conditions (1) and (2), the immersion may be recovered modulo translation in $\mathbb{R}^2$ by setting

$$x(\omega) + iy(\omega) = \int_0^{\omega} e^{z(\varpi)} d\varpi \tag{3}$$

Let Θ be the set of complex smooth functions of ω such that if $\lambda(\omega)+i\theta(\omega)\epsilon\Theta$, then, λ is 2π-periodic and $\theta(\omega + 2\pi) = \theta(\omega)$. Let

$$\Gamma = (\Gamma_0, \Gamma_1, \Gamma_2) : \Theta \rightarrow \mathbb{R}^3 \tag{4}$$

be defined by setting

$$\Gamma_0 = \frac{1}{2\pi}\int_S \theta e^{\lambda} d\omega - \pi, \qquad \Gamma_1 = \int \cos\theta e^{\lambda} d\omega, \qquad \Gamma_2 = \int \sin\theta e^{\lambda} d\omega \tag{5}$$

Then $\mathbb{E} = \Gamma^{-1}(0)$ is the space of smooth immersions and the orbit space $\mathbb{B}$ of $\mathbb{E}$ under the action of diffeomorphisms of S is the space of "smooth" planar curves with crossings, modulo translations.

The tangent space $T\Theta_z$ at a point $z\epsilon\Theta$ is isomorphic to Θ itself. A tangent vector at z is just a pair of functions μ and φ on S. The function μ specifies the stretching of the curve while φ specifies the rotation of the tangent vectors of the curve. Given tangent vectors $u_1 = (\mu_1 + i\phi_1)$ and $u_2 = (\mu_2 + i\phi_2)\epsilon T\Theta_z$, define a Riemannian metric $<< ., . >>$ on Θ and, by restriction also on $\mathbb{E}$, by setting

$$< u_1, u_2 >= \frac{1}{\ell}\int_S (A^2\mu_1\mu_2 + \varphi_1\varphi_2) e^{\lambda} d\omega \quad \text{where} \quad \ell = \int_S e^{\lambda} d\omega \tag{6}$$

Here A is a constant. The energy of infinitesimal deformation of z by a vector u is given by $\frac{1}{2} << u, u >>$. The larger the value of A, the larger the penalty for stretching. This metric is the same as that defined by Younes and Trouvé in [11] where A is equal to 1. The metric is invariant under the action of $Diff(S)$ and hence defines a metric on $\mathbb{B}$.

2.2 Closure of open curves in H^1 metric

A generic $z\epsilon\Theta$ defines an open curve by means of Equation (3). It may be closed by minimizing $\frac{1}{2}|\Gamma|^2_\alpha = \alpha\Gamma_0^2 + \Gamma_1^2 + \Gamma_2^2$ where α is a weight. The equation of gradient descent is

$$\left(\frac{\partial\lambda}{\partial t}, \frac{\partial\mu}{\partial t}\right) = \beta\left(\alpha\Gamma_0\nabla\Gamma_0 + \Gamma_1\nabla\Gamma_1 + \Gamma_2\nabla\Gamma_2\right) \tag{7}$$

where

$$\nabla\Gamma_0 = \frac{\ell}{2\pi}\left(\frac{\theta}{A^2}, 1\right) \quad \nabla\Gamma_1 = \ell\left(\frac{\cos\theta}{A^2}, -\sin\theta\right) \quad \nabla\Gamma_2 = \ell\left(\frac{\sin\theta}{A^2}, \cos\theta\right) \tag{8}$$

The vectors $\nabla\Gamma_0, \nabla\Gamma_1, \nabla\Gamma_2$ span the space normal to the level set $\Gamma^{-1}(\Gamma(z))$.

2.3 Geodesic equations in Θ and $\mathbb{E}$

Let $t \rightarrow z(t) = (\lambda(t), \theta(t))$ be a path in Θ. Let subscript t denote the partial derivative with respect to t. If f is function on S, let $\overline{f}$ denote its average at a point $z\epsilon\Theta$ defined as $\frac{1}{\ell}\int_S fe^\lambda d\omega$. Then the path is a geodesic in Θ if and only if satisfies the equations

$$\begin{aligned} \kappa_\lambda &= \ell(\lambda_t/\ell)_t + \frac{1}{2}\left(\lambda_t^2 + \overline{\lambda_t^2}\right) - \frac{1}{2A}\left(\theta_t^2 - \overline{\theta_t^2}\right) = 0 \\ \kappa_\theta &= (\theta_t/\ell)_t + \lambda_t\theta_t = 0 \end{aligned} \tag{9}$$

which are scale-invariant version of the equations given in [9]. If the path is in $\mathbb{E}$, then it is a geodesic if and only if the projection $P_{T\mathbb{E}}(\kappa)$ of κ onto the tangent space $T\mathbb{E}$ is zero. Let J be the matrix whose $(i,j)^{th}$ entry is $<< \nabla\Gamma_i, \nabla\Gamma_j >>$, $i, j = 0, 1, 2$. Let $\nabla\Gamma$ denote the matrix whose rows are $\nabla\Gamma_0$, $\nabla\Gamma_1$ and $\nabla\Gamma_2$. Then,

$$P_{T\mathbb{E}}(\kappa) = \kappa - (\nabla\Gamma)^T J^{-1} \begin{bmatrix} << \nabla\Gamma_0, \kappa >> \\ << \nabla\Gamma_1, \kappa >> \\ << \nabla\Gamma_2, \kappa >> \end{bmatrix} \tag{10}$$

(See [9] for details.) If the path $t \rightarrow z(t) = (\lambda(t), \theta(t))$ in $\mathbb{E}$ is not a geodesic, we can apply gradient descent to it to find a geodesic path:

$$\frac{\partial z}{\partial\tau} = \beta P_{T\mathbb{E}}(\kappa) \tag{11}$$

The constant β should be sufficiently small to ensure numerical stability of the equation.

3 H^2 metric

A straightforward way to define an H^2 metric on Θ is to introduce derivatives in the metric and set

$$<< u_1, u_2 >> = \ell\int_S (A^2\mu_1'\mu_2' + \varphi_1'\varphi_2')e^{-\lambda}d\omega \tag{12}$$

However, a simpler alternative is to replace the parameter θ by curvature [9]. An embedding $c(\omega) : S \to \mathbb{R}^2$ defines a function $k(\omega)$ on S where $k(\omega)$ is the curvature of the curve at $c(\omega)$. Since $k = d\theta/ds = e^{-\lambda}d\theta/d\omega$, the curvature function $k(\omega)$ satisfies the relation

$$\int_S ke^{\lambda}d\omega = 2\pi \tag{13}$$

The immersion $c(\omega)$ may be recovered modulo rotation and translation by setting $\theta(\omega) = \int_0^{\omega} ke^{\lambda}d\varpi$ and using Equation (3).

Let Ξ now be the set of vector valued functions $\varsigma = (\lambda, k)$ on S. Define

$$\Gamma = (\Gamma_0, \Gamma_1, \Gamma_2) : \Xi \to \mathbb{R}^3 \tag{14}$$

by setting

$$\Gamma_0 \;=\; \int_S ke^{\lambda}d\omega - 2\pi,\ \Gamma_1 = \int \cos\theta e^{\lambda}d\omega,\ \Gamma_2 = \int \sin\theta e^{\lambda}d\omega \tag{15}$$

Then $\mathbb{F} = \Gamma^{-1}(0)$ is the space of smooth immersions modulo translations and rotations in $\mathbb{R}^2$. The orbit space $\mathbb{D}$ of $\mathbb{F}$ under the action of diffeomorphisms of S is now the space of smooth planar curves with crossings, modulo translations and rotations.

The tangent space $T\Xi_{\varsigma}$ at a point $\varsigma \epsilon \Xi$ is again isomorphic to Ξ itself. Given tangent vectors $u_1 = (\mu_1 + ih_1)$ and $u_2 = (\mu_2 + ih_2)\ \epsilon T\Xi_{\varsigma}$, define a Riemannian metric $<< .,. >>$ on Ξ and, by restriction also on $\mathbb{F}$, by setting

$$< u_{1,} u_2 >= \int_S (\frac{A^2}{\ell}\mu_1\mu_2 + \ell h_1 h_2)e^{\lambda}d\omega \text{ where } \ell = \int_S e^{\lambda}d\omega \tag{16}$$

The energy of deformation now the stretching and bending energy of an elastica. The metric physically makes more sense than the H^1 metric which attaches cost to absolute rotation of the tangent vectors rather than their rotation relative to their immediate neighbors. The new metric is invariant under the action of $Diff(S)$ and hence defines a metric on $\mathbb{D}$.

3.1 Closure of open curves in H^2 metric

As before, an open curve corresponding to a generic $\varsigma \epsilon \Xi$ may be closed by minimizing $\frac{1}{2}|\Gamma|^2_{\alpha} = \alpha\Gamma_0^2 + \Gamma_1^2 + \Gamma_2^2$. The equation of gradient descent is

$$\left(\frac{\partial\lambda}{\partial t}, \frac{\partial h}{\partial t}\right) = \beta\left(\alpha\Gamma_0\nabla\Gamma_0 + \Gamma_1\nabla\Gamma_1 + \Gamma_2\nabla\Gamma_2\right) \tag{17}$$

where

$$\begin{aligned} \nabla\Gamma_0 = \left(\frac{\ell k}{A^2}, \frac{1}{\ell}\right) \qquad & \nabla\Gamma_1 = \left(\frac{\ell}{A^2}\{\cos\theta - k\gamma_2\}, -\frac{\gamma_2}{\ell}\right) \\ \nabla\Gamma_2 \;=\; & \left(\frac{\ell}{A^2}\{\sin\theta + k\gamma_1\}, \frac{\gamma_1}{\ell}\right) \\ \gamma_i(\omega) = x_i(2\pi) - x_i(\omega), \qquad & i = 1, 2 \qquad \text{and} \quad x = (x_1, x_2). \end{aligned} \tag{18}$$

3.2 Geodesic equations in Ξ and M

A straighforward calculation shows that a path $t \to \varsigma(t) = (\lambda(t), k(t))$ in Ξ is a geodesic if and only if it satisfies the equations

$$\begin{aligned} \kappa_\lambda &= \ell(\lambda_t/\ell)_t + \frac{1}{2}\left(\lambda_t^2 + \overline{\lambda_t^2}\right) - \frac{\ell^2}{2A^2}\left(k_t^2 + \overline{k_t^2}\right) = 0 \\ \kappa_k &= \frac{1}{\ell}(\ell k_t)_t + \lambda_t k_t = 0 \end{aligned} \tag{19}$$

A path in $\mathbb{F}$ is a geodesic if and only if the projection $P_{T\mathbb{F}}(\kappa) = 0$. The matrix J and the projection $P_{T\mathbb{F}}(\kappa)$ are given by the same formulae as in the case of H^1 metric. If the path $t \to \varsigma(t) = (\lambda(t), k(t))$ in $\mathbb{F}$ is not a geodesic, apply gradient descent to it to find a geodesic path:

$$\frac{\partial \varsigma}{\partial \tau} = \beta P_{T\mathbb{F}}(\kappa) \tag{20}$$

4 A numerical Example

The distance between two unparametrized curves may be found in two steps:

Step 1: Pick parametrizations for the two curves and find the geodesic distance between the two by constructing a minimal geodesic path between the two. Picking the two parametrizations essentially amounts to fixing a point-to-point correspondence between the two curves.

Step 2: Minimize the distance in Step 1 over the set of parametrizations of the second curve, keeping the parametrization of the first curve fixed. Note that if correspondence between a set of landmarks on the two curves is given a priori, it may be incorporated as a constraint on permissible reparametrizations.

The numerical example below finds a geodesic between two parametrized curves with no guarantee that the geodesic found in this way is minimal. One may employ dynamic programming to carry out the second step as described in [11].

There is also the question of choosing the value of the parameter A. It is easy to see that the metrics become degenerate if $A = 0$. For example, in the case of the H^2 metric, if $A = 0$, we can pick $\lambda(t)$ which, in the limit, blows up a point on S where the curvature equals the average curvature and shrinks the rest to a point, thus deforming the given curve into a circle with path length equal to zero. In the same way, in the case of the H^1 metric with $A = 0$, we can deform the curve into an equilateral triangle at no cost. Therefore, we should expect numerical difficulties if the value of A is too small. In fact, for small values of A, the length of the path continued to decrease during gradient descent while the value of κ blew up.

The results of our numerical experiment are shown in the following figure where each column shows a geodesic path constructed from bottom to top. After each step of gradient descent, the curve-closing algorithm was applied to prevent a slow drift away from the manifold of closed curves. The parameter

space S was divided into 250 equal parts and a path between the two curves was constructed by interpolating 49 curves between the initial curve and final curve. In the figure, the first two columns show the numerical geodesics in the H^1 metric with values of A equal to 1 (first column) and 10 (second column). The last two columns show the numerical geodesics in the H^2 metric using $A = 40$ (third column) and $A = 100$ (last column). In both cases, some protrusions of the fish are fashioned into horse's legs, head and the tail while the rest of the protrusions are smoothed out. In the case of the H^2 metric, there is a noticeable tendency for the curve to become circular, especially for the lower value of A. In the case of H^1 metric, the curve tends to remain polygonal.

5 References

1. V. Camion and L. Younes: "Geodesic interpolating splines", EEMCVPR01, LNCS 2134, pp. 513-527, 2001

2. E. Klassen, A. Srivastava and W. Mio: "Analysis of planar shapes using geodesic paths on shape spaces", PAMI 26(3) 2004, pp. 372-384.

3. P. Michor and D. Mumford: "Riemannian geometries on spaces of plane curves", J. Eur. Math. Soc. (JEMS) 8, 2006, pp.1-48, also arXiv:math. DG/0312384

4. P. Michor and D. Mumford: "An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach", arXiv:math.DG/0605009. 2006.

5. P. Michor and D. Mumford: "Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms", arXiv:math.DG/0409303, 2004.

6. P. Michor, D. Mumford and L. Younes: "Diffeomorphic shape comparison", Preprint, 2006.

7. M. Miller and L. Younes: "Group actions, homeomorphisms, and matching: A general framework", Int'l J. of Computer Vision, 41 (1/2), 2001, pp.61-84

8. M. Miller, A. Trouvé and L. Younes: "On the metrics and Euler-Lagrange equations of computational anatomy", Annual Review of Biomedical Engineering, 4, 2002, pp. 375-405.

9. W. Mio, A. Srivastava and S. Joshi: "On shape of plane elastic curves", Submitted to the International Journal of Computer Vision, 2005.

10. J. Shah: H^o-type Riemannian metrics on the space of planar curves", arXiv:math. DG/0510192.

11. L. Younes: "Computable elastic distances between shapes", SIAM J. Appl. Math. 58, 1998, pp. 565-586.

12. A. Trouvé and L. Younes: "Diffeomorphic matching in 1d: designing and minimizing matching functionals", ECCV 2000.

13. A. Yezzi and A. Mennucci: "Conformal Riemannian metrics in space of curves", EUSIPC004, MIA, 2004.

14. A. Yezzi and A. Mennucci: "Metrics in the space of curves", arXiv:math. DG/0412454

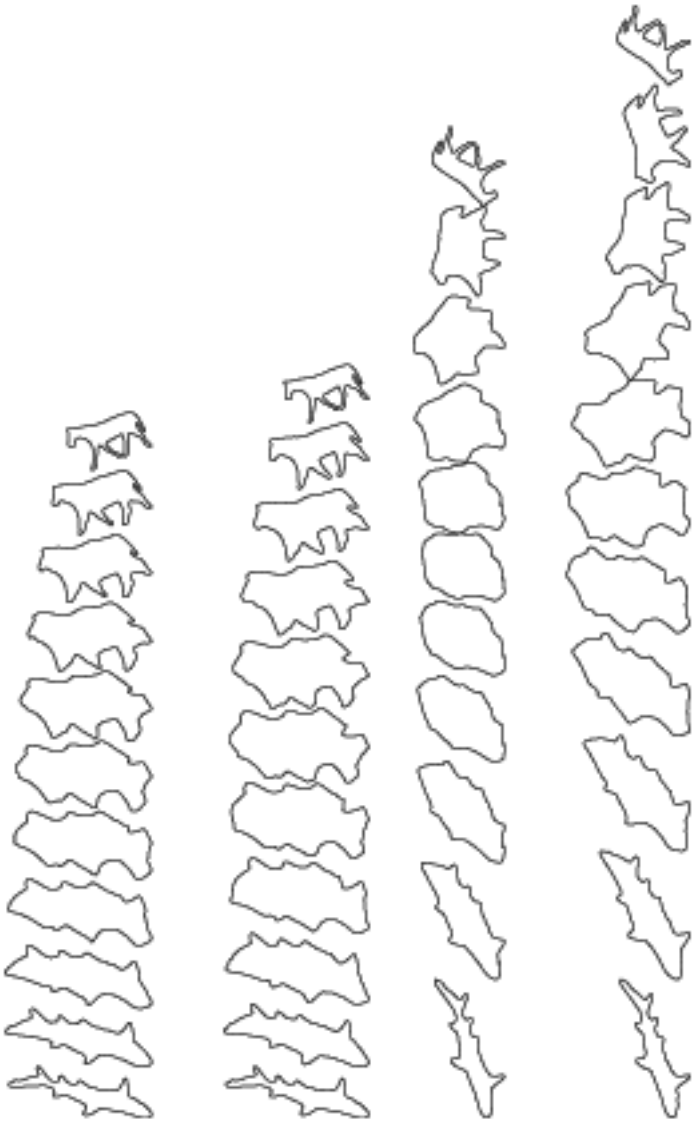

Figure 1: Geodesic paths. H^1, A=1,10, H^2, A=40,200.

Riemannian Metrics on the Space of Solid Shapes

P. Thomas Fletcher and Ross T. Whitaker

School of Computing
University of Utah
fletcher@sci.utah.edu, whitaker@cs.utah.edu

Abstract. We present a new framework for multidimensional shape analysis. The proposed framework represents solid objects as points on an infinite-dimensional Riemannian manifold and distances between objects as minimal length geodesic paths. Intershape distance forms the foundation for shape-based statistical analysis. The proposed method incorporates a metric that naturally prevents self-intersections of object boundaries and thus produces a well-defined interior and exterior along every geodesic path. This paper presents an implementation of the geodesic computations for 2D shapes and gives examples of geodesic paths that demonstrate the advantages of enforcing well-defined boundaries. This compares favorably with equivalent paths under a linear L^2 metric, which do not prevent self-intersection of the boundary, and thus do not produce valid solid objects.

1 Introduction

Shape analysis plays an important role in the understanding of anatomical variability from medical images. Statistics of shape is vital to applications ranging from disease diagnosis, treatment planning, and quantification of the effects of disease. While anatomy consists entirely of solid objects, many shape representations, such as landmarks, boundary curves, or harmonics, do not account for the solid nature of objects. In this paper we present a new shape metric that is well-suited to quantify shape changes in solid objects. We also show that representing shapes as solids results in a shape space that conforms qualitatively with some of our most natural intuitions about shape variabilities.

Quantitative study of shape begins with the formulation of a *shape space*, in which each shape is represented as a point. A distance metric on this shape space gives a measure of shape similarity between any two objects. When the shape space has the structure of a Riemannian manifold, this distance is given by minimal geodesics, or shortest length curve segments, between two shapes. The ability to compute distances between shapes is the foundation for statistical analysis. The first formulations of shape spaces, and the use of metrics to define shape statistics, were developed in the seminal works of Kendall [1] and Bookstein [2]. In their work a 2D object is represented as a discrete set of landmark points. By removing the effects of translations, rotations, and scalings of

these landmarks, the shape space is given the structure of a curved Riemannian manifold. The structure of this space can then be used to define probability distributions as in Mardia and Dryden [3]. The theory of landmark-based shape spaces is reviewed in several books [4], [5].

A typical strategy in shape analysis applications is to apply a linear metric on the parameters of the shape representation. The Active Shape Model (ASM) of Cootes and Taylor [6] represents objects as a dense sampling of their boundaries. They align objects using a Procrustes algorithm and perform principal components analysis (PCA) to capture the shape variability. Kelemen et al. [7] represent 3D objects by spherical harmonic (SPHARM) decompositions of their boundaries and quantify shape variation using PCA with a linear metric on the SPHARM coefficients. Shape variations under these metrics are characterized by straight line paths of object boundaries. Such deformations do not respect the solid properties of objects, and can create intersecting boundaries. Fletcher et al. [8] introduce a generalization of PCA to nonlinear manifolds and use it to compute shape statistics based on medial representations of objects. While nonlinear variations of shape provide a richer set of deformations, there is still no guarantee that shape boundaries will not intersect.

Recent work has focused on representing shape variations of continuous planar curves, where the shape spaces are infinite-dimensional manifolds. Klassen et al. [9] develop elastic curves based on an angle function of the tangent vector. Sharon and Mumford [10] design a metric based on conformal mappings between 2D objects. Michor and Mumford [11] investigate metrics on the space of smooth curves modulo reparameterizations. While this work lays a rigorous mathematical framework for comparing smooth curves, again there is no constraint that the curves not intersect. One method of shape analysis that does indeed constrain solid objects to be free of self-intersections is the diffeomorphic approach, first proposed by Grenander [12]. In this framework shape variations are represented as the actions of diffeomorphisms on a template. Miller and Younes [13] take this approach and define metrics on spaces of diffeomorphism groups, which are infinite-dimensional. While the diffeomorphism approach does preserve solid shapes, the metric is not defined directly on the shapes themselves. Rather, the metric is defined on the diffeomorphism group, which is combined with a matching term to deform one object into another.

In this paper we present a new shape representation which directly models solid objects. The shape space is an infinite-dimensional Riemannian manifold, with a metric designed to preserve non-intersecting boundaries of solid objects. Our framework is valid for objects in both 2D and 3D. The rest of this paper is organized as follows. In Section 2 we formulate the space of solid objects. We develop a Riemannian metric on this space in Section 3 and give a procedure to compute geodesics in Section 4. We present examples of these geodesics for 2D objects in Section 5 and demonstrate that they preserve non-intersecting object boundaries. We compare these geodesics to the equivalent minimal paths under a linear metric, which result in intersecting boundaries.

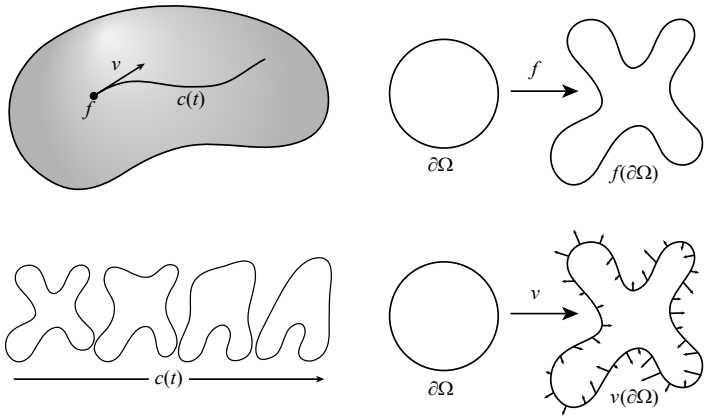

Fig. 1. A pictorial representation of the infinite-dimensional space of solid objects, $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$ (top left). A point f in this space is an embedding representing a solid object (top right). A path $c(t)$ in this space is a smooth deformation of shapes (bottom left). A tangent vector v at the point f is a vector field on the image $f(\partial\Omega)$ that represents an infinitesimal deformation of f (bottom right).

2 The Space of Solid Objects

The proposed formulation of the space of solid objects relies on a fixed template object, which is a compact region $\Omega \subset \mathbb{R}^n$. We require that Ω be a smooth n-dimensional manifold with boundary. The boundary of Ω will be denoted $\partial\Omega$. The compactness of Ω means that it is a closed and bounded set of $\mathbb{R}^n$. The fact that Ω is a manifold with boundary ensures that $\partial\Omega$ is a smooth, non-intersecting manifold that separates $\mathbb{R}^n$ into a distinct interior and exterior. These properties are designed to capture the essence of what it means for an object to be solid. As an example, Ω could be the closed unit ball $\overline{B^n} = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$. The boundary of the $\overline{B^n}$ is the unit sphere, i.e., $\partial\overline{B^n} = S^{n-1} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$.

We define a *solid object* to be an embedding of $\partial\Omega$ into $\mathbb{R}^n$ (Fig. 1). Recall that a mapping $f : \partial\Omega \to \mathbb{R}^n$ is an embedding if it is a diffeomorphism of $\partial\Omega$ onto its image $f(\partial\Omega)$. The space of all such embeddings forms an infinite-dimensional manifold, which is denoted $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$. Since each object f is defined to be an embedding of $\partial\Omega$, the image $f(\partial\Omega)$ is also a smooth, non-intersecting compact manifold that separate $\mathbb{R}^n$ into a distinct interior and exterior. In other words, f preserves our notion of what it means to be solid. Notice that the choice of Ω will determine a fixed topology for the possible objects in the space. However, if Ω and Ω' are two diffeomorphic template objects, then the resulting object spaces $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$ and $\mathcal{E}(\partial\Omega', \mathbb{R}^n)$ are equivalent. In other words, the definition of our object space is independent of the template object up to diffeomorphism.

A path in $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$ is a one-parameter family of embeddings, $c : (a, b) \times \partial\Omega \to \mathbb{R}^n$ (Fig. 1). For each real number $t \in (a, b)$, the point $c(t) \in \mathcal{E}(\partial\Omega, \mathbb{R}^n)$

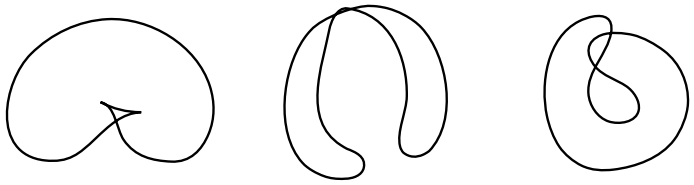

Fig. 2. The possible forms of self-intersections of an object's boundary: local singularity (left), global interior crossing (middle), and global exterior crossing (right).

is an object, and the path c is a smooth deformation of objects parameterized by t. For each $x \in \partial\Omega$ the path c generates a curve in $\mathbb{R}^n$: $t \mapsto c(t)(x)$. The t derivative of this curve, $c_t(t)(x)$, is a vector in $T_{c(t)(x)}\mathbb{R}^n$, the tangent space of $\mathbb{R}^n$ at the point $c(t)(x)$. Thus, the tangent vector $c_t(t)$ is a mapping that assigns to each $x \in \partial\Omega$ the vector $c_t(t)(x) \in T_{c(t)(x)}\mathbb{R}^n$ (Fig. 1). Intuitively, the tangent vector $c_t(t)$ can be thought of as an infinitesimal deformation of the object $c(t)$. The space of all tangent vectors to an object $f \in \mathcal{E}(\partial\Omega, \mathbb{R}^n)$ forms the vector space $T_f\mathcal{E}(\partial\Omega, \mathbb{R}^n) = C^\infty(\partial\Omega, f^*T\Omega_f)$, where f^* denotes the pull-back via f, and $\Omega_f = f(\Omega)$ is the current shape.

Notice that the parameterization of the object boundary is included in our definition of a solid shape. This is in contrast to recent work on planar curves, e.g., [9], [10], [11], where shape is defined modulo reparameterizations. Glaunes et al. [14] and Michor and Mumford [15] also show that metrics on the diffeomorphism group can be used to induce metrics on the space of unparameterized shapes, which they define as the quotient space $\mathit{Diff}(\mathbb{R}^n)/\mathit{Diff}(\mathbb{R}^n, S^{n-1})$, where $\mathit{Diff}(\mathbb{R}^n)$ is the space of diffeomorphisms on $\mathbb{R}^n$ and $\mathit{Diff}(\mathbb{R}^n, S^{n-1})$ are the diffeomorphisms that map the sphere S^{n-1} to itself. By including the parameterization of the boundary, our framework allows correspondences to be made between the boundaries of different objects. This is desirable in medical image applications, where it is typically necessary that corresponding anatomical features be compared across subjects.

3 Riemannian Metrics on $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$

We now define a new class of Riemannian metrics on the space of solid objects, $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$, that are particularly well-suited for solid shape analysis as they prevent self-intersections of shapes. A Riemannian metric on $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$ assigns a smoothly-varying inner product on each tangent space $T_f\mathcal{E}(\partial\Omega, \mathbb{R}^n)$. We denote the inner product of two tangent vectors $v, w \in T_f\mathcal{E}(\partial\Omega, \mathbb{R}^n)$ by $\langle v, w\rangle_f$. The length of a tangent vector v is given by $\|v\|_f = \sqrt{\langle v, v\rangle_f}$. A geodesic is a path γ that minimizes the energy $E(\gamma) = \int_a^b \|\gamma_t(t)\|^2_{\gamma(t)}\, dt$.

3.1 Preventing Boundary Intersections

Intersections of the boundary result when the object mapping $f : \partial\Omega \to \mathbb{R}^n$ fails to be an embedding, that is, it falls outside of the space $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$. There

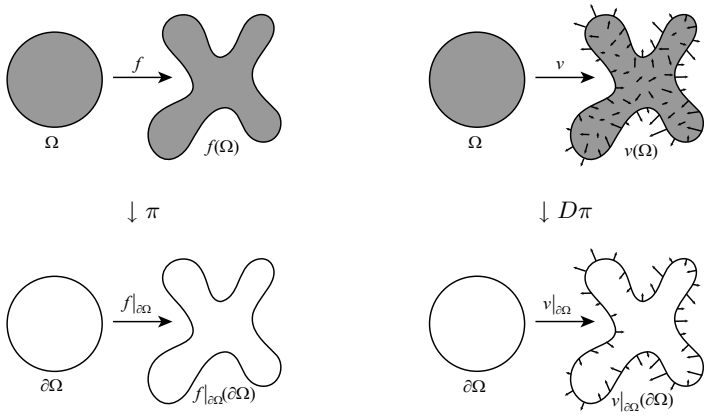

Fig. 3. Projection from the space of full embeddings to the space of boundary embeddings (left). Projection of a tangent vector (right).

are essentially three types of self-intersections of the object boundary that can occur (Fig. 2): local singularities, global interior crossings, and global exterior crossings. Local singularities happen when the derivatives of the object mapping become singular. Global interior crossings occur when the interior of the object collapses and the boundary touches itself. Global exterior crossings occur when the object boundary penetrates itself from the outside. Notice that the local singularities can be detected using only local information of the boundary, i.e., derivatives. However, the boundary of an object with a global intersection may still be smooth, and there is no way to detect these crossings with local information. In the next section we define a Riemannian metric on $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$ that is capable of preventing local singularities and global interior crossings by involving the interior of the object. The exterior global crossings, which are not prevented in our framework, present a unique difficulty in that they involve events that happen external to the object.

3.2 Metrics via Projection

If we are to keep objects from collapsing, it is only natural that velocities of interior points in the object should play a role in the metric. The challenge is to accomplish this while defining a metric that is determined uniquely by the boundary velocities. We do this by defining a metric on the space $\mathcal{E}(\Omega, \mathbb{R}^n)$ of embeddings of Ω in such a way that a vector field on the boundary of the object can be extended to a vector field on the interior with minimal norm. This will allow us to compute geodesics in $\mathcal{E}(\Omega, \mathbb{R}^n)$ that take into account the interior of the object, and then project them back down to geodesics on the solid object space $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$.

There is a natural projection $\pi : \mathcal{E}(\Omega, \mathbb{R}^n) \to \mathcal{E}(\partial\Omega, \mathbb{R}^n)$ given by the restriction to $\partial\Omega$. That is, for an embedding $f \in \mathcal{E}(\Omega, \mathbb{R}^n)$, we have $\pi(f) = f|_{\partial\Omega}$. The

derivative map of this projection, $D\pi : T\mathcal{E}(\Omega, \mathbb{R}^n) \to \mathcal{E}(\partial\Omega, \mathbb{R}^n)$, is also given by a restriction. Given a tangent vector $v \in T_f\mathcal{E}(\Omega, \mathbb{R}^n)$, we have $D\pi(v) = v|_{\partial\Omega}$. These projections are illustrated in Fig. 3. We define a metric on $\mathcal{E}(\Omega, \mathbb{R}^n)$ such that the projection π induces a unique metric on the space $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$.

If $f \in \mathcal{E}(\Omega, \mathbb{R}^n)$ is an embedding, then at each point $x \in \Omega$ the Jacobian matrix $Df(x)$ has positive determinant. Thus, $Df(x)$ is an element of $GL^+(n, \mathbb{R})$, the Lie group of all positive-determinant matrices. We now define the metric on $\mathcal{E}(\Omega, \mathbb{R}^n)$ at the point f as

$$\langle v, w \rangle_f = \int_\Omega \langle Dv, Dw \rangle_{Df} \, dx, \tag{1}$$

where the inner product inside the integral is a right-invariant Lie group metric on $GL^+(n, \mathbb{R})$. There are several possible right-invariant metrics on $GL^+(n, \mathbb{R})$, which lead to an entire class of metrics on $\mathcal{E}(\Omega, \mathbb{R}^n)$ of the form (1). In this paper we use the metric

$$\langle Dv, Dw \rangle_{Df} = \operatorname{tr}\left(Dv Df^{-1} \left(Dw Df^{-1} \right)^T \right). \tag{2}$$

Notice that the value of (2) approaches infinity if Df approaches a zero determinant matrix. This property of the metric means that geodesic paths generate shape mappings with positive Jacobians, and thus do not generate local singularities or interior self-intersections.

We now use the metric (1) on $\mathcal{E}(\Omega, \mathbb{R}^n)$ and the projection mapping π to induce a metric on the solid object space $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$. Given a tangent vector $v \in T_f\mathcal{E}(\partial\Omega, \mathbb{R}^n)$, we define an extension of v to the interior of Ω. An extension of v is a tangent vector $\tilde{v} \in T_{\tilde{f}}\mathcal{E}(\Omega, \mathbb{R}^n)$ such that $\pi(\tilde{f}) = f$ and $D\pi(\tilde{v}) = v$. The vector $\tilde{v}$ is chosen as the extension of v with minimal length in the metric (1), that is, $\tilde{v}$ is an extremal of the norm

$$\|\tilde{v}\|_{\tilde{f}} = \int_\Omega \operatorname{tr}\left(D\tilde{v} A D\tilde{v}^T \right) dx, \qquad A = D\tilde{f}^{-1} D\tilde{f}^{-1T}. \tag{3}$$

This is a variational problem that leads to the Euler-Lagrange equation

$$\operatorname{div}(A \nabla \tilde{v}) = 0. \tag{4}$$

This equation is an elliptic PDE since the matrix A is symmetric, positive-definite. The constraint that $\tilde{v}|_{\partial\Omega} = v$ provides the boundary conditions. Since it is an elliptic PDE with smooth boundary conditions, it has a unique solution. In other words, the vector v lifts to a unique vector $\tilde{v}$ with minimal norm. Therefore, we can define the metric on $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$ to be $\langle v, w \rangle_f = \langle \tilde{v}, \tilde{w} \rangle_{\tilde{f}}$.

The metric defined in (1) is dependent on the choice of the template object Ω and the mapping f from that template to the current shape. This dependence can be removed by including the Jacobian of f in the integral to give the metric

$$\langle v, w \rangle = \int_\Omega \langle Dv, Dw \rangle_{Df} |Df| dx,$$

where $|Df|$ denotes the absolute value of the determinant of Df. This metric is right-invariant, meaning that it is left unchanged by a diffeomorphism of Ω. In other words, the metric is independent of the choice of the mapping f or of the choice of Ω, up to a diffeomorphism. Another possibility is to keep the metric (1), but restrict to only volume-preserving maps, i.e., $|Df| = 1$. A right-invariant metric also has the desirable property that the projection mapping π becomes a Riemannian submersion [16]. This has the consequence that any geodesic γ in the space $\mathcal{E}(\Omega, \mathbb{R}^n)$ has projection $\pi(\gamma)$ that is also a geodesic in $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$. In this paper we focus on the metric given in (1), while the preferred right-invariant metrics are a current area of research.

4 Computing Geodesics

In this section we describe how to compute geodesics in $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$ equipped with the metric induced by (1). We lift the geodesic computations to the space $\mathcal{E}(\Omega, \mathbb{R}^n)$ and then project these geodesics back to $\mathcal{E}(\partial\Omega, \mathbb{R}^n)$ via the mapping π. A geodesic path γ on $\mathcal{E}(\Omega, \mathbb{R}^n)$ is an extremal of the energy functional $E(\gamma) = \int_a^b \int_\Omega \|D\gamma_t\|^2_{D\gamma}\ dx\ dt$, where the metric is defined as in (1), and we are given initial conditions $\gamma(0)|_{\partial\Omega} = f_0$ and $\gamma_t(0)|_{\partial\Omega} = v_0$. The first step is to lift these initial conditions to extensions $\tilde{f}_0$ and $\tilde{v_0}$ defined on all of Ω. Any extension of f_0 may be chosen, and the extension of v_0 is computed via (4).

4.1 Geodesics of Matrix Fields

Rather than solve the above variational problem directly for γ, we instead solve for the Jacobian matrix $D\gamma$. Then, at each time point t, we integrate $D\gamma(t)$ with respect to the spatial variable x to arrive at the final geodesic γ. Consider a time-varying matrix field, $M : (a, b) \times \Omega \to GL(n, \mathbb{R})$, which represents the Jacobian field for the geodesic γ, i.e., $M = D\gamma$. The energy functional now becomes

$$E(M) = \int_a^b \int_\Omega \|M_t\|^2_M\ dx\ dt, \tag{5}$$

with initial conditions $M(0) = Df_0$ and $M_t(0) = Dv_0$. However, solving for extremals M of the energy (5) is not equivalent to solving for extremals $D\gamma$ of the energy $E(\gamma)$. To make them equivalent, M must be constrained to be the Jacobian field of a mapping, that is, M must be kept integrable. The integrability condition is $\text{curl}(M) = 0$, where we define the curl of a matrix field as the component-wise curl of each of its row vectors.[1]

If we first consider the unconstrained variational problem in (5), the extremals are given by pointwise geodesics on $GL^+(n, \mathbb{R})$, that is, for each $x \in \Omega$ the curve $t \mapsto M(t)(x)$ is a geodesic on $GL^+(n, \mathbb{R})$ under the right-invariant

[1] For our purposes the curl of a 2D vector field v is the scalar field $\text{curl}(v) = \partial v_2/\partial x - \partial v_1/\partial y$. The curl of a 3D vector field v is the vector field $\text{curl}(v) = \nabla \times v$.

metric. These geodesics can be computed as the following system of first-order ODEs (see [17] and also [18] p. 277)

$$M_t = XM, \tag{6}$$

$$X_t = XX^T - X^TX, \tag{7}$$

where the initial conditions are given in the form $M(0) = M_0$ and $X(0) = X_0$. Actually, there is a closed-form solution to these equations. It is given by

$$M(t) = \exp(t(X_0 + X_0^T)) \exp(-tX_0^T)M_0. \tag{8}$$

Notice that if M is constant in x, that is, if the shape deformation is an affine transformation, then the integrability constraint is satisfied automatically. This results in an important property of this metric: affine transformations of objects can be computed as closed form geodesics using (8).

4.2 Projection

We now describe the projection step used to solve the integrability constraint . We only consider the 2D case here, although the 3D case is similar. We are given a matrix field M_t, and we want to project it onto the space of integrable matrix fields on Ω, denoted $I(\Omega)$. Again, a matrix field X is in $I(\Omega)$ if $\mathrm{curl}(X) = 0$. The projection needs to be orthogonal under the metric (1). The orthogonal subspace to $I(\Omega)$ under this metric, denoted $I^{\perp}(\Omega)$, consists of all matrix fields of the form $DwJA^{-1}$, where w is a smooth vector field, $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and A is given in (3). This is similar to the Helmholtz decomposition of a vector field into curl-free and divergence-free components. The difference is that the Helmholtz decomposition is orthogonal under the L^2 metric.

We now formulate the projection as a variational problem. Given a matrix field X that we want to project to $I(\Omega)$, we find the nearest matrix field in $I^{\perp}(\Omega)$ and subtract it from X. This is given by the matrix field $DwJA^{-1}$, where w minimizes the energy $E(w) = \int_{\Omega} \|DwJA^{-1} - X\|_{Df}^2 \, dx$. An extremal for this energy satisfies the Euler-Lagrange equation

$$\mathrm{div}\left(JA^{-1}J^TDw^T\right) = \mathrm{div}\left(JX^T\right). \tag{9}$$

This is a second-order elliptic PDE, and the orthogonality dictates that Dirichlet boundary conditions should be used. We solve this equation using a successive over-relaxation (SOR) method [19].

5 Results

In this section we give examples of the geodesic paths in the space of 2D solid objects, using the framework developed in this paper. To illustrate the power of our approach to prevent self-intersections, we compare our results to minimal

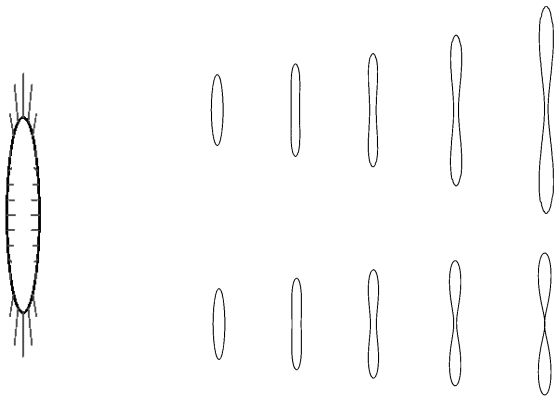

Fig. 4. Minimal paths of the pinching deformation for both the solid metric (top right) and the L^2 metric (bottom right). The initial velocity for both is shown on the left.

paths under a linear metric, i.e., the L^2 metric. The starting object for each example was an ellipse defined as $c(\theta) = ((1/6)\cos\theta, \sin\theta)$. This object was embedded in a uniformly-spaced 50×300 grid. The geodesic equations (6)-(7) were solved on a fixed grid using a second-order Runge-Kutta method [19].In each example we give an initial velocity vector on the boundary of the ellipse as the initial condition to the geodesic problem.

We first give an example of an affine transformation for which the solid shape geodesics can be computed in closed form using (8). The deformation is a stretching along the y-axis. The initial velocity vector (u, v) is given by $u(x, y) = -x, \quad v(x, y) = y$. The geodesic for the solid object metric is given by a path of embeddings $\gamma(t)(x) = M(t)x$, where M is the diagonal matrix with diagonal entries $\{e^{-t}, e^{t}\}$. It is clear that the matrix M has positive determinant for all t, and thus produces only valid, non-intersecting objects. In contrast, the minimal path under the L^2 metric for the same initial conditions is given by $c(t)(\theta) = ((1-t)(1/6)\cos\theta, (1+t)\sin\theta)$. This results in the ellipse collapsing to a vertical line at $t = 1$.

The second example is a pinch deformation (Fig. 4). The initial velocity vector (u, v) is given by $u(x, y) = -x^3 + 3xy^2 - x, \quad v(x, y) = -3x^2y + y^3$. The geodesic from the solid object metric nicely prevents the interior of the object from collapsing. Much like in the stretching example, the boundary slows down the closer it gets to itself. Under the L^2 metric the pinch eventually collapses into an interior global crossing.

The final example is a bending deformation (Fig. 5). The initial velocity vector (u, v) is given by $u(x, y) = (5/4)(x^2 - y^2) - x, \quad v(x, y) = (5/4)xy - y$. The resulting geodesic under the solid object metric produces qualitatively what we expect from a bending deformation. The minimal path under the L^2 metric starts out like a bending, but eventually begins to cross itself. This example shows how nonlinear deformations such as bending are not readily captured by

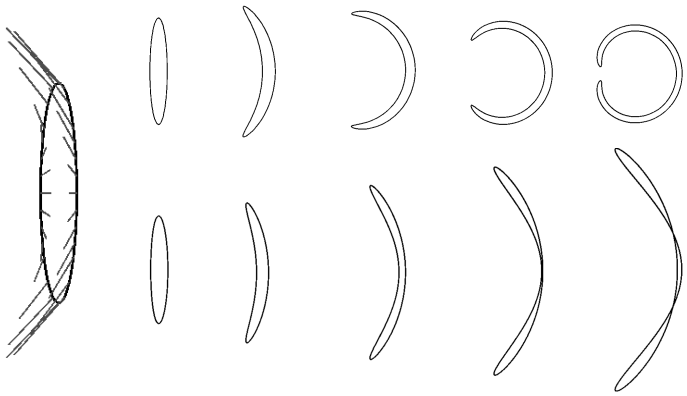

Fig. 5. Minimal paths of the bending deformation for both the solid metric (top right) and the L^2 metric (bottom right). The initial velocity for both is shown on the left.

linear metrics. The geodesics developed in this paper, on the other hand, are able to naturally model such nonlinear deformations.

6 Conclusion

We presented a new framework for shape analysis that directly models solid objects. Our method is based on representing the space of solid objects as an infinite-dimensional Riemannian manifold. We showed that the formulated metric possesses several desirable properties, including that it is valid for representing both 2D and 3D objects and that it can prevent certain types of self-intersections of object boundaries. We intend to pursue the use of this shape metric as a basis for statistical analysis of shape in computer vision and medical image analysis applications.

Acknowledgements

This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from `http://nihroadmap.nih.gov/bioinformatics`.

References

1. Kendall, D.G.: Shape manifolds, Procrustean metrics, and complex projective spaces. Bulletin of the London Mathematical Society **16** (1984) 18–121

2. Bookstein, F.L.: Size and shape spaces for landmark data in two dimensions (with discussion). Statistical Science **1**(2) (1986) 181–242
3. Mardia, K.V., Dryden, I.L.: Shape distributions for landmark data. Advances in Applied Probability **21** (1989) 742–755
4. Dryden, I., Mardia, K.: Statistical Shape Analysis. John Wiley and Sons (1998)
5. Small, C.G.: The statistical theory of shape. Springer (1996)
6. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models – their training and application. Comp. Vision and Image Understanding **61**(1) (1995) 38–59
7. Kelemen, A., Székely, G., Gerig, G.: Elastic model-based segmentation of 3-D neuroradiological data sets. IEEE Trans. on Medical Imaging **18**(10) (1999) 828–839
8. Fletcher, P.T., Lu, C., Joshi, S.: Statistics of shape via principal geodesic analysis on Lie groups. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2003) 95–101
9. Klassen, E., Srivastava, A., Mio, W., Joshi, S.H.: Analysis of planar shapes using geodesic paths on shape spaces. IEEE Transactions on Pattern Analysis and Machine Intelligence (2003) 372–383
10. Sharon, E., Mumford, D.: 2d-shape analysis using conformal mapping. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2004) 350–357
11. Michor, P.W., Mumford, D.: Riemannian geometries on spaces of plane curves. (to appear Journal of the European Mathematical Society)
12. Grenander, U.: General Pattern Theory. Oxford University Press (1993)
13. Miller, M.I., Younes, L.: Group actions, homeomorphisms, and matching: A general framework. International Journal of Computer Vision **41**(1–2) (2001) 61–84
14. Glaunès, J., Trouvé, A., Younes, L.: Modeling planar shape variation via Hamiltonian flows of curves. (preprint: `http://cis.jhu.edu/~younes/Preprints/glaunes.curves.pdf`)
15. Michor, P.W., Mumford, D.: An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. (preprint: `http://lanl.arxiv.org/PS_cache/math/pdf/0605/0605009.pdf`)
16. O'Neill, B.: The fundamental equations of a submersion. Michigan Mathematical Journal **13** (1966) 459–469
17. Wang, H.C.: Discrete nilpotent subgroups of Lie groups. Journal of Differential Geometry **3** (1969) 481–492
18. Helgason, S.: Differential Geometry, Lie Groups, and Symmetric Spaces. Academic Press (1978)
19. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C++: The Art of Scientific Computing. 2nd edn. Cambridge University Press (2002)

Geometric Surface and Brain Warping via Geodesic Minimizing Lipschitz Extensions*

Facundo Mémoli[1], Guillermo Sapiro[2], and Paul Thompson[3]

[1] Stanford Mathematics, Stanford University, Stanford, CA 94305, memoli@math.stanford.edu
[2] Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, guille@ece.umn.edu
[3] Laboratory of Neuro Imaging, UCLA School of Medicine, Los Angeles, CA 90095, thompson@loni.ucla.edu

Abstract. Based on the notion of Minimizing Lipschitz Extensions and its connection with the infinity Laplacian, a theoretical and computational framework for geometric nonrigid surface warping, and in particular the nonlinear registration of brain imaging data, is presented in this paper. The basic concept is to compute a map between surfaces that minimizes a distortion measure based on geodesic distances while respecting the provided boundary conditions. In particular, the global Lipschitz constant of the map is minimized. This framework allows generic boundary conditions to be applied and direct surface-to-surface warping. It avoids the need for intermediate maps that flatten the surface onto the plane or sphere, as is commonly done in the literature on surface-based non-rigid registration. The presentation of the framework is complemented with examples on synthetic geometric phantoms and cortical surfaces extracted from human brain MRI scans.

1 Introduction

Nonrigid surface warping is one of the most fundamental problems in geometric surface processing. This is particularly relevant for problems such as shape comparison, motion and deformation analysis, and shape morphing and interpolation. A particular important example is when the surface represents a brain, for example, the boundary between white matter and gray matter or between gray matter and CSF. Brain warping, which is a form of brain image registration and geometric pattern matching that will be used as test case in this work, is one of the most fundamental and thereby most studied problems in computational brain imaging [44]. Brain images are commonly warped, using 3-dimensional deformation fields, onto a common neuroanatomic template prior to cross-subject comparison and integration of functional and anatomical data. Images of the same subject may be warped into correspondence over time, to help analyze shape changes during development or degenerative diseases. Almost all the active research groups in this area have developed and/or have their favorite brain warping technique.[4] A few representative works can be found at [9, 11, 12, 17, 18, 21, 36, 41, 43, 44, 47, 48], this list being far from complete. In spite of this, the problem is still open and widely studied, since there is not a "ground truth" method to obtain a map between brains. The same is true for generic nonrigid surface matching. The criteria for matching different features (e.g., geometry or intensity) may also depend on the application, which range from recovering intraoperative brain change to mapping brain growth, or reducing cross-subject anatomical differences in group functional MRI studies.

The way the brain warping problem is addressed is critical for studies of brain diseases that are based on population comparisons. Examples of this application can be found at [18, 41], although these are a very non-exhaustive account of the rich literature on the subject. The interested reader may also check [42] for numerous applications of brain warping and population studies. As detailed in [44], shape and brain warping approaches can be divided into two classes, those based on volume-to-volume matching and those based on surface-to-surface matching. Our work belongs to the latter of the categories. Surface matching has recently received increasing attention as most functional brain imaging studies focus on the cortex, which varies widely in geometry across subjects. The power of these studies depends on the degree to which the functional anatomy of the cortex can be aligned across subjects, so improved cortical surface registration has become

* This work is partially supported by DARPA, the Office of Naval Research, the National Science Foundation, the National Institutes of Health, and the National Geospatial-Intelligence Agency.

[4] This includes groups at JHU, UCLA, U. Penn., INRIA, MGH, GATECH, Harvard-BW, and the University of Florida, to name just a few.

a major goal. In contrast with flow based works such as those in [9, 36, 41], our motivation is as in [1, 21–23, 26, 45, 46, 49]. That is, we aim to compute a map that preserves certain pre-defined geometric characteristics of the surfaces. While the literature has mainly attempted to preserve angles (e.g., [1, 21]), and areas (e.g., [11, 17]), we work with geodesic distances (see also [39]). Our work is inspired by the mathematically rich literature on Lipschitz minimizing maps and in its connection to the infinite Laplacian. The motivation for using these frameworks will be presented after some brief mathematical introduction below.

1.1 Key contributions

Let us conclude this section by explicitly formulating key contributions that make our work stand out from the rich literature in the area of surface, and in particular brain, warping:

- We introduce the theoretical and computational framework of Lipschitz minimizing maps into the area of nonrigid surface and brain mapping. This is the only map, following basic axioms [7, 8], that can simultaneously handle diverse boundary conditions, for example, a mixture of points and curves (see also [20]).
- We propose to preserve geodesic distances, being this the first time this is explicitly done for surface warping without intermediate projections.[5] Prior work has concentrated on preserving areas, angles, or Euclidean distances. It is indicated by our preliminary results shown in this paper, the vast literature on geometric surface matching, and the general knowledge in the brain imaging community, that geodesics and intrinsic brain measurements are critical.
- All the computations are done intrinsically on the surface, avoiding the distortions commonly introduced by arbitrary parametrizations or projections onto the plane or sphere.

Although the framework as here described is limited to matching natural geodesics distances, additional feature-matching terms could be easily added for example in the form of extra energy terms in the overall cost to be minimized, for which minimizations techniques different than the efficient one here suggested might need to be developed. Also, see for example concluding remarks, additional matching conditions can be simply considered as metric changes, thereby being included in the proposed framework and in the here suggested numerical implementation. To keep the exposition simpler, we do not include these terms in the current presentation.

2 Formal statement of the problem

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two cortical surfaces (2D surfaces in the three dimensional Euclidean space) which we consider smooth and endowed with the metric inherited from $\mathbb{R}^3$, so that $d_{\mathcal{B}_1}$ and $d_{\mathcal{B}_2}$ are the geodesic distances measured on $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively.[6] Let $\Gamma_1 \subset \mathcal{B}_1$ and $\Gamma_2 \subset \mathcal{B}_2$ be subsets which represent features for which a correspondence is already known (these sets could be empty). In general, the sets Γ_i are the union of smooth curves traced on the surfaces, e.g., sulcal beds lying between gyri, and/or a union of isolated points. A set of anatomical landmarks that occur consistently in all subjects can be reliably identified using standardized anatomical protocols or automated sulcal labeling techniques (see for example Brain VISA [4] by Mangin and Riviere and SEAL [40] by Le Goualher). The overall goal is to extend the given map between $\Gamma_1 \subset \mathcal{B}_1$ and $\Gamma_2 \subset \mathcal{B}_2$ to the whole surface $\mathcal{B}_1$ and $\mathcal{B}_2$, minimizing as much as possible the (geodesic) distortion.

Functional anatomy also varies with respect to sulcal landmarks, but sulci typically lie at the interfaces of functionally different cortical regions so aligning them improves the registration of functionally homologous

[5] After the original version of this paper from March 2005, http://www.ima.umn.edu/preprints/jan2006/jan2006.html, and motivated in part by our work on Gromov-Hausdorff for isometric shape comparison, [34], and by their work on multidimensional scaling (MDS), the authors of [5, 6] proposed a gradient descent technique to solve the classical STRESS function from MDS, where the attributes they select are geodesic distances.

[6] The work developed below applies to any desired intrinsic distances, where we could for example use non-uniform metrics on the surface as in [3, 33]. For simplicity of the presentation, we restrict the discussion to the usual geodesic distance.

areas. As commonly done in brain warping [44], we assume that a correspondence between Γ_1 and Γ_2 is pre-specified to the map (boundary conditions of the map). In this correspondence, internal point correspondences may be allowed to relax along landmark curves in the final mappings, e.g., [31].[7]

To fix ideas let's assume that $\Gamma_1 = \cup_{k=1}^N x_i$ and $\Gamma_2 = \cup_{k=1}^N y_i$, and that the correspondence is given by $x_i \mapsto y_i$ for $1 \leq i \leq N$. The x_i and y_i are elements on the surface, such as points or curves. We want to find a (at least continuous) map $\phi : \mathcal{B}_1 \to \mathcal{B}_2$ such that $\phi(x_i) = y_i$ for $1 \leq i \leq N$ and such that ϕ produces minimal distortion according to some functional $\mathbf{J}$. One possible way of interpreting this problem is that we are trying to extrapolate or extend the correspondence from Γ_1 to the whole of $\mathcal{B}_1$ in such a way that we achieve small distortion.

A common way to measure the distortion produced by a map ϕ is by computing the functionals ($1 \leq p < \infty$)

$$\mathbf{J}_p(\phi) = \left(\frac{1}{\mu(\mathcal{B}_1)} \int_{\mathcal{B}_1} \|D_{\mathcal{B}_1}\phi\|_2^p \, \mu(dx) \right)^{1/p} \tag{1}$$

where $D_{\mathcal{B}_1}$ denotes differentiation intrinsic to the surface $\mathcal{B}_1$ and μ is the area measure on $\mathcal{B}_1$. One immediate idea is then to consider, for a fixed $p \in (1, \infty)$, the following variational problem:

Problem 1 (minimize $\mathbf{J}_p$). Find $\phi \in \mathcal{S}$ such that $\mathbf{J}_p(\phi) = \inf_{\psi \in \mathcal{S}} \mathbf{J}_p(\psi)$, where $\mathcal{S}$ is a certain smoothness class of maps ϕ from $\mathcal{B}_1$ to $\mathcal{B}_2$ such that they respect the given boundary conditions $\phi(x_i) = y_i$ for all $x_i \in \Gamma_1$.

The case $p = 2$ corresponds to the Dirichlet functional and has connections with the theory of (standard) Harmonic Maps. In more generality, it is customary to call the solutions to Problem 1 p-Harmonic Maps, see for example [14, 15, 25]. It is easy to show, under mild regularity assumptions, that for a fixed ϕ, $\mathbf{J}_p(\phi)$ is nondecreasing as a function of p, and that [19]

$$\mathbf{J}_\infty(\phi) := \lim_{p \uparrow \infty} \mathbf{J}_p(\phi) = \operatorname{essup}_{x \in \mathcal{B}_1} \|D_{\mathcal{B}_1}\phi(x)\|_2, \tag{2}$$

which is the Lipschitz constant of ϕ.

In this paper we propose to use the functional $\mathbf{J}_\infty$ as a measure of distortion for maps between surfaces and to solve the associated variational problem in order to find a candidate mapping between the surfaces (constrained by the provided boundary conditions). Once again, although our working example are brain surfaces, the here proposed framework is general for non-rigid surface matching.

Let $\mathcal{L}$ denote the space of all Lipschitz continuous maps $\psi : \mathcal{B}_1 \to \mathcal{B}_2$ such that $\psi(x_i) = y_i$ for $1 \leq i \leq N$. We then propose to solve the following problem:

Problem 2 (minimize $\mathbf{J}_\infty$). Find $\phi \in \mathcal{L}$ such that $\mathbf{J}_\infty(\phi) = \inf_{\psi \in \mathcal{L}} \mathbf{J}_\infty(\psi)$.

We now argue in favor of this functional.

2.1 Why use $\mathbf{J}_\infty$?

Our first argument is that $\mathbf{J}_\infty$ measures distortion in a more global way than any of the $\mathbf{J}_p$ for $p \in (1, \infty)$, since instead of computing an averaged integral quantity, we are looking at the supremum of the local distortions, $\|D_{\mathcal{B}_1}\phi(x)\|_2$. Note also that $\mathbf{J}_p$ is bounded above by $\mathbf{J}_\infty$under mild regularity assumptions for all $p \in (1, \infty)$.

Another element to consider is that this problem is well posed for the kind of general boundary data we want to respect, provided both at curves and isolated points on the surfaces. At least for the case $p \leq 2$, this is not true in general, see [7].

We are then looking for a Lipschitz extension of the map given at Γ_1 whose Lipschitz constant is as small as possible. Let $L(\Gamma_1, \Gamma_2) := \max_{x_i, x_j \in \Gamma_1} \frac{d_{\mathcal{B}_2}(y_i, y_j)}{d_{\mathcal{B}_1}(x_i, x_j)}$, that is, the Lipschitz constant of the boundary data. In general, we have $\inf_{\psi \in \mathcal{L}} \mathbf{J}_\infty \geq L(\Gamma_1, \Gamma_2)$ (we cannot get lower distortion than the one already introduced by the pre-specified boundary conditions). This is related to Kirszbraun's Theorem, which in one of its many guises states that a Lipschitz map $f : S \to \mathbb{R}^D$, $S \subset \mathbb{R}^d$, has an extension $\bar{f} : \mathbb{R}^d \to \mathbb{R}^D$ with the same

[7] Extending the framework here introduced, in particular the matching of metric measurements, to handle flexibility in the boundary conditions, e.g., [20], is an interesting direction to pursue.

Lipschitz constant as f, see [16]. In the same vein, one has Whitney and McShane extensions which apply to the case when the domain is any metric space X and the target is $\mathbb{R}$. These extensions provide functions that agree with f where boundary conditions are given and preserve the Lipschitz constant throughout X, see for example [2, 27]. The more general problem of extending $f : S \to Y$ ($S \subset X$, X and Y any metric spaces) to all X with the same Lipschitz constant is not so well understood and only partial results are known, see for example [28–30].

The idea then is to keep the distortion at the same order as that of the provided boundary conditions. In general there might be many solutions for the Problem (2). One particular class of minimizers which has recently received a lot of attention is that of *absolute* minimizers, or *absolutely minimizing Lipschitz extensions* (AMLE). Roughly speaking, the idea here is to single out those solutions of Problem (2) that also possess minimal local Lipschitz constant, again, see [2] for a general exposition, and [27] for a treatment of the case when the domain is any reasonable metric space and the target is the real line. This is the direction we pursue with the computational approach detailed in the next section.[8]

3 Proposed computational approach

If we take for example the case of p-Harmonic maps, one way of dealing with the computation of the optimal map ϕ_p is by implementing the geometric p-heat flow associated with the Euler-Lagrange equation of the functional $\mathbf{J}_p$, starting from a certain initial condition. As was explained in [35], using an implicit representation for both $\mathcal{B}_1$ and $\mathcal{B}_2$, we could obtain the partial differential equation $\mathbf{PDE}_p$ we need to solve in order to find ϕ_p. By taking the formal limit as $p \uparrow \infty$ we would find $\mathbf{PDE}_\infty$, the PDE that characterizes the solution ϕ_∞ of the (variational) Problem (2).[9] All of this might work if we had a notion of solution for the resulting PDEs. Whereas this is feasible in the case of $\mathbf{PDE}_p$ for $1 < p < \infty$, to the best of our knowledge, there is no such notion of a solution for $\mathbf{PDE}_\infty$. One could of course still persist and try to solve these equations without the necessary theoretical foundations and call these plausible solutions ∞-Harmonic Maps. Nonetheless, this is certainly an interesting line of research.

A different direction is considered in this work. As a guiding example, we first concentrate on the case where $\mathcal{B}_1$ is any closed smooth manifold and $\mathcal{B}_2$ is replaced by $\mathbb{R}$, as considered in [8] (for scalar data interpolation on surfaces), and in [38]. In [8], the authors propose to follow a similar path to the one we have just described, and they do not obtain a convergent numerical discretization for the resulting PDE. Meanwhile, in [38], the author proposes a convergent discretization of the PDE, basing his construction on the original variational problem. We choose to follow this idea as our guiding principle, and extend it to the case where both $\mathcal{B}_1$ and $\mathcal{B}_2$ are surfaces.

We now explain this alternative approach. The basic idea is simple, instead of first obtaining the Euler-Lagrange equations for the energy $\mathbf{J}_\infty$ and then discretizing them, we first discretize the energy $\mathbf{J}_\infty$ and then proceed to solve the resulting discrete problem. Consider that the domain $\mathcal{B}_1$ is given discretely as a set of (different) points $\mathbb{B}_1 = \{x_1, \ldots, x_m\}$ together with a neighborhood relation (i.e., a graph). To fix ideas let's assume the neighborhood relation is a k-nearest neighbors one. Denote, for each $1 \leq i \leq m$, by $N_i = \{x_{j_1}, \ldots, x_{j_k}\} \in \mathbb{B}_1$ the set of k neighbors of the point x_i. We consider the discrete local Lipschitz constant of the map ϕ at x_i:

$$L_i(\phi) := \max_{x_j \in N_i} \frac{d_{\mathcal{B}_2}(\phi(x_i), \phi(x_j))}{d_{\mathcal{B}_1}(x_i, x_j)} \tag{3}$$

Upon noting that $L_i(\phi)$, which measures the local geodesic deformation for the N_i neighborhood of the point x_i, serves as a discrete approximation to $\|D_{\mathcal{B}_1}\phi(x_i)\|_2$, we see that a possible discretization of the functional $\mathbf{J}_\infty(\phi)$ is given by the discrete global Lipschitz constant of ϕ given by $\max_{1 \leq i \leq m} L_i(\phi)$. The author of [38] proposed, in the case when $\mathcal{B}_2$ is replaced by $\mathbb{R}$, solving the discrete version of Problem (2) by the following iterative procedure (here extended for $\mathcal{B}_2$ a surface as in our problem):

– Let ϕ_0 be an initial guess of the map (see below for details on this for our case).

[8] As commonly happens in the surface and brain warping literature, the map cannot be guaranteed to be invertible. We have not experienced any problems with this in our experiments.

[9] The case when the domain is a subset of $\mathbb{R}^d$ and the target is the real line leads to the so called infinity Laplacian, see [2, 13, 24].

– For each $n \geq 1$, if $x_i \notin \Gamma_1$, let

$$\phi_n(x_i) = \arg\min_{y \in \mathcal{B}_2} \max_{j \in N_i} \frac{d_{\mathcal{B}_2}(y, \phi_{n-1}(x_j))}{d_{\mathcal{B}_1}(x_i, x_j)} \tag{4}$$

– $\phi_n(x_i) = y_i$ for all $n \geq 0$ for $x_i \in \Gamma_1$.

With computational efficiency related modifications described below, this is the approach we follow in general. The intuition behind this iterative procedure is that, at each point of the domain, we are changing the value of the map in order to minimize the local Lipschitz constant, that is, the local (geodesic) distortion produced by the map. This is in agreement with the notion of AMLEs explained in §2.1. We should remark that since we are using intrinsic distances for the matching, we can let $L_i(\phi)$ play the role of (the norm of) the displacement field for analyzing the deformation,[10] see §5 ahead.

4 Implementation details

In addition to discretizing the domain $\mathbb{B}_1$, we also use a discretization $\mathbb{B}_2 = \{y_1, \ldots, y_{m'}\}$ of the target space $\mathcal{B}_2$ for our implementation. We endow $\mathbb{B}_2$ with a neighborhood relation given by the k-nearest neighbors of each point. For computational efficiency, we work at all times with two different scales in the discrete domain $\mathbb{B}_1$. For large data sets, additional scales can be used. We choose a subset F_1 of $\mathbb{B}_1$ such that $\#F_1 \ll m$ but still F_1 is an efficient (well separated) covering of $\mathbb{B}_1$ with small covering radius. We do this by using the well known (geodesic) Farthest Point Sampling (FPS) procedure, see [34, 37], which can be efficiently constructed based on optimal computational techniques. Roughly speaking, we apply the iterative procedure on this subset of points only and then extend the map to the rest of the points in the domain $\mathbb{B}_1$. We now show how to obtain a reasonable initial condition ϕ_0 and then discuss additional details regarding the implementation of the iterative procedure described in the previous section.

Building the initial condition: We compute, for all $x_r \in F_1 \backslash \Gamma_1$, $\phi_0(x_r) = \arg\min_{y \in \mathcal{B}_2} \max_{x_i \in \Gamma_1} \frac{d_{\mathcal{B}_2}(y, y_i)}{d_{\mathcal{B}_1}(x_r, x_i)}$. For this step we use the classical Dijkstra's algorithm for approximating the geodesic distances $d_{\mathcal{B}_1}$ and $d_{\mathcal{B}_2}$ since they might be evaluated at faraway points. This is of course run on the graphs obtained from connecting each point to its k-nearest neighbors.

The iterative procedure: After ϕ_0 is computed for all points in the set F_1, we run the iterative procedure from the previous section on this set of points. The main modification here is that whereas we still use Dijkstra's algorithm for approximating $d_{\mathcal{B}_2}$ in the target surface, since in the domain we must compute $d_{\mathcal{B}_1}$ only for neighboring points (F_1 was chosen to be dense enough), for computational efficiency we can approximate $d_{\mathcal{B}_1}(x_i, x_j) \simeq \|x_i - x_j\|$ for $x_j \in N_i$. We should also point out that for points in F_1, the neighborhood relation is defined to be that of k-nearest neighbors with respect to the metric on $\mathbb{B}_1$ defined by the adjacency matrix of $\mathbb{B}_1$. Let $\phi_* : F_1 \to \mathbb{B}_2$ denote the map obtained as the output of this stage.

Extension to the whole domain: After we have iterated over points in F_1 until convergence, we extend the map ϕ_* to all points x_i in $\mathbb{B}_1 \backslash \{F_1 \cup \Gamma_1\}$. This is done by computing $\phi_*(x_i) = \arg\min_{y \in \mathcal{B}_2} \max_{x \in F_1 \cup \Gamma_1} \frac{d_{\mathcal{B}_2}(y, \phi_*(x))}{d_{\mathcal{B}_1}(x_i, x)}$. For this step, and since we have already obtained the map for a relatively dense subset, we approximate both $d_{\mathcal{B}_1}$ and $d_{\mathcal{B}_2}$ by the Euclidean distance. Once again, the motivation for this is just computational efficiency.

5 Examples

In this section we present some computational examples of the ideas presented in previous sections. First, in Figure 1, top, the domain $\mathcal{B}_1$ is a *cube* ($m = 10086$) and the target $\mathcal{B}_2$ is a *sphere* ($m' = 17982$). For the purposes of visualizing the map, we assigned the clown texture (which can be thought of as a function $I : \mathbb{B}_2 \to \mathbb{R}$) to the sphere, which can be seen on the bottom-right corner of the figure. The sphere and the cube were concentric and of approximately the same size. We selected F_1 on the cube consisting of 1000 well

[10] One can imagine a situation in which two isometric surfaces are matched by our algorithm such that $L_i(\phi) = 1$ for all i, but the displacement field $\|x_i - \phi(x_i)\|$ is large since there may be no rigid motion that aligns the two surfaces. One simple example is a flat sheet of paper and the same sheet slightly bent.

separated points using the FPS procedure alluded to in §4. Also, we set $k = 6$ (number of neighbors). We then chose Γ_1 to be the first 100 points of the set and then projected them onto the sphere, obtaining in this way, the corresponding set Γ_2 to use as boundary conditions. We then followed the computational procedure detailed before. The top-left figure shows the composition $I \circ \phi_* : \text{cube} \to I\!R$ as a texture on the cube. Finally, the top-right and the bottom-left images show the histogram of $L_i(\phi_*)$ and its spatial distribution in the domain (we paint the cube at each point x_i with the color corresponding to $L_i(\phi_*)$), respectively. Ideally, we would like to obtain a δ-type histogram, meaning that the distances have been constantly scaled. Of course, this is not possible (unless one of the surfaces is isometric to a scaled version of the other), and we attempt to obtain histograms as concentrated as possible. This is quite nicely obtained for this and the additional examples in this paper.

Figure 1, bottom left and right, shows the construction of a map from the unit sphere S^2 into a cortical hemisphere $\mathbb{B}$ ($\mathcal{B}$). The boundary conditions consisted of 6 pairs of points. We first took the following 6 points on the sphere $\Gamma_1 = \{(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)\}$. We then constructed the intrinsic distance matrix $[d_{S^2}(p_i, p_j)]$ for all $p_i, p_j \in \Gamma_1$. Finally, we chose 6 points $\{q_1, \ldots, q_6\} = \Gamma_2$ in $\mathbb{B}$ such that $\max_{i \neq j} \frac{d_{\mathbb{B}}(q_i, q_j)}{d_{S^2}(p_i, p_j)}$ was as close as possible to $\frac{1}{\pi}\text{diam}(\mathbb{B})$. We painted $\mathbb{B}$ with a texture I_H depending on its mean curvature so as to more easily visualize the sulci/crests: If $H(x)$ stands for mean curvature of $\mathcal{B}$ at x, then $I_H(x) = (H(x) - \min_x H(x))^2$. See the caption for more details.

The example in Figure 2 is about computing a map Φ from a subject's left hemisphere $\mathbb{B}_1$ to another subject's left hemisphere $\mathbb{B}_2$. The boundary conditions were constructed in a way similar to the one used for the previous example, but in this case, 300 points were chosen. Note that, if available, hand traced curves could be used as commonly done in the literature (in other words, more anatomical/functional oriented boundary conditions). In the first two rows we show 4 different views of each cortical surface, and in the third row we show $\mathbb{B}_1$ colored with the values of $L_i(\Phi)$ which we interpret as a measure of the local deformation of the map needed to match $\mathbb{B}_1$ to $\mathbb{B}_2$. See the caption for more details.

6 Concluding remarks

In this paper we have introduced the notions of minimizing Lipschitz extensions into the area of surface and brain warping. These maps provide a more global constraint than ordinary p-harmonic ones, and allow for more general boundary conditions. The proposed computational framework leads to an efficient surface-to-surface warping algorithm that avoids distorting intermediate steps that are common in the brain warping literature. We are currently investigating the use of this new warping technique for creating population averages and applying it to disease and growth studies. In earlier work, the Jacobian of a deformation mapping over time has been used to map the profile of brain tissue growth and loss in a subject scanned serially (tensor-based morphometry [10, 41]). The discrete local Lipschitz constants of our computed mappings also provide a useful index of deformation that can be analyzed statistically across subjects. The framework here introduced can also be applied in 3D for volumetric warping and with weighted geodesic distances instead of natural ones, or different metrics in general, to include additional geometric characteristics in the matching. As recently shown in [32], the use of pairwise distances is of importance of other matching and computer vision tasks beyond the ones discussed in this paper. Results in these directions will be reported elsewhere.

Finally, a further validation and possible different application of the framework here proposed is presented in Figure 3. Here, the 3D shape on the left is mapped onto its given isometric 3D shape on the right. We used 2000 points per shape, with 20 pointwise boundary conditions. The bottom-left shows the shape on the top-left, colored according to the corresponding local Lipschitz constant. The bottom-right shows the histogram of the local Lipschitz constant, centered around one (average is 1.1917). The small deviation from the ideal case for isometric shapes, an histogram with all the points exactly at one, is expected due to the sampling (the shapes were first bended and then sampled), the algorithmic discretization, perturbations in the boundary conditions matches, and the extrapolation from a limited number of points. The detailed exploitation of this framework for metric-invariant shape warping and recognition, with side information in the form of boundary conditions, will be reported elsewhere.

References

1. S. Angenent, S. Haker, A. Tannenbaum, and R. Kikinis, "Conformal geometry and brain flattening," *Proc. MICCAI*, pp. 271-278, 1999.
2. G. Aronsson, M. Crandall and P. Juutinen, "A tour of the theory of absolutely minimizing functions." *Bull. Amer. Math. Soc.* **41:4**, pp. 439–505, 2004.
3. A. Bartesaghi and G. Sapiro, "A system for the generation of curves on 3D brain images," *Human Brain Mapping* **14:1**, pp. 1-15, 2001.
4. `http://brainvisa.info/`
5. A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching," *PNAS* **103**, pp. 1168-1172, January 2006.
6. A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Efficient computation of isometry-invariant distances between surfaces," *CIS Technical Report* **2006-02**, Technion, Israel, Januray 2006.
7. V. Caselles, J.M. Morel, and C. Sbert, "An axiomatic approach to image interpolation." *IEEE Trans. Image Process.* **7:3**, pp. 376–386, 1998.
8. V. Caselles, L. Igual, and O. Sander, "An axiomatic approach to scalar data interpolation on surfaces." *Numer. Math.* **102:3**, pp. 383–411, 2006.
9. G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "A deformable neuroanatomy textbook based on viscous fluid mechanics," *27th Ann. Conf. on Inf. Sciences and Systems*, pp. 211-216, 1993.
10. M.K. Chung, K.J. Worsley, S. Robbins, and A.C. Evans. "Tensor-based brain surface modeling and analysis," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* **I**, pp. 467-473, 2003.
11. M.K. Chung, K.J. Worsley, S. Robbins, T. Paus, J. Taylor, J.N. Giedd, J.L. Rapoport and A.C. Evans, "Deformation-based surface morphometry applied to gray matter deformation," *Neuroimage*, **18(2)**, pp. 198-213, 2003.
12. D. L. Collins, T. M. Peters, and A. C. Evans, "An automated 3D non-linear image deformation procedure for determination of gross morphometric variability in the human brain," *Proc. Visualization in Biomed. Comp. (SPIE)* **3**, pp. 180-190, 1994.
13. M. G. Crandall, L. C. Evans, and R. F. Gariepy, "Optimal Lipschitz extensions and the infinity Laplacian," *Calc. Var. Partial Differential Equations* **13:2**, pp. 123–139, 2001.
14. J. Eells and L. Lemaire, "A report on harmonic maps," *Bull. London Math. Soc.* **10:1**, pp. 1–68, 1978.
15. A. Fardoun and R. Regbaoui, "Heat flow for p-harmonic maps with small initial data." *Calc. Var. Partial Differential Equations* **16:1**, pp. 1–16, 2003.
16. H. Federer, *Geometric Measure Theory*, Springer-Verlag, 1969.
17. B. Fischl, M. I. Sereno, R. B. H. Tootell, and A. M. Dale, "High-resolution inter-subject averaging and a coordinate system for the cortical surface," *Hum Brain Mapp.* **8(4)**, pp. 272-84, 1999.
18. P. A. Freeborough and N. C. Fox, "Modeling brain deformations in Alzheimer's disease by fluid registration of serial 3D MR images," *J Comput Assist Tomogr* **22**, pp. 838-43, 1998.
19. D. Gilbarg and N. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Reprint of the 1998 edition, Classics in Mathematics, Springer-Verlag, Berlin, 2001.
20. J. Glaunes, A. Trouv and L. Younes, "Diffeomorphic Matching of Distributions: A New Approach for Unlabelled Point-Sets and Sub-Manifolds Matching." *CVPR* (**2**), pp. 712-718, 2004.
21. X. Gu, Y.L. Wang, T.F. Chan, P.M. Thompson, and S.T. Yau, "Genus zero surface conformal mapping and its application to brain surface mapping," *IEEE Transactions on Medical Imaging*, **23:7**, 2004.
22. S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, "Optimal mass transport for registration and warping," *International Journal on Computer Vision* **60:3**, pp. 225-240, 2004.
23. M. K. Hurdal, P. L. Bowers, K. Stephenson, D. W. L. Sumners, K. Rehm, K. Schaper and D. A. Rottenberg, "Quasi-conformally flat mapping the human cerebellum," in C. Taylor and A. Colchester (eds), *Medical Image Computing and Computer-Assisted Intervention - MICCAI'99*, Vol. 1679 of Lecture Notes in Computer Science, Springer, Berlin, pp. 279-286, 1999.
24. R. Jensen, "Uniqueness of Lipschitz extensions: Minimizing the sup norm of the gradient," *Arch. Rational Mech. Anal.* **123:1**, pp. 51–74, 1993.
25. J. Jost, *Riemannian Geometry and Geometric Analysis*, Springer-Verlag, Berlin, 2002.
26. L. Ju, J. Stern, K. Rehm, K. Schaper, M. Hurdal and D. Rottenberg, "Cortical surface flattening using least square conformal mapping with minimal metric distortion," *Proceedings of the Second IEEE International Symposium on Biomedical Imaging*, pp. 77-80, 2004.
27. P. Juutinen, "Absolutely minimizing Lipschitz extensions on a metric space," *Ann. Acad. Sci. Fenn. Math.* **27:1**, pp. 57–67, 2002.
28. U. Lang, "Extendibility of large-scale Lipschitz maps," *Trans. Amer. Math. Soc.* **351**, pp. 3975-3988, 1999.
29. U. Lang, B. Pavlovic, and V. Schroeder, "Extensions of Lipschitz maps into Hadamard spaces," *Geom. Funct. Anal.* **10**, pp. 1527-1553, 2000.

30. U. Lang and V. Schroeder, "Kirszbraun's theorem and metric spaces of bounded curvature," *Geom. Funct. Anal.* **7**, pp. 535–560, 1997.
31. A. Leow, C. L. Yu, S. J. Lee, S. C. Huang, R. Nicolson, K. M. Hayashi, H. Protas, A. W. Toga, and P. M. Thompson, "Brain structural mapping using a novel hybrid implicit/explicit framework based on the level-set method." *NeuroImage* **24:3**, pp. 910–927, 2005.
32. M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," *Proc. IEEE-ICCV*, Beijing, China, October 2005.
33. F. Mémoli and G. Sapiro, "Distance functions and geodesics on submanifolds of R^d and point clouds," *SIAM Journal Applied Math.* **65**, pp. 1227-1260, 2005.
34. F. Mmoli, and G. Sapiro, "A theoretical and computational framework for isometry invariant recognition of point cloud data." *Found. Comput. Math.* **5:3** pp. 313–347, 2005.
35. F. Mémoli, G. Sapiro, and S. Osher, "Solving variational problems and partial differential equations mapping into general target manifolds," *J. Comput. Phys.* **195**, pp. 263-292, March 2004.
36. M. I. Miller, A. Trouve, and L. Younes, "On the metrics and Euler-Lagrange equations of computational anatomy," *Ann. Rev. Biomed Eng.* **4**, pp. 375-405, 2002.
37. C. Moenning and N. A. Dodgson, "Fast marching farthest point sampling," `http://www.cl.cam.ac.uk/users/cm230/docs/pdfs/FastFPS.pdf`.
38. A. Oberman, "Convergent difference schemes for the infinity Laplacian: Construction of absolutely minimizing Lipschitz extensions," *Math. Comp.* **74**, pp. 1217-1230, 2005.
39. E. Schwartz, A. Shaw, and E. Wolfson, "A numerical solution to the generalized mapmaker's problem: Flattening nonconvex polyhedral surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11:9**, pp. 1005 -1008, 1989.
40. `http://www.bic.mni.mcgill.ca/~georges/`
41. P. M. Thompson, J. N. Giedd, R. P. Woods, D. MacDonald, A. C. Evans, and A. W. Toga, "Growth patterns in the developing brain detected by using continuum-mechanical tensor maps," *Nature* **404**, pp. 190-193., 2000.
42. P. Thompson, www.loni.ucla.edu/∼thompson/thompson.html
43. P. M. Thompson, R. P. Woods, M. S. Mega, and A. W. Toga, "Mathematical/computational challenges in creating population-based brain atlases," *Human Brain Mapping* **9(2)**, pp. 81-92, Feb. 2000.
44. A. W. Toga, *Brain Warping*, Academic Press, New York, 1998.
45. D. Zhang and M. Hebert, "Harmonic maps and their applications in surface matching," *Proc. CVPR '99*, Colorado, pp. 2524-2530, June 1999.
46. G. Zigelman, R. Kimmel, and N. Kiryati, "Texture mapping using surface flattening via multi-dimensional scaling," *Technion-CIS Technical Report* **2000-01**, 2000.
47. D. C. Van Essen, H. Drury, S. Joshi, and M. I. Miller, "Comparisons between human and macaque using shape-based deformation algorithms applied to cortical flat maps," *3rd Int. Conference on Functional Mapping of the Human Brain, NeuroImage* **5(4)**, Copenhagen, May 19-23, 1997.
48. D. C. Van Essen, H. Drury, S. Joshi, and M. I. Miller, "Functional and structural mapping of human cerebral cortex: Solutions are in the surfaces," *Proceedings of the National Academy of Science* **95**, pp. 788-795, February 1998.
49. Y.L Wang, X. Gu, T. Chan, P.M. Thompson, and S.T. Yau, "Intrinsic brain surface conformal mapping using a variational method," *Proc. Medical Imaging Computing and Computer Assisted Intervention (MICCAI)*, Canada, November 2003.

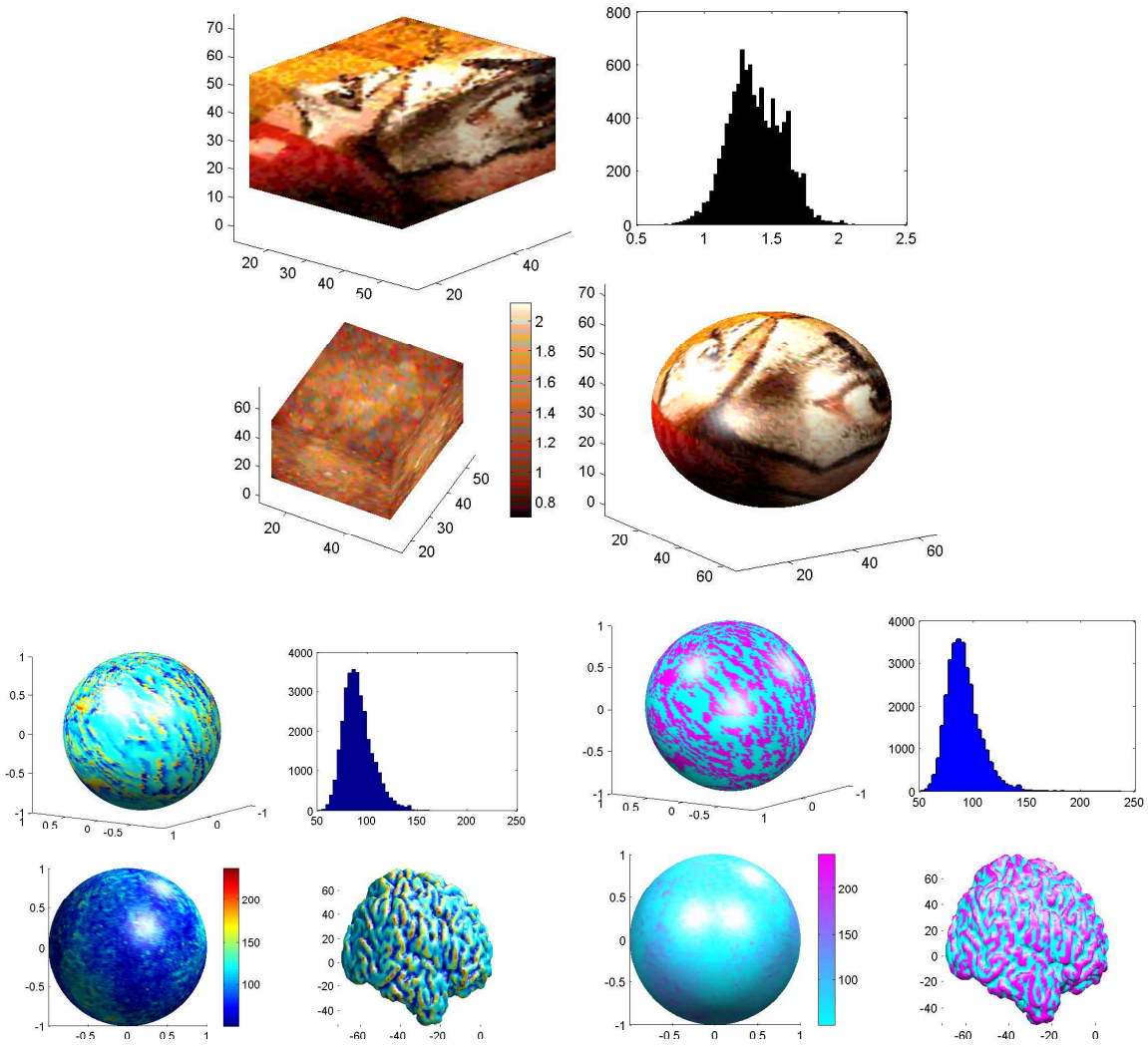

Fig. 1. Top: *Artificial example of the proposed warping algorithm. From top to bottom and left to right: The domain surface, with a picture painted on it to help in visualizing the computed map; histogram of the Lipschitz constant (note how it is concentrated around a single value); color coded distribution of the Lipschitz constant for the computed map; and mapped texture following the computed map.* **Bottom, left and right:** *Example of mapping between the cortex and a sphere. The order is the same as in the previous figure, but now the domain and target surface are colored with a curvature-based color code. Note once again the concentration of the Lipschitz constant for the computed map. On the middle, the texture map corresponding to* $I_H(x)$ *as described in the text is used. On the right, the texture is* $\max(I_H(x), \delta)$ *for a user selected value of the threshold* δ*, which highlights the gyral crests.*

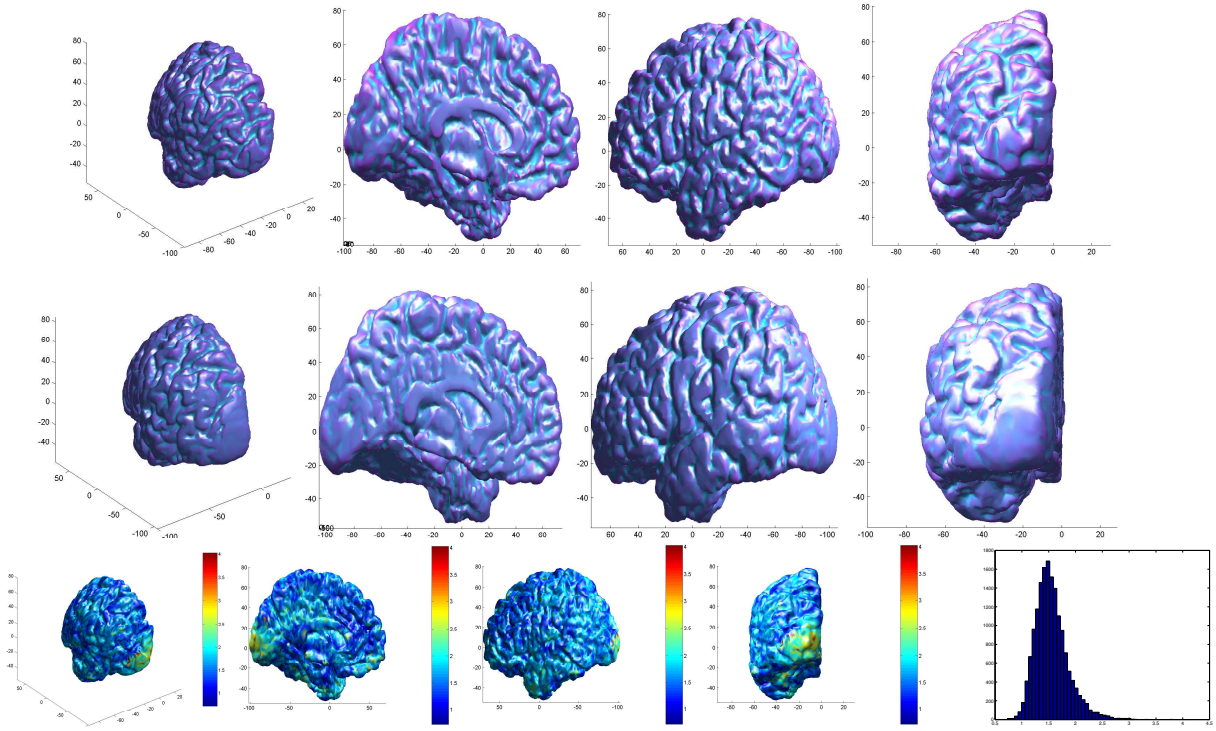

Fig. 2. *Warping between the cortical surfaces of two brains. In the first row, we show* 4 *views of* $\mathbb{B}_1$*: posterior, medial, lateral and directly viewing the occipital cortex. The corresponding* 4 *views of* $\mathbb{B}_2$ *are shown on second row. In the third row, we show* $\mathbb{B}_1$ *with texture* $I(x_i) = L_i(\Phi)$ *which can interpreted as a measure of local deformation needed to match* $x_i \in \mathbb{B}_1$ *to* $\Phi(x_i) \in \mathbb{B}_2$*. Relatively little deformation (blue colors) is required to match features across subjects on the flat interhemispheric surface (second image in the second row). This is consistent with the lower variability of the gyral pattern in the cingulate and medial frontal cortices. By contrast, there is significant expansion required to match the posterior occipital cortices of these two subjects, especially in the occipital poles which are the target of many functional imaging studies of vision. The final panel in the figure shows the corresponding histogram for* $L_i(\phi)$*, the local Lipschitz constants of the map.*

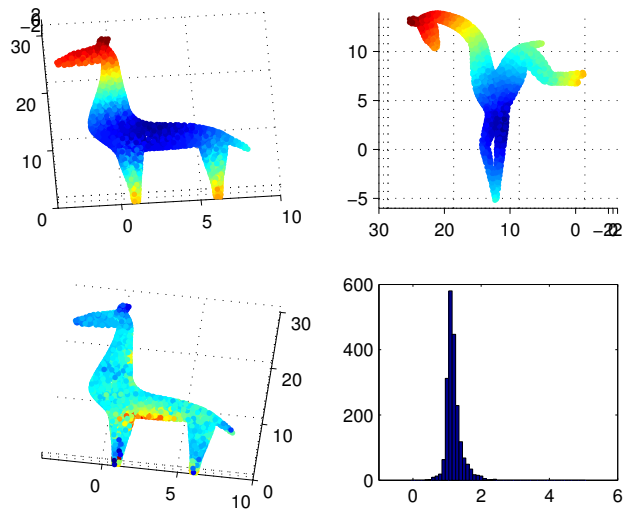

Fig. 3. *Mapping results for isometric invariant 3D shapes.*

Intrinsic and Extrinsic Analysis on Computational Anatomy

Anqi Qiu[1], Laurent Younes[1], Michael I. Miller[1]

Center for Imaging Science, The Johns Hopkins University

Abstract. We present intrinsic and extrinsic methods for studying anatomical coordinates in order to perform statistical inference on random physiological signals F across clinical populations. To do so, we introduce generalized partition functions of the coordinates, $\psi(x), x \in \mathcal{M}$, which are used to construct a random field model of F on $\mathcal{M}$. In the intrinsic analysis, such partition functions are defined intrinsically for individual anatomical coordinate based on Courant's theorem on nodal analysis via self adjoint operators. On the contrary to the intrinsic method, the extrinsic method needs only one set of partition functions for a template coordinate system, and then applied to each anatomical coordinate system via transformation. For illustration, we give clinical studies on cortical thickness for each of these methods.

1 Introduction

Computational Anatomy (CA) [1] is a discipline which is evolving rapidly worldwide. The three major areas in CA (i) construction of anatomical manifolds, (ii) metric comparison of anatomical manifolds, and (iii) large deviation testing and statistic inference are proceeding in many groups concurrently. This paper discusses two approaches to statistical inference in disease populations on anatomical manifolds, what we shall term intrinsic and extrinsic methods.

The specificity of the statistical analysis on anatomical manifolds comes from the fact that each observation consists of a physiological signal, F, defined on a manifold, $\mathcal{M}$, which is subject dependent. We refer to the observed signal , $F(x), x \in \mathcal{M}$ as the signal in anatomical coordinates. There are numerous examples, like functional activity, or cortical depth on brain surfaces, fiber orientation from DTI images etc.

To accommodate a statistical analysis, the most general approach is to consider that the observations arise from an infinite dimensional random process $(\mathcal{M}, F)$ that includes the manifold and the physiological observation that is carried by it. The anatomical part of the process $(\mathcal{M})$ can have important clinical implications, and its variations have been showed to be related to pathological states (e.g. [2, 3]). However, our focus is on the analysis of the variations of the physiological signal, F, independently from the anatomical variation that is considered as a nuisance component of the process. The question is how to characterize group differences in F that subsist disregarding variation in the anatomy.

A natural construction for the distribution of $(\mathcal{M}, F)$ is to first model the anatomical part, $\mathcal{M}$, then the physiological part conditional to $\mathcal{M}$. The latter distribution therefore models a random field on $\mathcal{M}$. In this paper, we will consider representations of this random field by a possibly infinite number of random variables given by

$$F_i = \int_{\mathcal{M}} F(x)\psi_i(x)ds(x) \tag{1}$$

where s is the volume form and ψ_i is a function on $\mathcal{M}$. The functions $\psi_1, \psi_2, \ldots$ obviously depend on $\mathcal{M}$, but we will assume that they are completely specified by it (they are deterministic for the conditional distribution given $\mathcal{M}$). We shall refer to them as *generalized partition functions* of the manifold of $\mathcal{M}$. A simple case can be associated to the decomposition model $F(x) = \sum_i F_i\psi_i(x)$ where the ψ_i form an orthonormal basis of $L^2(\mathcal{M})$. They can also be a simple set of indicator functions, leading to a real decomposition of F on a partition. We do not assume here that the representation $F_1, F_2, \ldots$ is exact (i.e., that F can be reconstructed from it), but we expect that it will conserve the necessary information to perform statistical analysis, that will amount to performing inference on the random physiological measurements reduced to the coefficients $F_i, i = 1, 2, \ldots$. The real challenge here is of course the construction of the generalized partition ψ in anatomical coordinates (i.e., defining $\psi_1, \psi_2, \ldots$ from $\mathcal{M}$), and to ensure that the construction is done so that the features F_i retain comparable qualitative interpretation across multiple individuals $\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(n)}$. Such an approach will be called an *intrinsic analysis*, because it relied on computation that only depend on the observation $(\mathcal{M}, F)$.

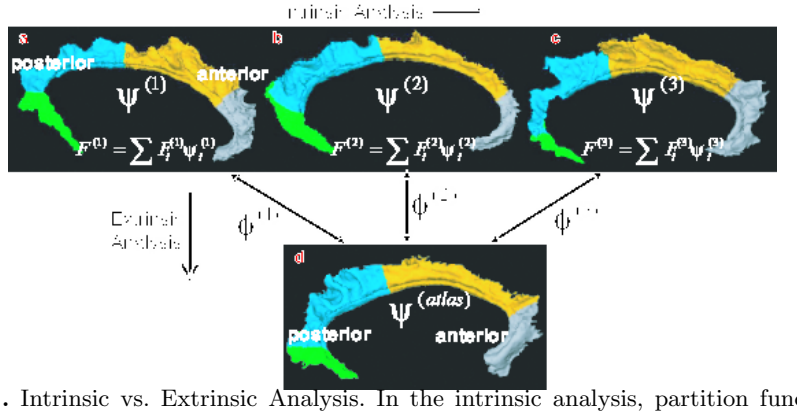

Fig. 1. Intrinsic vs. Extrinsic Analysis. In the intrinsic analysis, partition functions $\psi^{(j)}$ are constructed independently for individual anatomical coordinates shown in panels (a-c). In extrinsic analysis, only one partition function is needed for template coordinates shown in panel (d) and partition functions of other anatomical coordinates are carried onto the template coordinate through mapping ϕ.

We shall develop an example of intrinsic analysis, and also compare it to template based methods that we and many groups (e.g. [4–6]) have been following in the past 10 years. As shown in Figure 1, for each anatomical configuration $\mathcal{M}$ there is a correspondence ϕ which carries the template coordinates

to the anatomy. The implicit statistical model in this framework is that $\mathcal{M}$ is a (random) deformation of the template $\mathcal{M}_{atlas}$. In the approach we have used extensively, this deformation is modeled as a diffeomorphism ϕ of the ambient space (an open subset of $\mathbb{R}^3$), such that $\mathcal{M} = \phi(\mathcal{M}_{atlas})$. The random process F is then pulled back to $\mathcal{M}_{atlas}$ by the transformation $F \rightarrow F \circ \phi$ (the later being defined on the template). It is the pulled-back version that is modeled as a random field on the fixed set $\mathcal{M}_{atlas}$, making its statistical analysis much easier. To define the generalized partition model, one starts with defining $\psi_1, \psi_2, \ldots$ once for all on $\mathcal{M}_{atlas}$, and, given an observation $(\mathcal{M}, F)$, compute the coefficient F_i by first estimating the registration ϕ such that $\mathcal{M} = \phi(\mathcal{M}_{atlas})$, then setting

$$F_i = \int_{\mathcal{M}_{atlas}} F \circ \phi(x)\psi_i(x)ds(x). \tag{2}$$

The power of this approach is of course that only one partition needs to be created, indexed over the template. We call this approach *extrinsic* analysis, since it depends on the relation between the observation and the template. The issues that can be raised on such an approach are that the resulting features depend on the choice of the template, and on the computed correspondences ϕ. The difficulty with this strategy is that for many anatomical configurations the correspondence may not be well defined. For highly curved surfaces in the cortex, the additional transformation required by the mappings may itself introduce errors which can directly influence the statistics being inferred. The results of the analysis may therefore depend on the *algorithm* that is used for estimating ϕ.

For this reason, it is attractive if possible to be able to study variations in the structure and function associated with anatomical coordinates without having to generate these bijective correspondences, that is, to use the *intrinsic* analysis. We provide such an approach in section 3, based on a classical theorem from Courant on nodal analysis and the partition of domains based on self adjoint operators.

An intrinsic approach should address several principles that sustain its well-foundness and efficiency.

1. Intrinsicality: the generalized partition functions ψ_j must be obtained directly from the anatomical coordinates as shown in Figure 1 (a-c), without the availability of the common extrinsic template coordinates and correspondence.
2. Reliability: the functions ψ_j computed from similar manifolds must also be similar.
3. Locality: to be able to obtain statistical conclusions that concern specific regions of the manifold, the functions ψ_j must be supported by a subregion of $\mathcal{M}$.

2 Intrinsic and Extrinsic Random Field Models

An interesting illustration of the difference between the intrinsic and extrinsic approaches comes from the two random field models they respectively lead to.

Intrinsic Model. Denote $L^2(\mathcal{M})$ for the set of square integrable functions on $\mathcal{M}$. Let $\psi^{\mathcal{M}} = (\psi_1, \psi_2, \ldots)$ form an orthonormal basis of $L^2(\mathcal{M})$. Then, the conditional distribution of F given $\mathcal{M}$ can be defined via the decomposition

$$F = F_0 + \sum_{k=1}^{\infty} F_k \psi_k,$$

F_0 being a given, average function, and $F_1, F_2, \ldots$ being uncorrelated Gaussian variables of respective variances $\sigma_1^2, \sigma_2^2, \ldots$. Such models have been considered in [7], with the ψ_i being the eigenfunctions of the Laplace-Beltrami operator on $\mathcal{M}$.

Extrinsic Model. In the extrinsic approach, modeling focuses on the template. Defining $\psi^{atlas} = (\psi_1, \psi_2, \ldots)$ to be an orthonormal basis of $L^2(\mathcal{M}_{atlas})$, one can model a function F defined on $\mathcal{M} = \phi(\mathcal{M}_{atlas})$ by

$$F = F_0 \circ \phi^{-1} + \sum_{k=1}^{\infty} F_k \psi_k i \circ \phi^{-1}$$

with independent F_k. Equivalently, the model is that the registered physiological measure, $F \circ \phi$ follows a distribution (as a random field on $\mathcal{M}_{atlas}$) similar to the intrinsic distribution we have discussed above.

It is a powerful tool for analyzing brain function, and has successfully been used by several groups including ours (e.g. [2, 3]). It is robust, since the choice of the basis ψ is done once for the atlas, and then transported to the observed manifolds. As we have already noticed, however, one important drawback is that the model depends on the registration ϕ which is not observed. It therefore depends on the implemented registration algorithm: different algorithms yield different models and possibly different conclusions.

This last drawback is, by construction, absent from the intrinsic approach. The issue there is that, since the ψ_i are computed independently for each manifold, it is not sure that they can be given a universal interpretation on a given class of observations (on cingulates, or planum temporale, for example). Even if the distribution of F can be assumed to be robust as a whole (by letting, in the case of a decomposition of the Laplace-Beltrami operator, the variance of F_i be a fixed function of the eigenvalue associated to ψ_i), it is not always the case for the computed values of F_i. This robustness depends on fact on the considered class of manifolds, as will be illustrated later.

Both the intrinsic and extrinsic constructions above are unsatisfactory, however, because they do not address the locality constraint that is desirable for the interpretability of the results. We now describe how the eigenfunctions of the Laplacian can be used to define intrinsic regions on the manifolds, through the notion of nodal decomposition. Nodal decomposition based on the Laplacian has been used in other areas, such as image segmentation, graph decomposition, et al. [8–10]. Moreover, the Laplace-Beltrami operator has been used in the brain studies, such as surface and function smoothing, conformal mapping [7, 11–13]

3 Intrinsic Study of Anatomical Coordinates

To generate an intrinsic partition, we start with a classical result from Courant on the partition of coordinates based on nodal domains. The nodal (or zero-crossing) lines of the eigenfunctions of a self-adjoint operator defined on the manifold separates it in a partition with a predictable number of connected components. We are particularly interested in cortical surfaces for our partition.

Courant's nodal theorem [14]: *Let L be a self adjoint differential operator and consider the differential equation $L\psi + \lambda\psi = 0$, for a domain $\mathcal{M}$ with arbitrary homogeneous boundary conditions; if its eigenfunctions are ordered according to increasing eigenvalues, then the nodes of the ith eigenfunction ψ_i divide the domain into no more than i subdomains. No assumptions are made about the number of independent variables.*

Such subdomains are called *nodal domains*. Lines where ψ_i changes sign are defined as *nodal lines*. On a triangulated mesh, ψ_i is defined on each vertex of the mesh. It may change from positive to negative without passing through zero. The discrete analogue of a *nodal domain* is a connected set of vertices on which the eigenfunction has the same, strict or loose, sign. For our partition we

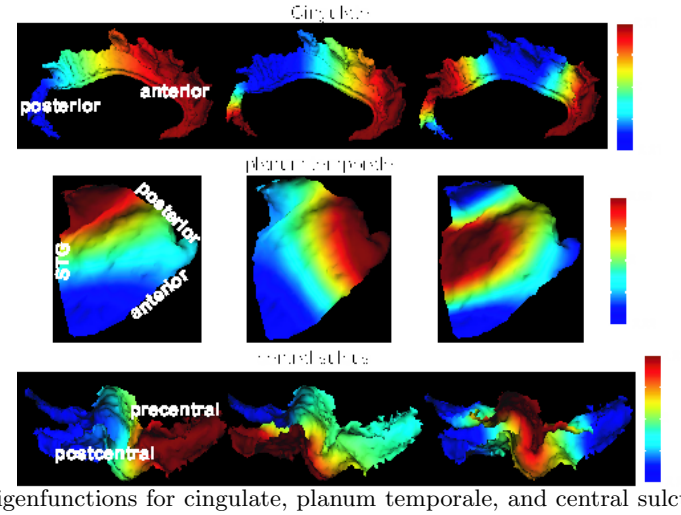

Fig. 2. Eigenfunctions for cingulate, planum temporale, and central sulcus are shown from top to bottom. The second, third, and fourth eigenfunctions are shown from left to right.

follow our previous work [7] deriving them as an orthonormal basis constructed from the Laplace-Beltrami (LB) operator, which is the extension of the Laplace operator from a regular grid to an arbitrary surface. The LB incorporates the intrinsic geometric properties of a surface, such as angle between two curves on the surface, length of a curve, and area, so that subregions of the cortical surface defined by the LB eigenfunctions are based on the geometric information of the cortex itself.

The LB spectral problem with Neumann boundary conditions for surface $\mathcal{M}$ is posed as

$$\begin{aligned} \Delta\psi(\mathbf{u}) + \lambda\psi(\mathbf{u}) &= 0, \text{ in } \mathcal{M}, \\ \int_{\mathcal{M}} |\psi(\mathbf{u})|^2 d\mathcal{M} &= 1, \\ < \nabla\psi(\mathbf{u}), \mathbf{n} >|_{\partial\mathcal{M}} &= 0, \end{aligned} \tag{3}$$

where Δ is the LB operator; $\mathbf{u}$ is the local coordinates with entries u_1 and u_2 on $\mathcal{M}$. $\mathbf{n}$ is the normal vector on the boundary of $\mathcal{M}$. A numerical solution, using a finite element implementation, of this eigenvalue problem is provided in [7].

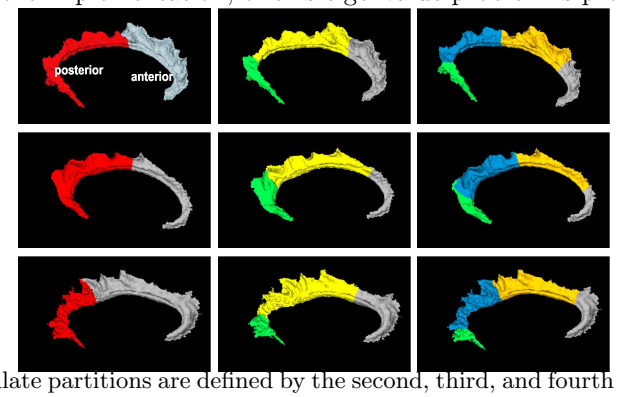

Fig. 3. Cingulate partitions are defined by the second, third, and fourth eigenfunctions. Each row gives cingulate partitions for one subject.

Eigenfunctions are ordered according to increasing eigenvalues. The first eigenfunction is a constant function, corresponding to the zero eigenvalue. Shown in Figure 2 are examples of eigenfunctions generated via the LB operator on three different cortical gyri. The automated partition on the cortical surface subregions is obtain via the nodal domains of the LB eigenfunctions. For instance, the top row in Figure 2 shows examples of the second, third and fourth eigenfunctions on cingulate. The region in red has positive values, while the region in blue has negative values. As one goes to the higher order of eigenfunctions, red and blue regions alternate rapidly. The number of nodal domains is bounded by Courant's nodal theorem and each nodal domain has at least D_j vertices. Such nodal domains are labeled as index j, $1, 2, \cdots$, from anterior to posterior of cingulate. The nodal lines are sets of vertices as the form $\mathcal{L}(\phi_i) = \{v_n : e_{mn} \in \mathcal{M},\ \phi_i(v_m)\phi_i(v_n) < 0 \text{ and } l_m > l_n\}$, where l_m and l_n are labels of the nodal domains that vertices v_m and v_n belong to. e_{mn} is an edge between v_m and v_n on surface M. Figure 3 below illustrates the nodal domains in cingulate gyri generated from the multiple eigenfunctions. The intrinsic methodology associates to each surface a triangular sequence of domains of the form $N_{11}, N_{21}, N_{22}, N_{31}, N_{32}, N_{33}, \ldots, N_{n1}, \ldots, N_{nn}, \ldots$ where $N_{n1}, \ldots, N_{nn}$ is the nodal partictions associated to the nth eigenfunction of the Laplacian. Then the partition functions are characteristic functions of these domains, namely $\psi_{nk}(x) = 1_{N_{nk}}(x)$.

4 Extrinsic Study of Anatomical Coordinates

The previous definition clearly addresses the intrinsicality and locality principle. Like for the random field model, the reliability issue is not as straightforward and can fail to be true for some surfaces. This can be related to possible multiplicity of eigenvalues (see Figure 4) for the Laplacian, which is often associated to symmetries within the manifold.

The extrinsic version of the nodal analysis simply involves carrying the generalized partition from the template onto each of the individual anatomical configurations, according to equation (2).

The correspondences (ϕ) are constructed using Large Deformation Diffeomorphic Metric Mapping methods with landmark data. Landmark data is one among several instances of modalities that can be aligned with LDDMM, together with images, tensors (DTI), unlabelled points and measures, courants and surfaces ... [15–18].

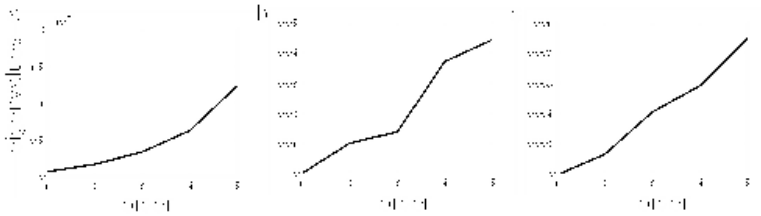

Fig. 4. Panels (a-c) show the first 5 eigenvalues of cingulate, planum temporale, and central sulcus that are shown in Figure 2. Panel (b) shows that the second and third eigenvalues of the planum temporale are close to each other. This evidence is also observed on the second row of Figure 2.

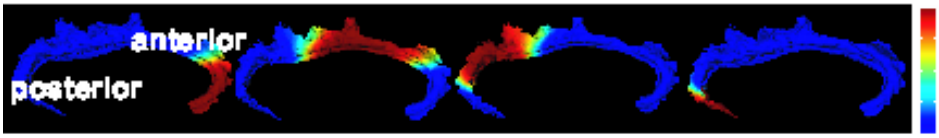

Fig. 5. Probability maps of each region as a function of cortical location.

In this experiment, we work with a database of 113 landmarked cingulates, each of them provided with its own collection of (intrinsic) nodal functions, that will be denoted $\psi_{nk}^{(j)}, n \geq 2, k \leq n$, j indexing the cingulates.

For the extrinsic representation, we selected the cingulate of one healthy subject as template, then used the landmark matching procedure[15, 19] to register the dataset on the template. Letting $\phi^{(j)}$ be the deformation carrying the left (or right) template to the left (or right) cingulate of subject j, the intrinsic nodal functions $\psi_{nk}^{(j)}$ were remapped as $\psi_{nk}^{(j)} \circ \phi^{(j)}$ to the cingulate template. We then measured how each class of remapped domains overlaps by computing, for all points x in each the template.

$$P_{nk}(x) = \frac{1}{J} \sum_{j=1}^{J} \psi_{nk}^{(j)} \circ \phi^{(j)}(x) \ , \ k = 1, \ldots, n \ .$$

Figure 5 shows the probability maps for each nodal domain determined by the fourth LB eigenfunction in the case $n = 4, k = 1, 2, 3, 4$. For instance, the left top panel shows that the region in red certainly belongs to the first nodal domain N_{41} while the region in blue is not in this domain and the region colored from red, yellow, green to blue is the transition region between two nodal domains due to the variation of the brain anatomy across subjects.

5 Statistics in Nodal Domains

When trying to characterize the neuroanatomical and functional abnormalities associated with a specific neuropsychiatric disorder, certain fundamental questions always arise. The most important one is asked to detect group difference in anatomical structure and function of a particular brain region. This can be quite challenging due to highly variable brain structure across subjects and high dimensionality of data (e.g. thickness function defined on the cortical surface) compared to the small number of individuals in each group. Partitioning the cortex into different domains seems a way to overcome these issues by reducing the complexity of brain structure into domains and reducing dimensionality of data. Below we define a statistic corresponding to each nodal domain for detecting group differences.

For a scalar function $F^{(j)}(x)$ defined on cortical surface j, such as cortical thickness, curvature, or functional response, we define the normalized scalar measurement $\overline{F}_{nk}^{(j)}$ within nodal domain $N_{nk}^{(j)}$ by

$$\overline{F}_{nk}^{(j)} = \frac{\int_{\mathcal{M}^{(j)}} F^{(j)}(x)\psi_{nk}^{(j)}(x)ds(x)}{\int_{\mathcal{M}^{(j)}} \psi_{nk}^{(j)}(x)ds(x)} \ , k = 1, 2, \cdots, n \ , \tag{4}$$

where $\psi_{nk}^{(j)}$ is as before the indicator function of domain N_{nk}. Again, $\psi_{nk}^{(j)}$ is computed from individual cortical surface in the intrinsic analysis, while only one set of partition functions $\psi_{nk}^{(atlas)}$ is used in the extrinsic analysis and then transported to the other surfaces via diffeomorphic regsitration. Group difference of $\overline{F}_{nk}^{(j)}$ within each nodal domain is detected to perform statistical analysis, such as t-test, rank sum test.

5.1 Clinical Studies

Extrinsic Study. Our first example discusses the use of the extrinsic random field model (section 2) to characterize group differences of structural or functional measurement on the cortical surface. We give an example of the clinical study in schizophrenia for the subcortical structure – left planum temporale (lPT), which is the associate auditory system located on the left superior temporal gyrus. We have assessed the lPT thickness in 10 healthy subjects, 10 subjects with schizophrenia, matched with age and gender. We first chose one lPT of a healthy subject as template and all others were deformed into the template space using the diffeomorphic surface matching approach [16]. The cortical thickness maps on the other lPTs were then remapped to the template. The

average thickness maps within the healthy and schizophrenic groups are shown in Figure 7(a,b), respectively. These two panels suggest that there are similar patterns in both groups, that is, the lPT is thin at the bottom of Heschl's sulcus (HS), then progressively thicker away from HS, and finally thinner towards the posterior ramus. Figure 7(c) shows pairwise differences in the average thickness maps between the two groups. Red denotes the region where the lPT is thicker in the healthy control group than in the schizophrenic group while blue represents the region where the lPT is thinner in the healthy group than in the schizophrenic group. To demonstrate the group difference, each thickness map was expanded as a linear combination of eigenfunctions shown in Figure 6 with a coefficient vector $f^{(i,j)}$, where i is the index of groups and j is the index of subjects in group i. The Hotelling's T^2 test was incrementally performed on $F^1 = \{f^{(1,j)}, j = 1, 2, \cdots, N\}$ and $F^2 = \{f^{(2,j)}, j = 1, 2, \cdots, N\}$ when $N = 1, 2, 3, \cdots$. We found that the group difference in thickness occurs in the second, third, and fourth eigenfunctions shown in Figure 6(a,b,c), respectively. The significant group difference reconstructed by these eigenfunctions is shown in panel (d) of Figure 7. Several previous studies have shown that reduced superior temporal gyrus volume is associated with hallucinations/delusions or positive formal thought disorder [20]. This with our thickness results thus implies that the region showing thickness decrease relative to healthy comparison controls is associated with these dysfunctions (hallucinations/delusions or positive formal thought disorder).

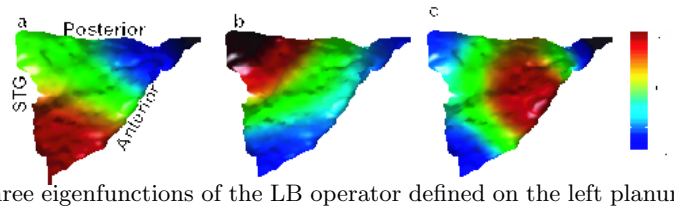

Fig. 6. Three eigenfunctions of the LB operator defined on the left planum temporale.

Intrinsic study. We study the thickness variation of cingulate in schizophrenia using the approach described in Section 5 in the intrinsic analysis. The cingulate gyrus is the part of the cerebrum that lies closest to the limbic system, above the corpus callosum. It provides a pathway from the thalamus to the hippocampus, is responsible for focusing attention on emotionally significant events, and for associating memories to smells and to pain. It has been considered as prominent brain structure related to schizophrenic symptoms. The cortical thickness of the cingulate gyrus estimated from MRI is one of quantitative morphometric measurements, which indirectly implies cellular changes in density as well as soma size and directly indicates the change in gray matter volume and anatomical shape that may be associated with schizophrenia. Figure 8 gives examples of thickness maps of a control and schizophrenic patient. The interesting questions arising are how to detect group difference on the cingulate surface and (if there is one) how to localize such a change on the cingulate surface. To answer these questions, we applied the intrinsic method. Two-sided rank sum tests were performed on the coefficients $\overline{F}_{nk}$ to detect and quantify nonuniform abnormalities

of the cortical thickness on cingulate surfaces in 20 subjects with schizophrenia as compared to 20 healthy subjects matched for gender and age. We found distinct pattern of thickness on the left cingulate gyrus. In terms of power of statistical testing, the most significant change in thickness between healthy and schizophrenic groups is on the region colored in blue shown in the last column of Figure 3. As for the right side of the cingulate gyrus, distinct pattern of thickness is shown in both anterior and posterior segments of the cingulate gyrus as well.

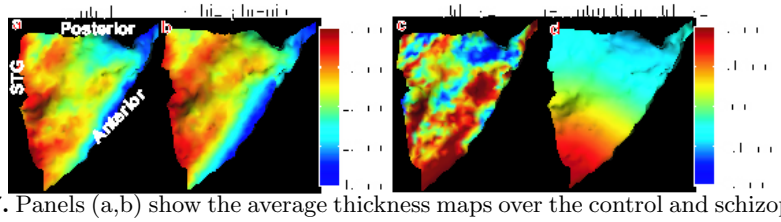

Fig. 7. Panels (a,b) show the average thickness maps over the control and schizophrenic groups on the left planum temporale. Panel (c) shows the difference in thickness between the control and schizophrenic groups. Panel (d) illustrate the significantly different pattern in thickness between these two groups, which is constructed using eigenfunctions shown in Figure 6.

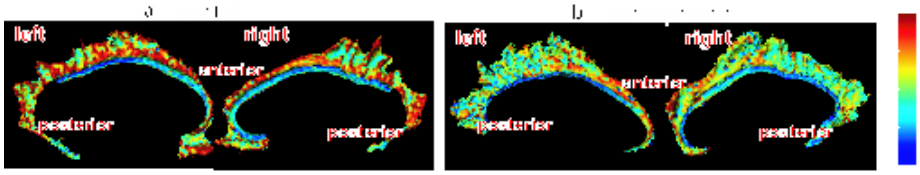

Fig. 8. Panel (a) shows cortical thickness maps on the left and right cingulates of one healthy control subject. Panel (b) shows thickness maps for one schizophrenic patient.

6 Conclusion

This paper presents the intrinsic and extrinsic methods both of which are powerful tools to study statistical inference on physiological random signals F in anatomical coordinates. The extrinsic method has been used to study the anatomical variability in the last decade. An advantage of the extrinsic analysis is that only one set of partition functions is needed for the template coordinate. But this requires to find the correspondence between anatomies that carries these partition functions to each anatomical coordinate. The intrinsic method may become an alternative way to study statistical inference on F across populations if such a correspondence between anatomies is not well defined. The tradeoff of this method is that the set of partition functions has to be found for every anatomical coordinate, which is solved based on Courant's theorem on nodal analysis via the LB operator in this paper.

Acknowledgements

This work was supported by NIH grants: R01 MH 064838, R01 EB00975, P41 RR15241, and P20 MH071616. The authors would like to thanks Dr. Csernansky and Dr. Wang from Washington University in St. Louis for cingulate data,

Dr. Ratnanather and Dr. Barta from the Johns Hopkins University for planum temporale and central sulcus data.

References

1. Grenander, U., Miller, M.I.: Computational anatomy: An emerging discipline. Quarterly of Applied Mathematics **LVI**(4) (1998) 617–694
2. Csernansky, J.G., Joshi, S., Wang, L., Hallerparallel, J.W., Gado, M., Miller, J.P., Grenander, U., Miller, M.I.: Hippocampal morphometry in schizophrenia by high dimensional brain mapping. PNAS **95** (1998) 11406–11411
3. Csernansky, J.G., Wang, L., Joshi, S.C., Ratnanather, J.T., Miller, M.I.: Computational anatomy and neuropsychiatric disease: probabilistic assessment of variation and statistical inference of group difference, hemispheric asymmetry, and time-dependent change. NeuroImage (2004) S139–S150
4. Thompson, P.M., Hayashi, K.M., Sowell, E.R., Gogtay, N., Giedd, J.N., Rapoport, J.L., de Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., Doddrell, D.M., Wang, Y., van Erp, T.G., Cannon, T.D., Toga, A.W.: Mapping cortical change in alzheimer's disease, brain development, and schizophrenia. NeuroImage **23** (2004) S2–S18
5. Van Essen, D.: Surface-based approaches to spatial localization and registration in primate cerebral cortex. NeuroImage **23** (2004) s97–s107
6. Fischl, B., Sereno, M.I., Dale, A.M.: Cortical surface-based analysis II: inflation, flattening, and a surface-based coordinate system. NeuroImage **9** (1999) 195–207
7. Qiu, A., Bitouk, D., Miller, M.I.: Smooth functional and structural maps on the neocortex via orthonormal bases of the Laplace-Beltrami operator. submitted to IEEE transactions on medical imaging (2005)
8. Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Nadler, B., Warner, F., Zucker, S.W.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. PNAS **102** (2005) 7426–7431
9. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Tran PAMI **22** (2000) 888–C905
10. Belkin, M., Niyogi, P.: Semi-supervised learning on riemannian manifolds. Machine Learning **56** (2004) 209–239
11. Angenent, S., Haker, S., Tannenbaum, A., Kikinis, R.: Laplace-beltrami operator and brain surface flattening. IEEE Trans. Medical Imaging **18** (1999) 700–711
12. Gu, X., Wang, Y., Chan, T., Thompson, P., Yau, S.: Genus zero surface conformal mapping and its application to brain surface mapping. Information Processing Medical Imaging (2003)
13. Memoli, F., Sapiro, G., Thompson, P.: Implicit brain imaging. NeuroImage **23** (2004) S179–S188
14. Courant, R., Hilbert, D.: Methods of Mathematical Physics. Volume 1. Interscience Publishers, New York and London (1953)
15. Joshi, S.C., Miller, M.I.: Landmark matching via large deformation diffeomorphisms. IEEE Trans. Image Processing **9**(8) (2000) 1357–1370
16. Vaillant, M., Glaunès, J.: Surface matching via currents. Lecture Notes in Comp. Sci.: Inform. Proc. in Med. Imaging **3565** (2005) 381–392
17. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision **61**(2) (2005) 139–157

18. Cao, Y., Miller, M., Winslow, R., Younes, L.: Large deformation diffeomorphic metric mapping of vector fields. IEEE Trans. Med. Imag. **24** (2005) 1216–1230
19. Allassonnière, S., Trouvé, A., Younes, L.: Geodesic shooting and diffeomorphic matching via textured meshes. In: EMMCVPR. (2005) 365–381
20. Shenton, M.E., Dickey, C.C., Frumin, M., McCarley, R.W.: A review of MRI findings in schizophrenia. Schizophr Res. **49** (2001) 1–52

A Continuous 3-D Medial Shape Model With Branching

Timothy B. Terriberry[1] and Guido Gerig[1,2]

[1] Dept. of Computer Science
[2] Dept. of Psychiatry
Univ. of North Carolina, Chapel Hill, NC 27599, USA
{tterribe,gerig}@cs.unc.edu

Abstract. We present a new, continuously defined three-dimensional medial shape representation based on subdivision surfaces. The shape is modeled via its medial axis, and the associated boundary is computed directly from this axis at every point. Our model is parameterized over a fixed domain, so comparison among different shapes is possible. It is the first such model to support branch curves, which allows it to represent complex medial axes with more than one medial sheet.

1 Introduction

Medial shape models are powerful tools for shape analysis since they can represent complex biological shapes using a relatively small number of parameters. Defined as the locus of the centers and radii of the maximally inscribed balls of an object, they also have a natural connection with the cognitive processes of the human visual system [1]. They were originally introduced by Blum for 2-D shape analysis, where they can be organized into a tree structure with a finite number of singular points [2]. They are also well-suited for 2-D shape synthesis, as they can be modeled by a series of continuous curves with a finite number of constraints at their end-points [3]. A model of the centers and radii of the inscribed balls is sufficient to reconstruct the complete boundary of the shape.

The singular structure of the medial axis has been thoroughly analyzed in three dimensions [4], and the differential geometry of its associated boundary has undergone rigorous mathematical treatment in arbitrary dimension [5]. However, its adaptation to shape synthesis in three dimensions has proven difficult.

The primary challenge is that in 3-D, unlike in 2-D, the medial axis contains infinitely many singular points, e.g., along its edge. Each of these points requires a boundary condition to be satisfied in order for the associated boundary of the object to be closed. However, without a shape model designed explicitly to enforce these boundary conditions, a straightforward model design will only have a finite number of parameters to adjust to satisfy them.

Yushkevich introduced the first three-dimensional continuous medial model, which was based on B-spline curves, and supported only a single medial sheet [6]. Instead of trying to adjust control-point values to force the boundary conditions

to be true on the edge of the spline domain, he adjusted the radius function on the edge to force the boundary conditions to be true somewhere in the interior. Yushkevich then implicitly solved for the curve where this occurred, and defined this to be the edge of the axis.

Unfortunately, this meant that the edge of every shape model had a different location in the domain, and due to the unconstrained, irregular shape, there was no straightforward mapping between the interiors of different models. This makes the representation unsuitable for comparing different shapes or performing shape statistics across a population. It also makes it impossible to join multiple sheets, as there is no way to make the sheets intersect along a specific curve without solving high degree polynomial equations, much less obey boundary conditions.

Yushkevich et al. later describe an approach that produces an explicit domain by interpolating a potential function ρ across the medial axis instead of the radius [7]. The radius is recovered by numerically solving a differential equation with boundary conditions via finite element methods up to sufficient accuracy to get a finely sampled shape model. However, when connecting multiple sheets, using a potential function instead of an explicit representation of the radius does not leave enough free parameters to ensure both that the necessary boundary conditions are satisfied and that the radii are equal where the sheets meet. We are unaware of an existing solution to this problem.

We take a different approach. We use Catmull-Clark subdivision surfaces with what we shall term an *ordinary, corner-free boundary* to model the medial sheets. The boundary conditions are then enforced by modifying the patches at the edges of the sheet to use an interpolating spline. This allows us to reduce the degree of the polynomial equations for the boundary conditions from 12 to 2, yielding an explicit and efficient solution. The solutions of these equations are then used to construct a "control curve" which replaces the outer layer of control points, effectively providing the infinite number of free parameters needed to enforce the condition everywhere. These two key ideas—interpolating splines and control curves—are what makes this approach possible. We begin by giving a mathematical description of the medial axis and the necessary boundary conditions, and then outline our method of enforcing them.

2 Geometry of the Medial Axis

Damon formally describes the medial axis as a special kind of Whitney stratified set [5], a stratification into smooth manifold pieces of codimension one, with their boundaries and corners classified into different smooth strata of higher codimension. An associated radius function can be used to compute a multivalued *radial vector field*, S, over the axis. These radial vectors, also called *spokes*, point from a point on the axis to the corresponding points on the associated boundary, and are normal to the boundary at these points. The medial axis and its associated radial vector field are a special type of what Damon terms a *skeletal structure* [5].

For a generic genus zero object, the medial axis is composed of smooth, two-dimensional manifolds called *medial sheets*, whose boundaries form one-

dimensional curves called *edge curves*. Three sheets may be connected via a smooth one-dimensional curve called a *branch curve*. Branch curves end when they intersect an edge curve of one of the three sheets, at a *fin creation point*. Finally, four branch curves joining six medial sheets may intersect at a *six-junction point*. These are all of the singularities that may occur in three dimensions [4]. We will address them all in this work, except for six-junction points.

On a medial sheet's interior, there are exactly two values of the radial vector field, one pointing out from each side. Denoting the field S in terms of the radius and a unit radial vector field, $S = rU$, the two values of U at each smooth point and the corresponding points on the associated boundary are given by

$$U^{\pm} = -\nabla r \pm \sqrt{1 - \|\nabla r\|^2} \cdot \mathcal{N} \ . \qquad \mathcal{B} = m + rU \tag{1}$$

Here $\mathcal{N}$ is the unit normal vector for the medial sheet, and ∇r is the *Riemannian gradient* defined on the manifold. Let $m(u, v)$ be a local parameterization of the manifold and $\mathbf{I}_m$ its first fundamental form. Then ∇r is given by

$$\nabla r \triangleq \begin{bmatrix} m_u \ m_v \end{bmatrix} \mathbf{I}_m^{-1} \begin{bmatrix} r_u \\ r_v \end{bmatrix} \ , \qquad \mathbf{I}_m \triangleq \begin{bmatrix} E_m \ F_m \\ F_m \ G_m \end{bmatrix} \triangleq \begin{bmatrix} m_u \cdot m_u \ m_u \cdot m_v \\ m_v \cdot m_u \ m_v \cdot m_v \end{bmatrix} \ . \tag{2}$$

These radial vectors are arranged symmetrically about the tangent plane. One can see that the vector $U^+ + U^-$ points in the $-\nabla r$ direction, which is tangent to the medial sheet, and $U^+ - U^-$ points in the normal direction.

The values at singular points are obtained by smooth extension from the locally neighboring medial sheets. Along the edge of a medial sheet, the top spokes must meet the bottom spokes, or the surface will not be closed. As can be seen from the coefficient of $\mathcal{N}$ in (1), this occurs precisely when

$$\|\nabla r\| = 1 \ . \tag{3}$$

This "edge constraint" makes the normal component zero and the tangential components equal. Everywhere else on the medial axis, $\|\nabla r\| < 1$ must hold.

Along a branch curve, if one has three medial sheets $m^{(i)}$ oriented so that the bottom of one is adjacent to the top of the next, then the restriction becomes [3]

$$\nabla r^{(i \oplus 2)} - \nabla r^{(i \oplus 1)} = \mathcal{N}^{(i)} \cdot \sqrt{1 - \|\nabla r^{(i)}\|^2} \ , \tag{4}$$

where $\oplus$ denotes addition modulo 3. At fin creation points, $\|\nabla r^{(i)}\|$ goes to 1 for one of the sheets, causing the constraint in (4) to disappear.

A medial axis (m, r) which satisfies these boundary conditions will be closed, but there may be singularities on the reconstructed boundary itself, causing kinks and overfoldings. These occur in areas of concave curvature, where the radius extends "too far" into the concavity. Damon outlines a set of conditions to check which ensure no such illegalities occur [5]. They are based on the *radial shape operator*, S_{rad}, and a related *edge shape operator*, S_E, which measure the change in U for an infinitesimal step along a medial sheet. The radial shape operator is not a shape operator in the traditional differential geometric sense, but it *is* a

linear operator on the tangent space of m, although not necessarily self-adjoint. A matrix representation can be computed from second derivatives of m and r, and its eigenvalues, called the *principal radial curvatures*, can be used to check for overfolding of the boundary. When combined with a measure defined on the medial axis, it can also be used to compute *skeletal integrals* over the boundary or interior of the object in medial coordinates [8]. One of the advantages of a continuous representation is that this S_{rad} can be computed analytically.

3 Catmull-Clark Subdivision on the Medial Axis

Catmull-Clark subdivision surfaces are a generalization of B-spline knot insertion to meshes of arbitrary topology [9]. They were initially created as a graphics primitive to represent the boundary of closed objects. A continuous surface, called the *limit surface*, is created from an initial, discrete mesh by recursive subdivision. After the first level of subdivision, all of the faces are quadrilaterals, and every new vertex will have valence four.

Any vertex of valence four is thus called an *ordinary* vertex, while the remaining vertices are *extraordinary*. After the second level of subdivision, every face has at most one extraordinary vertex. Away from an extraordinary vertex, the limit surface behaves exactly like a B-spline patch, and thus the surface and its derivatives can be evaluated efficiently at arbitrary points. Stam showed how, with some one-time setup, the limit surface could also be efficiently evaluated near an extraordinary vertex [10].

Catmull-Clark surfaces are everywhere C^2 continuous, except at extraordinary vertices, where they are C^1 continuous. C^1 continuity ensures that ∇r, and thus the spoke field, are continuous everywhere. Since the spoke field is normal to the boundary, a continuous spoke field will generate a G^1 continuous boundary. C^2 continuity everywhere except on a set of measure zero ensures that we can compute S_{rad} almost everywhere, and thus can check for illegalities and compute medial integrals. Subdivision surfaces are also appealing because they offer *local control*. That is, a control point only influences a small local region of the surface surrounding it, instead of the whole surface. This becomes advantageous when trying to fit a model to an image, since the derivative of most points on the boundary with respect to one of the parameters is zero.

Originally restricted to closed surfaces, there have been several strategies for handling meshes with edges and creases. Initially, Hoppe et al. proposed a set of rules for triangular subdivision surfaces [11], which have a straightforward adaptation to quadrilateral subdivision surfaces—including Catmull-Clark surfaces—as described by, e.g., DeRose et al. [12] and Warren and Schaeffer [13]. The latter also describe a simple method of implementing these rules, however neither proved that the rules produced C^1 limit surfaces. A set of rules that are provably C^1 everywhere were proposed by Biermann et al., addressing problems with extraordinary edge vertices and convex and concave corners [14].

For simplicity, we require that the edge of the mesh contain only ordinary vertices (which will be valence three, not four). For reasons that will become

clear in the next section, we also disallow corners, i.e., no two adjacent edges on a crease or edge curve can belong to the same face. Besides being a practical requirement, eliminating corners is also desirable for many of the objects we wish to model. Under these restrictions, all of the previous rules for handling edges and creases are identical. They effectively cause the edge of the mesh to be interpolated as a one-dimensional B-spline curve, unaffected by other control points in the interior of the mesh. This makes joining three medial sheets along a branch curve straightforward. At a fin creation point, the fin still ends in a corner and a *dart vertex* is used to merge the crease smoothly into the surface.

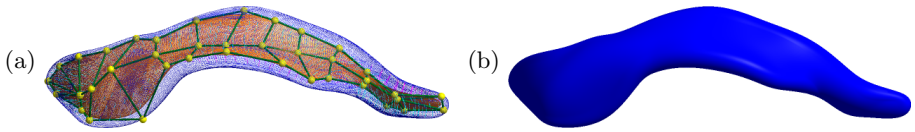

Fig. 1. A continuous medial model defined on a mesh with ordinary, corner-free boundary, depicting a healthy left lateral ventricle. (a) The mesh, medial axis, and radial vector field. (b) The enclosing boundary.

We call the result a mesh with ordinary, corner-free boundary. We perform two levels of subdivision, and then construct B-spline patches, except near extraordinary vertices, which are handled via Stam's evaluation method. The next two sections address patches adjacent to an edge or branch curve.

4 Edge Patches

First, we address patches along the edge of a medial sheet. Without loss of generality, we assume the edge lies along the line $u = 0$, as illustrated in Figure 2(a).

A full 4×4 matrix of B-spline control points $\mathbf{P}$ that interpolate the limit surface $\mathcal{M} \triangleq (m, r)$ of this patch for both m and r can be constructed by adding $P_{0,j}$ points out past the boundary, where[3] $P_{0,j} \triangleq 2P_{1,j} - P_{2,j}$. Then the local portion of the medial locus is defined by the equation

$$\mathcal{M}(u,v) \triangleq \begin{bmatrix} 1 & u & u^2 & u^3 \end{bmatrix} \mathbf{B}\mathbf{P}\mathbf{B}^T \begin{bmatrix} 1 & v & v^2 & v^3 \end{bmatrix}^T \quad , \quad \mathbf{B} \triangleq \frac{1}{6} \begin{bmatrix} 1 & 4 & 1 & 0 \\ -3 & 0 & 3 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{bmatrix} \quad . \tag{5}$$

The medial sheet m is interpolated as usual, but in order to enforce the edge condition (3) along this edge, we interpolate r using a *control curve* $r_0(v)$ instead

[3] Using $P_{0,j} \triangleq \frac{1}{2}(P_{1,j} + P_{2,j})$ and replacing $P_{1,j}$ with $\tilde{P}_{1,j} \triangleq \frac{11}{8}P_{1,j} - \frac{3}{8}P_{2,j}$ helps prevent the right side of the patch from overfolding, at the cost of a small modification to the neighboring patch, while interpolating the same $u = 0$ edge.

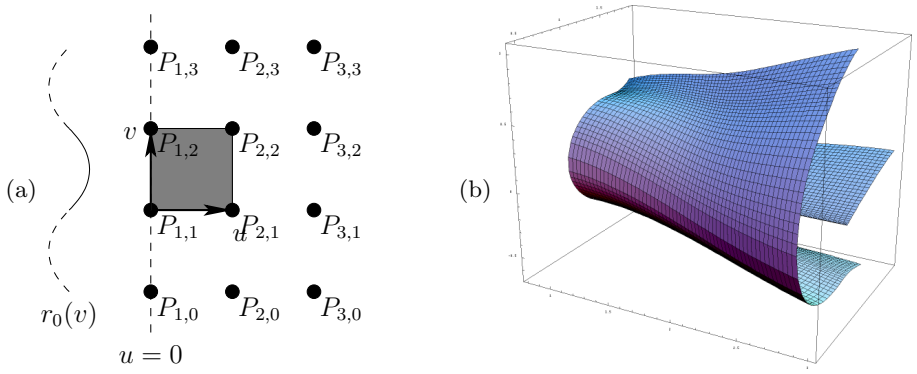

Fig. 2. (a) The control curve, $r_0(v)$, on an edge patch. (b) An example of interpolated edge patches and their associated boundary.

of a few isolated control points. This lets us ensure the condition holds at every point on the edge, which would not be possible with only a finite number of control point values as free parameters. It is important to emphasize that this curve does not represent a set of parameters explicitly defined by the modeler, but is implicitly defined by the boundary condition.

Unfortunately, the use of such a curve makes the interpolation fundamentally asymmetric with respect to the two variables u and v. We must first interpolate in the v direction, obtaining new control points for a spline in u. Then, we use a change of basis to convert the spline in u to an *interpolating spline*, which passes through its control points. Finally, we replace the control point on the left with the point from our control curve, and perform interpolation in the u direction. Because of this asymmetry, we cannot, for example, add a control curve to both the $u = 0$ and $v = 0$ edges of the same patch, which is why we disallow corners on the edge of the mesh.

The particular interpolating spline we use is the C^1 Catmull-Rom spline. We perform the change of basis by modifying the spline defined above. Replacing $\mathbf{BP}$ in equation (5) by $\mathbf{CP'}$, we use a new set of control points, $\mathbf{P'}$, given by

$$\mathbf{P'} \triangleq \mathbf{C}^{-1}\mathbf{BP} \;, \qquad \mathbf{C} \triangleq \frac{1}{2}\begin{bmatrix} 0 & 2 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -5 & 4 & -1 \\ -1 & 3 & -3 & 1 \end{bmatrix} \;. \tag{6}$$

Now, we can replace the row $P'_{0,j}$ with our control curve, and the result will still pass through the limit surface at either end of the patch. Furthermore, note that $P'_{0,j}$ has no influence on the derivative with respect to u on the right edge of the patch. Hence, whatever we use for a control curve, we will retain C^1 continuity along the right edge. We now derive an equation for the control curve, and show that we also retain C^1 continuity along the top and bottom edges.

4.1 Solving for $r_u(0, v)$

In order to ensure that the top and bottom spokes meet, we proceed to enforce the constraint in (3). With some algebra, we can decompose ∇r into a component in the m_v direction and a component in the $m_v \times (m_u \times m_v)$ direction:

$$\nabla_r^{(v)} = \frac{r_v}{\sqrt{G_m}} , \qquad \nabla_r^{(\perp v)} = \frac{r_u G_m - r_v F_m}{\sqrt{G_m(E_m G_m - F_m^2)}} . \tag{7}$$

Now, we hold $m_u(0, v)$, $m_v(0, v)$, and $r_v(0, v)$ fixed, and solve for $r_u(0, v)$. The key observation is that using a spline interpolative in u means that $r(0, v)$, and hence $r_v(0, v)$, is completely determined by the control points $P'_{1,j}$ along the edge. In particular, they do *not* depend on the values of our unknown control curve. Then substituting (7) into (3) and solving for r_u produces

$$r_u = \frac{1}{G_m} \left(r_v F_m \pm \sqrt{(G_m - r_v^2)(E_m G_m - F_m^2)} \right) . \tag{8}$$

There are two possible solutions, one of which corresponds to spokes along the crest pointing outward from the medial sheet, and the other corresponding to spokes pointing inwards. The latter is clearly illegal for a Blum medial axis, but as we can see by substituting (8) into (7), the sign of the component perpendicular to m_v is determined entirely by the sign of the square root term. The plus solution always corresponds to outward-pointing spokes.

4.2 The Complete Control Curve

Now, given $r_u(0, v)$, it is a simple matter to solve for the value of the control curve $r_0(v)$ that produces this derivative. Let $r_1(v)$, $r_2(v)$, and $r_3(v)$ be the interpolation of the three lines of ordinary control points in the v direction:

$$r_i(v) \triangleq \left[P'_{i,0}\ P'_{i,1}\ P'_{i,2}\ P'_{i,3}\right] \mathbf{B}^T \left[1\ v\ v^2\ v^3\right]^T , \qquad i = 1 \ldots 3 . \tag{9}$$

Then solving for $r_0(v)$ in terms of $r_i(v)$ yields

$$r_0(v) = \frac{1}{C_{1,0}} \left(r_u(0, v) - C_{1,1} r_1(v) - C_{1,2} r_2(v) - C_{1,3} r_3(v) \right) , \tag{10}$$

where $C_{i,j}$ is the (0-indexed) i, jth element of $\mathbf{C}$. It is important to emphasize that this formulation only works for interpolating splines, since otherwise $r_u(0, v)$ would not be independent of $r_0(v)$ in the right-hand side of (10).

Note that $r_0(v)$ is a function of $m_u(0, v)$, $m_v(0, v)$, and $r_v(0, v)$ only. As these are all first derivatives of functions obtained from B-spline interpolation, they are C^1 across patches, and hence so is $r_0(v)$. This ensures that $r(u, v)$ is C^1 with neighboring patches along the top and bottom edges, as well as the right edge. In practice, we can regain C^2 continuity on the right by using a fourth-order C^2 interpolating cubic spline (omitted for brevity). Such a spline even contains enough free parameters to enforce a second boundary condition on the same edge, although we do not make use of that in this work.

5 Branch Curve Patches

Although (4) is a succinct expression of the boundary conditions along a branch curve, it is not obvious how they should be enforced. We motivate the solution with a geometric approach. Without loss of generality, we assume our three patches are oriented so that they meet at the $u = 0$ curve in each. Then r_v and G_m are the same in each patch, and hence so is $\nabla r^{(v)}$.

Now, since all three $\mathcal{N}^{(i)}$s are perpendicular to m_v, they lie in the same plane, and hence so do the endpoints of $\nabla r^{(i)}$ and $U^{(i)\pm}$. We project everything into this plane, as illustrated in Figure 3(a). As one can see, the sums of the angles $\theta^{(i)}$ must be π, and the in-plane components of the tangent vectors $m_u^{(i)}$ must bisect these angles. This provides a full set of geometric constraints.

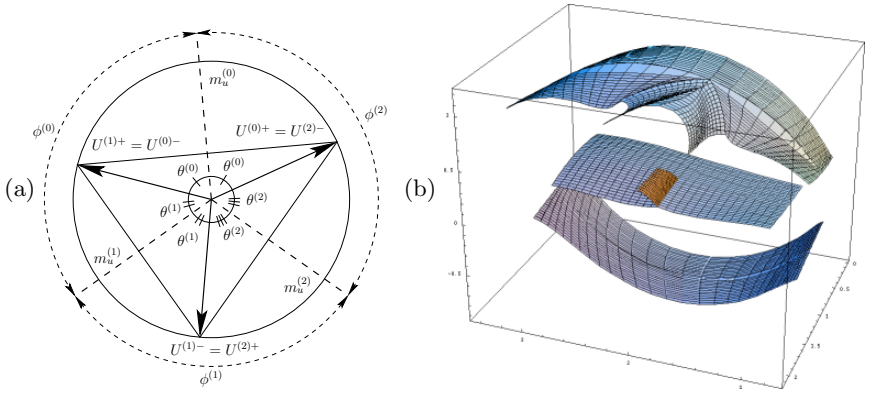

Fig. 3. (a) The cross-section of a branch curve, with m_v pointing out of the page. (b) An example of the interpolated patches near a fin creation point.

5.1 Away From Fin Creation Points

Away from a fin creation point, we can enforce these constraints by noting that $\nabla r^{(\perp v,i)} \propto \cos\theta^{(i)}$. Let $\phi^{(i)}$ be the counter-clockwise angle between the in-plane components of the tangent vectors $m_u^{(i)}$ and $m_u^{(i+1)}$, so $\phi^{(i)} = \theta^{(i)} + \theta^{(i+1)}$. Then solving for $\theta^{(i)}$ and $\nabla r^{(\perp v,i)}$ gives

$$\theta^{(i)} = \frac{1}{2}(\phi^{(i)} + \phi^{(i\oplus 2)} - \phi^{(i\oplus 1)}) \ , \qquad \nabla r^{(\perp v,i)} = -\sqrt{1 - \frac{r_v^2}{G_m}}\cos\theta^{(i)} \ . \tag{11}$$

These constraints can be enforced with a control curve exactly as in Section 4.

5.2 At a Fin Creation Point

Catmull-Clark subdivision does not produce B-spline patches around a dart vertex, but after the first two levels of subdivision, we approximate the medial surface with B-spline patches that interpolate the same branch curve. Without loss of generality, we will assume $m^{(0)}$ is the fin patch. At the fin point itself, we adjust $r_u^{(0)}(0,0)$ while holding $r^{(0)}(0,v)$ fixed so that $\|\nabla r^{(0)}\| = 1$ at $(0,0)$. Since this is a constraint at just one point, this is easily done by adjusting control points. We also note that we can rotate $m_u^{(i)}$ around m_v arbitrarily without affecting $\|\nabla r^{(i)}\|$. Thus, we rotate $m_u^{(0)}(0,0)$ around m_v until $U^{(0)-}$ and $U^{(1)+}$ coincide. This can also be done without affecting $m^{(0)}(0,v)$ by adjusting control points. We now enforce the $\|\nabla r^{(0)}\| = 1$ condition on the entire $v = 0$ edge as normal. However, this prevents us from using a control curve along the other edge. Hence, we hold the values of $r_u^{(0)}(0,v)$ and $m_u^{(0)}(0,v)$ fixed.

This fixes $\theta^{(0)}$, so we adjust $\theta^{(1)}$ and $\theta^{(2)}$. Let $\theta^{(1)} = \phi^{(0)} - \theta^{(0)} - \alpha$ and $\theta^{(2)} = \phi^{(2)} - \theta^{(0)} - \beta$. We require $\theta^{(0)} + \theta^{(1)} + \theta^{(2)} = \pi$ and for a second constraint choose $\alpha\phi^{(2)} = \beta\phi^{(0)}$, yielding

$$\alpha = \phi^{(0)} \cdot \frac{\phi^{(0)} + \phi^{(2)} - \pi - \theta^{(0)}}{\phi^{(0)} + \phi^{(2)}} \ , \quad \beta = \phi^{(2)} \cdot \frac{\phi^{(0)} + \phi^{(2)} - \pi - \theta^{(0)}}{\phi^{(0)} + \phi^{(2)}} \ . \tag{12}$$

To make the tangent planes form the correct angles between them, we rotate $m_u^{(1)}$ by α and $m_u^{(2)}$ by $-\beta$ around m_v. This can be accomplished with a control curve in a way exactly analogous to the procedure in Section 4.2 for r_u.

At the fin creation point, $\phi^{(0)}(0) + \phi^{(2)}(0) = \pi$ and $\theta^{(0)}(0) = 0$, so $\alpha(0) = \beta(0) = 0$. However, to ensure C^1 continuity, we require $\alpha_v(0) = \beta_v(0) = 0$. This occurs precisely when $\phi_v^{(0)}(0) + \phi_v^{(2)}(0) + \theta_v^{(0)}(0) = 0$. Since $\phi^{(0)} + \phi^{(2)} = \pi$ past the fin creation point, and these are functions of the unrotated $m^{(i)}$s, which join the rest of the surface with C^2 continuity at $v = 0$, this implies $\phi_v^{(0)}(0) + \phi_v^{(2)}(0) = 0$, and hence $\theta_v^{(0)}(0) = 0$ must hold. This can be enforced by adjusting control points to modify $r_{uu}^{(0)}(0,0)$ so that $\frac{\partial \|\nabla r\|^2}{\partial v} = 0$ at the fin creation point, while holding $r_u^{(0)}(0,0)$ and $r^{(0)}(0,v)$ fixed.

To transition between the procedure at a fin creation point and the procedure elsewhere on the branch curve, we smoothly blend between the two strategies in an adjacent patch using the weight function $\omega(v) = 1 - 3v^2 + 2v^3$. Note that $\omega(0) = 1$, $\omega(1) = 0$, and $\omega_v(0) = \omega_v(1) = 0$. Then in the transition region we use

$$\theta^{(i)} = \omega \cdot \theta^{(i)}_{\text{fin}} + (1 - \omega) \cdot \theta^{(i)}_{\text{branch}} \ , \qquad \alpha = \omega \cdot \alpha_{\text{fin}} \ , \qquad \beta = \omega \cdot \beta_{\text{fin}} \ . \tag{13}$$

6 Conclusion

We have presented a generative shape model using the medial axis in three dimensions. It is based on cubic splines defined on subdivision surfaces, and so can be implemented efficiently. It is the first such model to support branching. In future work, we also plan to incorporate six-junction points.

Acknowledgments

The research was supported by the NIH NIBIB grant P01 EB002779. The authors would also like to thank Paul Yushkevich and Jim Damon for several enlightening discussions, and the reviewers for their helpful comments.

References

1. Burbeck, C.A., Pizer, S.M., Morse, B.S., Ariely, D., Zauberman, G.S., Rolland, J.P.: Linking object boundaries at scale: a common mechanism for size and shape judgements. Vision Research **36** (1996) 361–372
2. Blum, H.: A transformation for extracting new descriptors of shape. In Wathen-Dunn, W., ed.: Models for the Perception of Speech and Visual Form, Cambridge, MA, MIT Press (1967) 362–380
3. Yushkevich, P.A., Fletcher, P.T., Joshi, S.C., Thall, A., Pizer, S.M.: Continuous medial representations for geometric object modeling in 2D and 3D. Image and Vision Computing **21** (2003) 17–28
4. Giblin, P.J., Kimia, B.B.: A formal classification of 3D medial axis points and their local geometry. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004) 238–251
5. Damon, J.N.: Smoothness and geometry of boundaries associated to skeletal structures I: Sufficient conditions for smoothness. Annales de l'Institut Fourier **53** (2003) 1941–1985
6. Yushkevich, P.A.: Statistical shape characterization using the medial representation. PhD thesis, University of North Carolina at Chapel Hill Department of Computer Science (2003)
7. Yushkevich, P.A., Zhang, H., Gee, J.C.: Parametric medial shape representation in 3-D via the Poisson partial differential equation with non-linear boundary conditions. In Christensen, G.E., Sonka, M., eds.: Information Processing in Medical Imaging (IPMI'05). (2005) 162–173
8. Damon, J.N.: Global geometry of regions and boundaries via skeletal and medial integrals. Preprint: http://www.math.unc.edu/Faculty/jndamon/Skel.Str.IV.v5.pdf (2005)
9. Catmull, E.E., Clark, J.H.: Recursively generated B-spline surfaces on arbitrary topological meshes. Computer Aided Design **10** (1978) 350–355
10. Stam, J.: Exact evaluation of Catmull-Clark subdivision surfaces at arbitrary parameter values. Computer Graphics **32** (1998) 395–404
11. Hoppe, H., DeRose, T., Duchamp, T., Halstead, M., Jin, H., McDonald, J., Scheitzer, J., Stuetzle, W.: Piecewise smooth surface reconstruction. Technical Report TR 94-01-01, University of Washington Department of Computer Science and Engineering (1994)
12. DeRose, T., Kass, M., Truong, T.: Subdivision surfaces in character animation. In: Proceedings of the 25th Annual SIGGRAPH Confererence on Computer Graphics and Interactive Techniques. (1998) 85–94
13. Warren, J., Schaefer, S.: A factored approach to subdivision surfaces. IEEE Computer Graphics and Applications **24** (2004) 74–81
14. Biermann, H., Levin, A., Zorin, D.: Peicewise smooth subdivision surfaces with normal control. In: Proceedings of the 27th Annual SIGGRAPH Confererence on Computer Graphics and Interactive Techniques. (2000) 113–120

Entropy-Based Particle Systems for Shape Correspondence

Joshua Cates, Miriah Meyer, P. Thomas Fletcher and Ross Whitaker

Scientific Computing and Imaging Institute
University of Utah
Salt Lake City, Utah

Abstract. This paper presents a new method for constructing statistical representations of ensembles of similar shapes. The proposed method relies on an optimal distribution of a large set of surface point correspondences, rather than the manipulation of a specific surface parameterization. The optimal configuration is defined as one in which the entropy or information content of each shape is balanced against the entropy of the ensemble of shapes. The correspondences are modeled as sets of dynamic particles that are manipulated using a gradient descent on the entropies of the shapes and the ensemble, but constrained to lie on a set of implicit surfaces. The proposed, particle-based method for finding correspondences requires very little preprocessing of data or parameter tuning, and therefore makes the problem of shape analysis more practical for a wider range of problems. This paper presents the formulation and several synthetic and real shape examples in two and three dimensions.

1 Introduction

Computing statistics on sets of shapes requires quantification of shape differences, which is a fundamentally difficult problem. A widely-used strategy for computing shape differences is to compare the positions of corresponding points among sets of shapes. Medical or biological shapes, however, are typically derived from the interfaces between organs or tissue types. Such surfaces are usually defined implicitly in the form of *segmented volumes*, rather than from explicit parameterizations or surface point samples. Thus, no defined, a priori relationship between points across surfaces exists. Correspondences must therefore be inferred from the shapes themselves, giving rise to the difficult, yet very important, *correspondence problem*.

Until recently, correspondences for shape statistics were established manually by choosing small sets of anatomically significant landmarks on organs or regions of interest, which would then serve as the basis for shape analysis. The demand for more detailed analyses on ever larger populations of subjects has rendered this approach unsatisfactory. Recently, Davies et al. [1] present methods for automatically establishing relatively dense sets of correspondences based on the information content of the set of points needed to describe an ensemble of similar shapes. These methods, however, rely on mappings between

fixed parameterizations, and because most shapes in medicine or biology are not derived parametrically, this reliance on a parameterization presents some significant drawbacks. Automatic selection of correspondences for nonparametric, point-based shape models has been explored in the context of surface registration [2], but such methods are not sufficient for shape analysis because they are typically concerned only with pairwise correspondences, and not correspondences across populations of points. Furthermore, these methods also assume a fixed set of samples on surfaces, whereas, in the context of this paper, we are given implicit surfaces (volume segmentations) and dynamically resample them as part of the correspondence selection process.

This paper presents a new method for extracting dense sets of correspondences that describe ensembles of similar shapes. The method is nonparametric and borrows technology from the computer graphics literature for representing surfaces as discrete point sets. The proposed method iteratively modifies a system of dynamic particles so that they follow trajectories across the surfaces to find positions that optimize the information content of the system. This strategy is motivated by a recognition of the inherent tradeoff between geometric accuracy and statistical simplicity. Our assertion is that each unit of complexity, or information, across the ensemble should be balanced against a unit of information on the surface. This approach provides a natural equivalence of information content and reduces the dependency on ad-hoc regularization strategies and free parameters. Since the points are not tied to a specific parameterization, the method operates directly on volumetric data, extends easily to higher dimensions or arbitrary shapes, and provides a more homogeneous geometric sampling as well as more compact statistical representations. The method draws a clear distinction between the objective function and the minimization process, and thus can more readily incorporate additional information such as adaptive surface sampling and high-order geometric information.

2 Related Work

The strategy of finding of parameterizations that minimize information content across an ensemble was first proposed by Kotcheff and Taylor [3]. They represent each two-dimensional contour as a set of N samples taken at equal intervals from a parameterization. Each shape is treated as a point sample in a $2N$-dimensional space, with associated covariance Σ and cost function, $\sum_k \log(\lambda_k + \alpha)$, where λ_k are the eigenvalues of Σ, and α is a regularization parameter that prevents the very *thinnest* modes (smallest eigenvalues) from dominating the process. This is the same as minimizing $\log|\Sigma + \alpha I|$, where I is the identity matrix, and $|\cdot|$ denotes the matrix determinant.

Davies et al. [1] propose *minimum description length* (MDL) as a cost function. In that work they use arguments regarding quantization to limit the effects of thin modes and to determine the optimal number of components that should influence the process. They propose a piecewise linear reparameterization and a hierarchical minimization scheme. Monotonicity in the reparameterizations en-

sures that those composite mappings are diffeomorphic. Davies et al. [4] propose a 3D extension to the MDL method, which relies on spherical parameterizations and subdivisions of an octahedral base shape, and smoothed updates, represented as Cauchy kernels. The parameterization must be obtained through another process such as [5], which relaxes a spherical parameterization onto an input mesh. The overall procedure is empirically satisfactory, but requires significant data preprocessing, including a sequence of optimizations—first to establish the parameterization and then on the correspondences—each of which entails a set of free parameters or inputs in addition to the segmented volumes. A significant concern with the basic MDL formulation is that the optimal solution is often one in which the correspondences all collapse to points where all the shapes in the ensemble happen to be near (e.g., crossings of many shapes). Several solutions have been proposed [4, 6], but they entail free parameters and assumptions about the quality of the initial parameterizations.

The MDL formulation is mathematically related to the min-log $|\Sigma + \alpha I|$ approach, as noted by Thodberg [6]. Styner et al. [7] describe an empirical study that shows ensemble-based statistics improve correspondences relative to pure geometric regularization, and that MDL performance is virtually the same as that of min-log $|\Sigma + \alpha I|$. This last observation is consistent with the well-known result from information theory that MDL is, in general, equivalent to minimum entropy [8].

Another body of relevant work is the recent trend in computer graphics towards representing surfaces as scattered collections of points. The advantage of so-called *point-set surfaces* is that they do not require a specific parameterization and do not impose topological limations; surfaces can be locally reconstructed or subdivided as needed [9]. A related technology in the graphics literature is the work on particle systems, which can be used to manipulate or sample [10] implicit surfaces. A particle system manipulates large sets of particles constrained to a surface using a gradient descent on radial energies that typically fall off with distance. The proposed method uses a set of interacting particle systems, one for each shape in the ensemble, to produce optimal sets of surface correspondences.

3 Methods

3.1 Entropy-Based Surface Sampling

We treat a surface as a subset of $\Re^d$, where $d = 2$ or $d = 3$ depending whether we are processing curves in the plane or surfaces in a volume, respectively. The method we describe here is limited to smooth, closed manifolds of codimension one, and we will refer to such manifolds as *surfaces*. We sample a surface $\mathcal{S} \subset \Re^d$ using a discrete set of N points that are considered random variables $Z = (X_1, X_2, \ldots, X_N)$ drawn from a probability density function (PDF), $p(X)$. We denote a realization of this PDF with lower case, and thus we have $z = (x_1, x_2, \ldots, x_N)$, where $z \in \mathcal{S}^N$. The probability of a realization x is $p(X = x)$, which we denote simply as $p(x)$.

The amount of information contained in such as random sampling is, in the limit, the differential entropy of the PDF, which is $H[X] = -\int_S p(x) \log p(x) dx = -E\{\log p(X)\}$, where $E\{\cdot\}$ is the expectation. When we have a sufficient number of samples from p, we can approximate the expectation by the sample mean [8], which gives $H[X] \approx -(1/N) \sum_i \log p(x_i)$. We must also estimate $p(x_i)$. Density functions on surfaces can be quite complex, and so we use a nonparametric, Parzen windowing estimation of this density using the samples themselves. Thus we have

$$p(x_i) \approx \frac{1}{N(N-1)} \sum_{\substack{j=1 \\ j \neq i}}^{N} G(x_i - x_j, \sigma) \tag{1}$$

where $G(x_i - x_j, \sigma)$ is a d-dimensional, isotropic Gaussian with standard deviation σ. The cost function C, is therefore an approximation of (negative) entropy: $H[X] \approx -C(x_1, \ldots, x_N) = \sum_i \log \frac{1}{N(N-1)} \sum_{j \neq i} G(x_i - x_j, \sigma)$,

In this paper, we use a gradient-descent optimization strategy to manipulate particle positions. The optimization problem is given by:

$$\hat{z} = \arg\min_z E(z) \text{ s.t. } x_1, \ldots, x_N \in \mathcal{S}. \tag{2}$$

The negated gradient of E is

$$-\frac{\partial E}{\partial x_i} = \frac{1}{\sigma 2} \frac{\sum_{\substack{j=1 \\ j \neq i}}^{N} (x_i - x_j) G(x_i - x_j, \sigma)}{\sum_{\substack{j=1 \\ j \neq i}}^{N} G(x_i - x_j, \sigma)} = \sigma^{-2} \sum_{\substack{j=1 \\ j \neq i}}^{N} (x_i - x_j) w_{ij}, \tag{3}$$

where $\sum_j w_{ij} = 1$. Thus to minimize C, the samples (which we will call *particles*) must move away from each other, and we have a set of particles moving under a repulsive force and constrained to lie on the surface. The motion of each particle is away from all of the other particles, but the forces are weighted by a Gaussian function of inter-particle distance. Interactions are therefore local for sufficiently small σ. We use a Jacobi update with forward differences, and thus each particle moves with a *time* parameter and positional update $x_i \leftarrow x_i - \gamma \frac{\partial C}{\partial x_i}$, where γ is a time step and $\gamma < \sigma 2$ for stability.

The surface constraint is specified by the zero set of a scalar function $F(x)$. This constraint is maintained, as described in several papers [10], by first projecting the gradient of the cost function onto the tangent plane of the surface (as prescribed by the method of Lagrange multipliers), and then by iterative reprojection of the particle onto the nearest root of F by the method of Newton-Raphson. Another aspect of this particle formulation is that it computes Euclidean distance between particles (in the ambient space), rather than the geodesic distance on the surface. Thus, we assume sufficiently dense samples so that nearby particles lie in the tangent planes of the zero sets of F. This is an important consideration; in cases where this assumption is not valid, such as highly convoluted surfaces, the distribution of particles may be affected by neighbors that are outside of the true manifold neighborhood. The question of particle interactions with more general distance measures remains for future work.

Finally, we must choose a σ for each particle, which we do automatically, before the positional update, using the same optimality criterion described above. The contribution to C of the ith particle is simply the probability of that particle position, and optimizing that quantity with respect to σ gives a maximum likelihood estimate of σ for the current particle configuration. We use Newton-Raphson to find σ such that $\partial p(x_i, \sigma)/\partial\sigma = 0$, which typically converges to machine precision in several iterations.

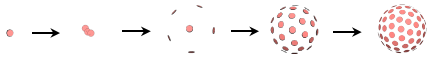

Fig. 1. A system of 100 particles (right) produced by successive splitting of a single particle (left).

There are a few important numerical considerations. We must truncate the Gaussian kernels, and so we use $G(x, \sigma) = 0$ for $|x| > 3\sigma$. This means that each particle has a finite radius of influence, and we can use a spatial binning structure to reduce the computational burden associated with particle interactions. If σ for a particle is too small, a particle will not interact with its neighbors at all, and we cannot compute updates of σ or position. In this latter case, we update the kernel size using $\sigma \leftarrow 2\sigma$, until σ is large enough for the particle to interact with its neighbors. Another numerical consideration is that the system must include bounds $\sigma_{\min}$ and $\sigma_{\max}$ to account for anomalies such as bad initial conditions, too few particles, etc. These are not critical parameters, so as long as they are set to include the minimum and maximum resolutions the system operates reliably.

The mechanism described in this section is, therefore, a self contained, self tuning system of particles that distribute themselves using repulsive forces to achieve optimal distributions. For this paper we initialize the system with a single particle that finds the nearest zero of F, which then splits (producing a new, nearby particle) at regular intervals until a specific number of particles are produced and they reach a steady state. Figure 1 shows this process for a sphere.

3.2 The Entropy of The Ensemble

An ensemble $\mathcal{E}$ is a collection of M surfaces, each with their own set of particles, i.e. $\mathcal{E} = z1, \ldots, z^M$. The ordering of the particles on each shape implies a correspondence among shapes, and thus we have a matrix of particle positions $P = x_j^k$, with point samples along the rows and shapes across the columns. We model $z^k \in \Re^{Nd}$ as an instance of a random variable Z, and propose to minimize the combined ensemble and shape cost function

$$Q = H(Z) - \sum_k H(P^k), \tag{4}$$

which favors a compact ensemble representation, balanced against a uniform distribution of particles on each surface. The different entropies are commensurate so there is no need for ad-hoc weighting of the two function terms.

For this discussion we assume that the complexity of each shape is greater than the number of examples, and so we would normally choose $N > M$. Given

the low number of examples relative to the dimensionality of the space, we must impose some conditions in order to perform the density estimation. For this work we assume a normal distribution and model $p(Z)$ parametrically using a Gaussian with covariance Σ. The entropy is then given by

$$H(Z) \approx \frac{1}{2} \log |\Sigma| = \frac{1}{2} \sum_{j=1}^{Nd} \log \lambda_j, \tag{5}$$

where $\lambda_1, ..., \lambda_{Nd}$ are the eigenvalues of Σ.

In practice, Σ will not have full rank, in which case the entropy is not finite. We must therefore regularize the problem with the addition of a diagonal matrix αI to introduce a lower bound on the eigenvalues. We estimate the covariance from the data, letting Y denote the matrix of points minus the sample mean for the ensemble, which gives $\Sigma = (1/(M-1))YY^T$. Because $N > M$, we perform the computations on the dual space (dimension M), knowing that the determinant is the same up to a constant factor of α. Thus, we have the cost function G associated with the ensemble entropy:

$$\log |\Sigma| \approx G(P) = \log \left| \frac{1}{M-1} Y^T Y, \right| \text{ and } -\frac{\partial G}{\partial P} = Y(Y^T Y + \alpha I)^{-1}. \tag{6}$$

We now see that α is a regularization on the inverse of $Y^T Y$ to account for the possibility of a diminishing determinant. The negative gradient $-\partial G/\partial P$ gives a vector of updates for the entire system, which is recomputed once per system update. This term is added to the shape-based updates described in the previous section to give the update of each particle:

$$x_j^k \leftarrow \gamma \left[-\partial G/\partial x_j^k + \partial E^k/\partial x_j^k \right]. \tag{7}$$

The stability of this update places an additional restriction on the time steps, requiring γ to be less than the reciprocal of the maximum eigenvalue of $(Y^T Y + \alpha I)^{-1}$, which is bounded by α. Thus, we have $\gamma < \alpha$, and note that α has the practical effect of preventing the system from slowing too much as it tries to reduce the thinnest dimensions of the ensemble distribution. This also suggests an annealing approach for computational efficiency (which we have used in this paper) in which α starts off somewhat large (e.g., the size of the shapes) and is incrementally reduced as the system iterates.

The choice of a Gaussian model for $p(Z = z)$ is not critical for the proposed method. The framework easily incorporates either nonparametric, or alternate parametric models. In this case, the Gaussian model allows us to make direct comparisons with the MDL method, which contains the same assumptions. Research into alternative models for Z is outside the scope of this paper and remains of interest for future work.

The method outlined above assumes a population of surfaces that are in alignment with one another. For medical image datasets, this is often not the case, and some surface registration technique must be applied as a part of the

algorithm for finding correspondences. Goodall [11], for example, suggests the point-based Procrustes method. For the results given in the following section, we assume the surface data is in alignment and leave the analysis of the stability and interplay between the proposed method and any surface registration techniques for future work. Preliminary results do suggest, however, that Procrustes alignment may be effectively applied at intervals in the proposed correspondence optimization.

4 Results and Conclusions

We begin with two experiments on closed curves in a 2D plane and a comparison with the 2D open-source Matlab MDL implementation given by Thodberg [6]. In the first experiment, we used the proposed, particle method to optimize 100 particles per shape on 24 *box-bump* shapes, similar to those described in [6]. Each shape was constructed from a set of point samples using cubic b-splines with the same rectangle of control, but with a bump added at a random location along the top of its curve. Distance transforms from these shapes were constructed using the fast-marching algorithm [12], which forms implicit contours suitable for input to the proposed algorithm. MDL correspondences were computed using 128 nodes and *mode 2* of the Matlab software, with all other parameters set to their defaults (see [6] for details). Both methods identified a single dominant mode of variation, but with different degrees of leakage into orthogonal modes. MDL lost 0.34% of the total variation from the single mode, while the proposed method lost only 0.0015%. Figure 2 illustrates the mean and three standard deviations of the first mode of the two different models. Shapes from the particle method remain more faithful to those described by the original training set, even out to three standard deviations where the MDL description breaks down. A striking observation from this experiment is how the relatively small amount of variation left in the minor modes of the MDL case produce such a significant effect on the results of shape deformations.

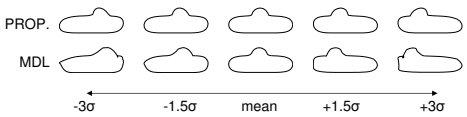

Fig. 2. A comparison of the mean and three standard deviations of the *box-bump* experiment.

The second experiment was conducted on the set of 18 hand shape contours described in [1], again applying both the particle method and MDL using the same parameters as described above. As with the box-bump data, distance transforms were generated from the spline-based contour models for input to the correspondence algorithm. In this case, we also compared results with a set of *ideal*, manually selected correspondences, which introduce anatomical knowledge of the digits. Figure 3 compares the three resulting models in the top three modes of variation to ± 3 standard deviations. A detailed analysis of the principle components showed that the proposed particle method and the manually

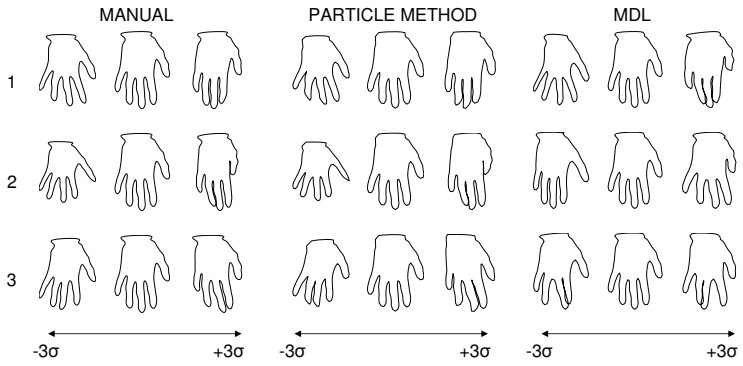

Fig. 3. The mean and three standard deviations of the top three modes of the hand models.

selected points both produce very similar models, while MDL differed significantly, particularly in first three modes.

Existing 3D MDL implementations rely on spherical parameterizations, and are therefore only capable of analyzing shapes topologically equivalent to a sphere. The particle-based method does not have this limitation. We applied the proposed method to a set of randomly chosen tori, drawn from a 2D distribution that is parameterized by the small radius r and the large radius R. Samples were chosen from a distribution with mean $r = 1, R = 2$ and $\sigma_r = 0.15, \sigma_R = 0.30$, with a rejection policy that excluded invalid tori (e.g., $r > R$). Figure 4 shows the particle system distribution across two of the torus shapes in the sample set with 250 correspondences. An analysis of the correspondences showed that the particle system discovered the two pure modes of variation, with only 0.08% leakage into smaller modes.

We applied the proposed method to a set of 20, volumetric hippocampus segmentations chosen at random from a larger data set described in Styner, et al. [13]. Using the fast-marching algorithm, this time in 3D, we generated distance transforms from the boundaries of these segmentations for input to the method. Fig. 4 shows the particle system distributed across two of the shapes after optimizing 300 particles per shape. We used the modes from the resulting Gaussian model to construct a set of surface reconstructions for the three largest principle components. These modes are illustrated in Fig. 5 to three standard deviations, with percentage of total variation of 38.78%, 26.31%, and 12.29% for the modes, respectively. The surface meshes shown were generated by the tight cocone algorithm for sur-

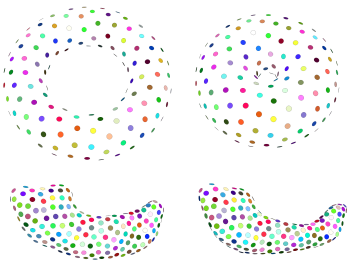

Fig. 4. Particle correspondences in two tori (left) and two hippocampus (right) shapes. Corresponding inter-shape particles have matching colors.

face reconstruction from point clouds [14], using the implementation provided by the authors of that work.

Because the proposed method is completely generalizable to higher dimensions, we were able to perform both the 2D and 3D experiments using the same *C++* software implementation, which is templated by dimension. All experiments were run on a 2Ghz processor with run times of approximately 20 minutes for 2D cases and 45 minutes the 3D cases. In each case, the numerical parameter $\sigma_{\min}$ was set to machine precision and $\sigma_{\max}$ was set to the size of the domain. For the annealing parameter α, we used a starting value roughly equal to the diameter of an average shape and reduced it to machine precision over several hundred iterations. The results presented in this section are typical of reliably similar results produced during several experimental runs, suggesting that the proposed method is fairly robust to the initialization.

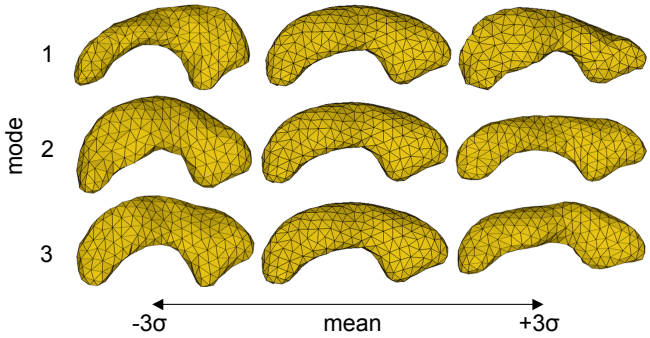

Fig. 5. The mean and three standard deviations of the top three modes of the hippocampus model.

The proposed nonparametric method for shape correspondences produces results that compare favorably with the state of the art. The method works directly on volumes, requires very little parameter tuning, and generalizes easily. The particle-based shape correspondence offers a practical solution for a wider range of shape analysis problems relative to the work in the literature. Although the energy functions push the system toward consistent configurations of particles, the approach does not guarantee diffeomorphic mappings between shapes. The system might be extended to include neighborhood configurations as a part of the model, but this remains a topic of further investigation.

Acknowledgments.

The authors wish to thank Hans Henrik Thodberg for the use of his open source MDL Matlab code.

This work was funded by the Center for Integrative Biomedical Computing, National Institutes of Health (NIH) NCRR Project 2-P41-RR12553-07.

This work is also part of the National Alliance for Medical Image Computing (NAMIC), funded by the NIH through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from `http://nihroadmap.nih.gov/bioinformatics`.

References

1. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J.: A minimum description length approach to statistical shape modeling. IEEE Trans. Med. Imaging **21**(5) (2002) 525–537
2. Audette, M., Ferrie, F., Peters, T.: An algorithmic overview of surface registration techniques for medical imaging. Medical Image Analysis **4** (2000) 201–217
3. Kotcheff, A., Taylor, C.: Automatic Construction of Eigenshape Models by Direct Optimization. Medical Image Analysis **2** (1998) 303–314
4. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J.: 3d statistical shape models using direct optimisation of description length. In: ECCV (3). (2002) 3–20
5. Brechbuhler, C., Gerig, G., Kubler, O.: Surface parameterization and shape description. Visualisation in Biomedical Computing (1992) 80–89
6. Thodberg, H.H.: Minimum description length shape and appearance models. In: IPMI. (2003) 51–62
7. Styner, M., Rajamani, K.T., Nolte, L.P., Zsemlye, G., Székely, G., Taylor, C.J., Davies, R.H.: Evaluation of 3d correspondence methods for model building. In: IPMI. (2003) 63–75
8. Cover, T., Thomas, J.: Elements of Information Theory. Wiley and Sons (1991)
9. Boissonnat, J.D., Oudot, S.: Provably good sampling and meshing of surfaces. Graphical Models **67** (2005) 405–451
10. Meyer, M.D., Georgel, P., Whitaker, R.T.: Robust particle systems for curvature dependent sampling of implicit surfaces. In: In Proceedings of the International Conference on Shape Modeling and Applications. (2005) 124–133
11. Goodall, C.: Procrustes methods in the statistical analysis of shape. J. R. Statistical Society B **53** (1991) 285–339
12. Sethian, J.: Level Set Methods and Fast Marching Methods. Cambridge University Press (1996)
13. Styner, M., Lieberman, J., Gerig, G.: Boundary and medial shape analysis of the hippocampus in schizophrenia. In: MICCAI. (2003) 464–471
14. Dey, T.K., Goswami, S.: Tight cocone: a water-tight surface reconstructor. In: SM '03: Proceedings of the eighth ACM symposium on Solid modeling and applications, New York, NY, USA, ACM Press (2003) 127–134

A New Closed-Form Information Metric for Shape Analysis

Adrian Peter[1] and Anand Rangarajan[2]

[1]Dept. of ECE, [2]Dept. of CISE, University of Florida, Gainesville, FL

Abstract. Recently, a unifying framework was introduced for shape matching that uses mixture-models and the Fisher-Rao metric to couple both the shape representation and deformation. A fundamental drawback of the Fisher-Rao metric is that it is NOT available in closed-form for the mixture models making shape comparisons computationally very expensive. Here, we propose a new Riemannian metric based on generalized ϕ- entropy measures. In sharp contrast to the Fisher-Rao metric, our new metric is available in closed-form.

1 Introduction

Shape analysis is a key ingredient to many medical imaging applications that seek to study the intimate relationship between the form and function of medical and biological structures. In recent work [1], it was shown that shape representation and deformation can be unified within a probabilistic framework – using a mixture of densities to represent landmark shapes and then directly using the Fisher information matrix to establish an intrinsic, Riemannian metric and subsequently a shape distance measure.

To address many of the computational inefficiencies that arise when using the standard Fisher-Rao information metric, the current work introduces a new Riemannian metric based on the generalized notion of a ϕ-entropy functional.

2 Differential Metrics for Parametric Distributions

The parametric, GMM representation for 2-D shapes is given by

$$p(\mathbf{x}|\theta) = \frac{1}{2\pi\sigma^2 K}\sum_{a=1}^{K}\exp\{-\frac{\|\mathbf{x}-\phi_a\|^2}{2\sigma^2}\} \tag{1}$$

where θ is our set of landmarks, $\phi_a = [\theta^{(2a-1)}, \theta^{(2a)}]^T$, $\mathbf{x} = [x^{(1)}, x^{(2)}]^T \in \mathbb{R}^2$ and equal weight priors are assigned to all components, i.e $\frac{1}{K}$. Basically, a shape with K landmarks is represented as a K-component GMM where the landmarks are the means of each component.

Burbea and Rao demonstrated that the notion of distances between parametric models can be extended to a large class of generalized metrics [2] using the *ϕ-entropy functional*

$$H_\phi(p) = -\int_\chi \phi(p)d\mathbf{x} \tag{2}$$

where χ is the measurable space (for our purposes $\mathbb{R}^2$), and ϕ is a C^2-convex function defined on $\mathbb{R}_+ \equiv [0,\infty)$. The metric on the parameter space is obtained by finding the Hessian of (2) along a direction in its tangent space. This directly leads to the following differential metric satisfying Riemannian metric properties

$$ds^2_\phi(\theta) = -\Delta_\theta H_\phi(p) = \sum_{i,j=1}^{n} g^\phi_{i,j} d\theta^i d\theta^j \ , \tag{3}$$

where

$$g^{\phi}_{i,j} = \int_{\chi} \phi''(p)(\frac{\partial p}{\partial \theta^i})(\frac{\partial p}{\partial \theta^j})d\mathbf{x}\,. \tag{4}$$

The metric tensor in (4) is called the ϕ-*entropy matrix*. We can define distances between probability densities by finding a geodesic between their parameter values as determined by (3).

The new metric can be obtained by selecting a ϕ such that ϕ'' becomes a constant in (4). In [3], Havrda and Charvát introduced the notion of a α-order entropy using the convex function

$$\phi(p) = (\alpha - 1)^{-1}(p^{\alpha} - p),\, \alpha \neq 1\,. \tag{5}$$

We set $\alpha = 2$ which results in $\frac{1}{2}\phi'' = 1$. The one-half scaling factor does not impact the metric properties. The new metric is defined as

$$g^{\alpha}_{i,j} = \int_{\chi} (\frac{\partial p}{\partial \theta^i})(\frac{\partial p}{\partial \theta^j})d\mathbf{x} \tag{6}$$

and we refer to it as the α-*order entropy metric tensor*.

The new metric can be used to find the distance between two shapes represented as $p(\mathbf{x}|\theta_1)$ and $p(\mathbf{x}|\theta_2)$. The geodesic can be obtained by solving

$$g^{\alpha}_{ki}\ddot{\theta}^i + \Gamma_{k,ij}\dot{\theta}^i\dot{\theta}^j = 0. \tag{7}$$

Using this new metric, we now have closed-form solutions to $g^{\alpha}_{i,j}$ and $\frac{\partial g^{\alpha}_{kj}}{\partial \theta^i}$.

3 Experimental Results and Analysis

For applications in medical imaging, we have evaluated both metrics and compared to the standard thin-plate spline (TPS) bending energy on real data consisting of nine corpora callosa with 63-landmarks each. (Please see the conference paper for test results and analysis.)

The new ϕ-entropy metric avoids an $O(N^2)$ computational hit on the metric tensor computation and gains another $O(N^2)$ savings when computing the derivative of the metric tensor.

4 Conclusion

In this paper, we introduced a new Riemannian metric for landmark-based shape matching that combines both shape representation and deformation. This improved on the previous work of [1] where the Fisher-Rao metric was used to establish an intrinsic metric between landmark shapes. The new α-order entropy metric enables us to obtain closed-form solutions to the metric tensor and its derivatives. Our new approach also illustrated the possibilities of using information metrics besides Fisher-Rao and their benefits.

References

1. Peter, A., Rangarajan, A.: Shape matching using the Fisher-Rao Riemannian metric: Unifying shape representation and deformation. ISBI 2006 (2006)
2. Burbea, J., Rao, R.: Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. Journal of Multivariate Analysis **12** (1982) 575–596
3. Havrda, M.E., Charvát, F.: Quantification method of classification processes: Concept of structural α-entropy. Kybernetica **3** (1967) 30–35

Statistical linear models in Procrustes shape space

Matias N. Bossa and Salvador Olmos

Communications Technologies Group (GTC), Aragon Institute of Engineering Research (I3A), Zaragoza University, Spain, {bossa,olmos}@unizar.es*

Abstract. The configuration matrix of a set of labeled landmarks is one the most used shape representations. However, it is well-known that the configuration matrix is not invariant under translation, scaling and rotation. This problem is revisited in this work where a local tangent shape space characterization at a reference shape is obtained as the null space of the subspace spanned by the reference shape and the set of translation and rotation generators. This local linear description of the shape space allows us to compute mean and variance of shapes as well as apply classical mutivariate statistical techniques such as Principal Component Analysis. Our proposal is compared with previous approaches, such as the seminal work [1] and more recents works [2] and [3].

1 Introduction

Shape analysis is concerned with the study of the geometrical descriptors that are invariant to position, size and orientation. Shape analysis has proven to be very useful in several tasks of computer vision and medical imaging, such as segmentation of anatomical structures or detecting and quantifying shape differences driven by pathology.

Several shape descriptors have been proposed in the literature. Many authors use a set of landmarks on the shape boundary as relevant geometric features [1, 2]. A recent comprehensive survey on shape analysis with landmarks can be found in [4]. However it is well known that the configuration matrix is not invariant under translation, scaling and rotation.

When using a set of labeled landmarks, one way to achieve a shape description invariant to position, size and orientation is by means of Procrustes alignment. A complete analysis of Procrustes shape space was done in [5, 6], where a deep understanding of the topology of Riemannian manifolds is used.

In this work we propose a procedure to obtain a local shape characterization, by using a tangent space projection and a Riemannian exponential mapping. This shape description allows us to apply classical multivariate statistical tools in shape space with accuracy. For example we applied Principal Component

* This work was funded by grants TEC2005-07801-C03:02 from CICYT, PI04/1795 and PI05/2006 from FIS, and PIP113/05 from DGA. The work of M. Bossa is supported by DGA under the FPI grant B097/2004.

Analysis (PCA) on several characterizations of shape space and compared their performance.

2 Shape characterization and distances

Shape is usually defined as the geometrical information of an object that is invariant under a similarity transformation, i.e. location, translation and scale [7]. In this work we will focus on a particular shape description, that is the set of k labeled points in $\mathbb{R}^m$. The coordinates of these points can be arranged in a $k \times m$ configuration matrix $\mathbf{X}$, or equivalently on a $km \times 1$ configuration vector $\mathbf{x} = \mathrm{vec}(\mathbf{X})$. We will use the notation $\mathbf{z}_i$ to denote the $k \times 1$ vector containing the ith column of any $k \times m$ matrix $\mathbf{Z}$, and $\mathbf{z}^j$ to denote different $km \times 1$ vectors indexed by j.

A configuration matrix $\mathbf{X}$ is not a proper shape descriptor, because it is not pose invariant. For any similarity transformation, i.e. $s \in \mathbb{R}^+$, $\mathbf{R} \in \mathbf{SO}(m)$ (the special orthogonal group) and $\mathbf{t} \in \mathbb{R}^m$, the configuration given by $s\mathbf{X}\mathbf{R} + \mathbf{1}_k\mathbf{t}^{\mathrm{T}}$ describes the same shape than $\mathbf{X}$, where $\mathbf{1}_k$ is the $k \times 1$ vector $[1\ 1\ \ldots\ 1]^{\mathrm{T}}$.

In order to get a shape descriptor invariant under scale and translation several constraints on $\mathbf{X}$ can be used. For example, if $\mathbf{X}$ is forced to be a unity norm matrix $\|\mathbf{X}\|^2 = \mathrm{tr}(\mathbf{X}\mathbf{X}^{\mathrm{T}}) = 1$ (or equivalently $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$), and to have null centroid $\mathbf{1}_k^{\mathrm{T}}\mathbf{X} = \mathbf{0}_m^{\mathrm{T}}$, then scaling and translation effects are removed from $\mathbf{X}$.[1] As the set of unitary vectors $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$ corresponds to the unit sphere $\mathbb{S}^{mk-1}$, the condition of having null centroid is equivalent to intersect $\mathbb{S}^{mk-1}$ with the null space of the matrix $[(\mathbf{1}_k^{\mathrm{T}}\ \mathbf{0}_k^{\mathrm{T}}\ \ldots\ \mathbf{0}_k^{\mathrm{T}})^{\mathrm{T}}, (\mathbf{0}_k^{\mathrm{T}}\ \mathbf{1}_k^{\mathrm{T}}\ \ldots\ \mathbf{0}_k^{\mathrm{T}})^{\mathrm{T}}, \ldots, (\mathbf{0}_k^{\mathrm{T}}\ \mathbf{0}_k^{\mathrm{T}}\ \ldots\ \mathbf{1}_k^{\mathrm{T}})^{\mathrm{T}}]$. The intersection corresponds again to a unit sphere, $\mathbb{S}^{m(k-1)-1}$, named pre-shape space S_m^k.

The shape space Σ_m^k is the set of equivalence classes of S_m^k under the action of $\mathbf{SO}(m)$. The mapping that takes a configuration vector $\mathbf{x} \in S_m^k$ to shape space is $x = \pi(\mathbf{x}) : S_m^k \to \Sigma_m^k$. All the elements in the equivalence class of $\mathbf{x}$, also called fiber, are the set $\{\pi^{-1}(\pi(\mathbf{x}))\}$.

The distance between shapes can be defined in several ways, depending on the problem at hand. The Riemannian distance in pre-shape space $\rho(\mathbf{x}, \mathbf{y})$ is the length of the shortest curve segment in $\mathbb{S}^{m(k-1)-1}$ that connects $\mathbf{x}$ and $\mathbf{y}$ (great-circle): $\rho(\mathbf{x}, \mathbf{y}) = 2\arcsin(\frac{1}{2}\|\mathbf{x} - \mathbf{y}\|)$. However, in this work we are interested in the Procrustes distance, i.e. the distance in shape space, defined as $d(\pi(\mathbf{x}), \pi(\mathbf{y})) = \inf_{\mathbf{R} \in \mathbf{SO}(m)} \rho(\mathbf{x}, \mathbf{R}\mathbf{y})$, with $\mathbf{x}, \mathbf{y} \in S_m^k$.

2.1 Local tangent parametrisation of shape space

Given a reference configuration vector $\boldsymbol{\mu} \in S_m^k$ and $\pi(\boldsymbol{\mu}) \in \Sigma_m^k$ its corresponding shape, shapes in a finite neighbourhood of $\pi(\boldsymbol{\mu})$ can be characterized by variations in S_m^k from $\boldsymbol{\mu}$.

[1] Another way to remove translation can be obtained by multiplying $\mathbf{X}$ by the $(k-1) \times k$ Helmert sub-matrix [2].

The tangent space at $\boldsymbol{\mu}$ of pre-shape space, $T_{\boldsymbol{\mu}}S_m^k$, is a linear space that can be used as a linear approximation, as well as a parametrization of S_m^k, via the exponential-logarithm mapping. We will see that a linear subspace of $T_{\boldsymbol{\mu}}S_m^k$ can be found that is also the tangent space at $\pi(\boldsymbol{\mu})$ of Σ_m^k.

Any vector $\mathbf{x} \in S_m^k$ close to the reference shape $\boldsymbol{\mu} \in S_m^k$, can be written as

$$\mathbf{x} = \cos(\theta)\boldsymbol{\mu} + \sin(\theta)\mathbf{v} \tag{1}$$

with $\mathbf{x}^{\mathrm{T}}\mathbf{x} = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} = \mathbf{v}^{\mathrm{T}}\mathbf{v} = 1$, and $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{v} = 0$. As $\boldsymbol{\mu}$ and $\mathbf{x}$ are centered, $\mathbf{x}^{\mathrm{T}}\mathbf{t}^i = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{t}^i = 0$, where the pure translation vectors $\{\mathbf{t}^i\}_{i=1}^m$ are given by $\mathbf{t}_i^i = \mathbf{1}_k$ and $\mathbf{t}_j^i = \mathbf{0}_k$ for $i \neq j$. It can be checked that also $\mathbf{v}^{\mathrm{T}}\mathbf{t}^i = 0$.

The tangent space at $\boldsymbol{\mu}$, $T_{\boldsymbol{\mu}}S_m^k$, is formed by the set of vectors $\mathbf{u} = \theta\mathbf{v}$. In order to extend classical statistics on the shape space a mapping between the manifold and its tangent space that preserves distances and angles is required. The exponential map

$$\exp_{\boldsymbol{\mu}}(\mathbf{u}) = \cos(u)\boldsymbol{\mu} + \sin(u)\hat{\mathbf{u}} \tag{2}$$

with $u = \|\mathbf{u}\|$ and $\hat{\mathbf{u}} = \mathbf{u}/u$, generates geodesics with initial velocity $\mathbf{u}$, and accomplishes that u is the Riemannian distance between $\boldsymbol{\mu} = \exp_{\boldsymbol{\mu}}(\mathbf{0})$ and $\exp_{\boldsymbol{\mu}}(\mathbf{u})$, i.e. $\rho(\exp_{\boldsymbol{\mu}}(\mathbf{0}), \exp_{\boldsymbol{\mu}}(\mathbf{u})) = u$. The angle between geodesics is the angle between their corresponding initial velocities. The logarithm map is the inverse of the exponential map: $\log_{\boldsymbol{\mu}}(\mathbf{x}) = \mathbf{u}$, where $\hat{\mathbf{u}} = \mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\mu}^T\mathbf{x})/\|\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\mu}^T\mathbf{x})\|$ and $\|\mathbf{u}\| = 2\arcsin(1/2\|\mathbf{x} - \boldsymbol{\mu}\|)$, therefore:

$$\rho(\boldsymbol{\mu}, \mathbf{x}) = \|\log_{\boldsymbol{\mu}}(\mathbf{x})\| \tag{3}$$

In order to get rid of rotations, the variations from $\boldsymbol{\mu}$ should be orthogonal to the fibers $\pi^{-1}(\boldsymbol{\mu})$. Let $\boldsymbol{\Omega}^{i,j}(\theta) \in \mathbf{SO}(m)$, $i < j \leq m$, be a rotation matrix in the plane $\{i, j\}$ with rotation angle θ. It can be seen that the fiber $\mathbf{R}^{i,j}(\theta)\boldsymbol{\mu}$, with $\mathbf{R}^{i,j}(\theta) = (\mathbf{I}_k \otimes \boldsymbol{\Omega}^{i,j}(\theta))$, is a curve in S_m^k but not a geodesic for $m > 2$. It was shown in [5, 8, 9] that if one moves away from $\boldsymbol{\mu}$ along geodesics orthogonally to fibers (no matters fibers are not geodesics), a Riemannian submersion from S_m^k onto Σ_m^k is obtained. Roughly, this means that identifying parts of Σ_m^k with submanifolds of S_m^k orthogonal to fibers, is a smooth mapping with the same metric. This allows us to work in S_m^k as if it were Σ_m^k, in a neighbourhood of $\boldsymbol{\mu}$. The tangent directions to fibers $\mathbf{R}^{i,j}(\theta)\boldsymbol{\mu}$ at $\boldsymbol{\mu}$ are given by their derivatives $\mathbf{r}^{i,j}$: with $\mathbf{r}_i^{i,j} = \boldsymbol{\mu}_j$, $\mathbf{r}_j^{i,j} = -\boldsymbol{\mu}_i$, and $\mathbf{r}_l^{i,j} = \mathbf{0}_k$ for $i, j \neq l$. The same result can be found in [6], from a more theoretical point of view. Any configuration $\mathbf{x}$ Procrustes aligned to $\boldsymbol{\mu}$, fulfils the following linear constraints: $\mathbf{x}^{\mathrm{T}}\mathbf{r}^{i,j} = 0$. This is also true for $\boldsymbol{\mu}$, then $\mathbf{v}^{\mathrm{T}}\mathbf{r}^{i,j} = 0$.

Fig. 1 shows schematically the pre-shape space S_m^k as a 2-sphere. The fiber $\mathbf{R}^{i,j}(\theta)\boldsymbol{\mu}$ is a small circle labeled as $\mathbf{R}\boldsymbol{\mu}$, and its tangent direction as $\mathbf{r}$. The shape space Σ_m^k is obtained by intersecting the pre-shape sphere with a plane orthogonal to $\mathbf{r}$.

Summing up all the constraints on $\mathbf{v}$ can be written as $\mathbf{v}^{\mathrm{T}}\mathbf{N} = \mathbf{0}_{km}$, with $\mathbf{N} = (\boldsymbol{\mu}, \mathbf{t}^1, \mathbf{t}^2, \ldots, \mathbf{t}^m, \mathbf{r}^{1,2}, \mathbf{r}^{1,3}, \ldots, \mathbf{r}^{(m-1),m})$. The intersection of the pre-shape

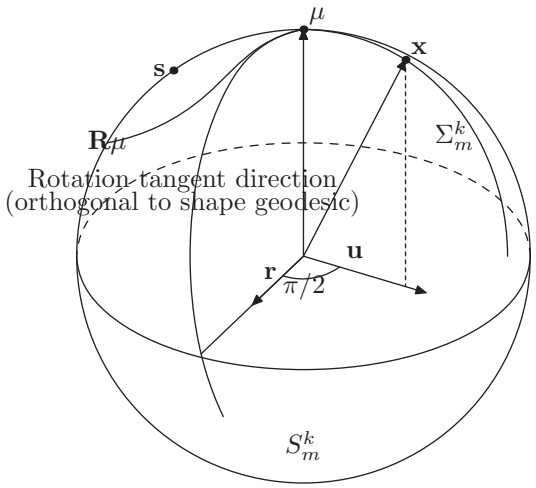

Fig. 1. Schematic representation of pre-shape space S_m^k. Shape space Σ_m^k is identified with the submanifold orthogonal to fiber $\mathbf{R}$ in a neighbourhood of reference shape .

sphere $\mathbb{S}^{m(k-1)-1}$ and the hyperplane defined by the null space of $\{\mathbf{r}^{i,j}\}$ is again a smaller dimension sphere, in this case $\mathbb{S}^{m(k-1-(m-1)/2)-1}$. Therefore, shape space, although being a manifold with singularities for $m > 2$, it behaves like a sphere in a finite neighbourhood of any nonsingular reference shape $\boldsymbol{\mu}$, with the metric defined in (3), i.e. $d(\pi(\mathbf{x}), \pi(\mathbf{y})) = \rho(\mathbf{x}, \mathbf{y})$.

We have explicitly removed the m degrees of freedom of translation, the $m(m-1)/2$ ones of rotation and the one of size. If $\mathbf{M}$ denotes the matrix containing a basis of the null space of $\mathbf{N}$, any shape can be written as

$$\mathbf{x} = \cos(b)\boldsymbol{\mu} + \sin(b)\mathbf{M}\hat{\mathbf{b}}, \tag{4}$$

with $\hat{\mathbf{b}} = \mathbf{b}/b$, $b = \|\mathbf{b}\|$ and $\mathbf{b}$ an arbitrary vector of $\mathbb{R}^{m(k-1-(m-1)/2)-1}$. The parameter vector $\mathbf{b}$ is a linear, non-redundant representation of shape space, that preserve distances to the reference shape $\boldsymbol{\mu}$ and angles between geodesics. Therefore it can be used as an approximate and nice characterization of the shape space. Note that $\mathbf{M}$ only depends on $\boldsymbol{\mu}$, and it is not required a training set to compute $\mathbf{M}$.

The above conditions are valid at least within a ball around $\boldsymbol{\mu}$ with radius smaller than the distance from $\boldsymbol{\mu}$ to the closest singularity, which is usually far away enough in real problems. The singularities (labeled as $\mathbf{s}$ in Fig. 1) arise when $\text{rank}(\mathbf{X}) < m-1$, . For $m = 3$, this means that all the landmarks are located over a line, and for $m = 2$ there are no singularities. Fig. 2 shows a simple example of a geodesic passing through a singularity. It can be seen on bottom left panel that shapes further away than a singularity accomplish $d(\pi(\mathbf{x}), \pi(\mathbf{y})) \neq \rho(\mathbf{x}, \mathbf{y})$.

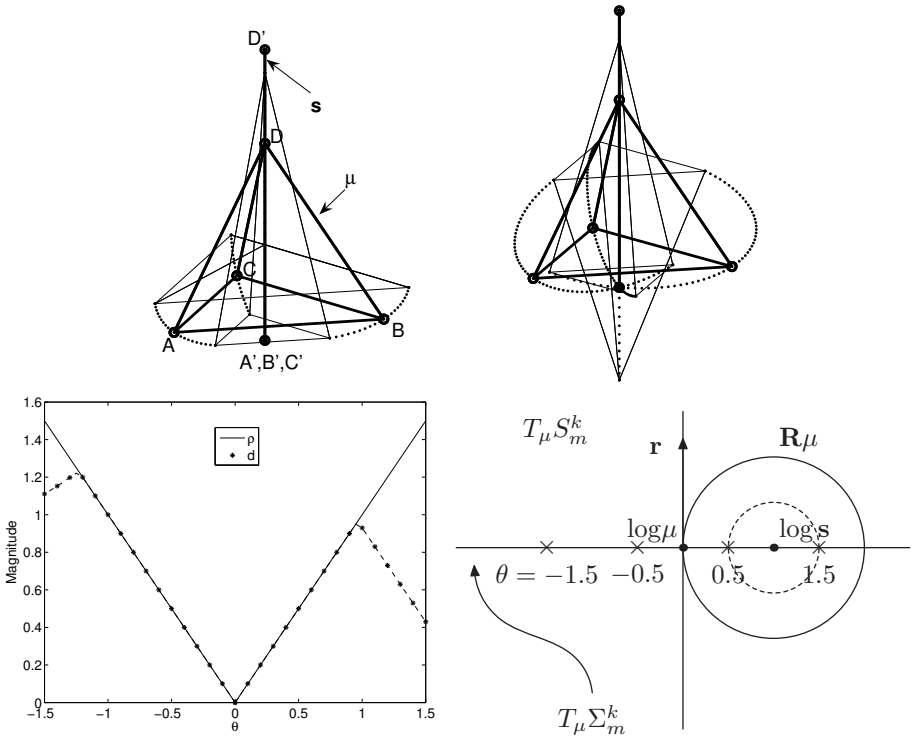

Fig. 2. Up: The reference shape (a tetrahedron) is denoted as and its closest singular shape as **s** (vertical line). The Procrustes distance between and **s** is about 1. The vertexes ABC collapse at A'B'C'. Geodesics from towards **s** are illustrated with dotted lines. On the left upper panel, θ runs in the interval (-0.5,0.5), and on the right (-1.5,1.5). Note that the configuration obtained for $\theta = 1.5$ corresponds to an inverted tetrahedron, which is not Procrustes aligned to . Bottom left: Values of the pre-shape Riemannian distance ρ (solid line) and Procrustes distance d (dashed line) from for running values of θ. Bottom right: schematic view of the tangent space T S_m^k.

3 Multivariate statistics on shape space

Classical multivariate statistical techniques such as PCA have been applied to some linear approximations of shape space. PCA involves estimating the mean and the modes of variation from a training set.

The mean $\boldsymbol{\mu}_E$ of a set of vectors $\mathbf{x}^n \in S_m^k$ is computed as an iterative process including averaging and alignment. At each iteration, classical averaging is performed on landmark coordinates and the vectors are aligned to the normalized average. The final result $\boldsymbol{\mu}_E$ is the normalized average after convergence. $\boldsymbol{\mu}_E$ provides the shape that minimizes the Euclidean distance $\sum_n \|\boldsymbol{\mu} - \mathbf{x}^n\|^2$, but not the Procrustes distance $\sum_n d^2(\boldsymbol{\mu}, \mathbf{x}^n)$ that we are interested in.

An alternative procedure to compute the mean of a set of elements in Riemmanian manifolds was proposed in [10]. The method consists on iteratively estimating the mean as $\boldsymbol{\mu}^{j+1} = \exp(1/n\Sigma_n \log_{\boldsymbol{\mu}^j}(\mathbf{x}^n))$. We used this method with $\boldsymbol{\mu}_E$ as initialization. This iterative procedure has a fast convergence.

Modes of variation are computed in the tangent space at $\boldsymbol{\mu}$, by means of SVD of the residues $\mathbf{u}^n = \log_{\boldsymbol{\mu}}(\mathbf{x}^n)$. The principal geodesic directions are given by the singular vectors $\mathbf{p}^k$. Singular vectors are collected in the matrix $\mathbf{P}$. The singular values σ^k account for their relative importance. The procedure of computing PCA in the tangent space of a Riemannian manifold is called Principal Geodesic Analysis (PGA) [10, 11]. Principal geodesics are $\mathbf{y}^k(b_k) = \exp_{\boldsymbol{\mu}}(\mathbf{p}^k b_k) = \cos(b_k)\boldsymbol{\mu} + \sin(b_k)\mathbf{p}^k$, where b_k measures the deviation from the mean. An arbitrary vector of the PGA model is:

$$\mathbf{y}(\mathbf{b}) = \exp_{\boldsymbol{\mu}}(\mathbf{P}\mathbf{b}) = \cos(b)\boldsymbol{\mu} + \sin(b)\mathbf{P}\hat{\mathbf{b}}, \tag{5}$$

where $\hat{\mathbf{b}} = \mathbf{b}/b$, and $b = \|\mathbf{b}\|$.

3.1 Discussion of previous approaches

Previous approaches proposed for PCA on shape space failed on finding the best linear approximation of the shape space. Cootes' proposal [1] is one of the most referenced approaches. Shape was modeled as

$$\mathbf{y}_C(\mathbf{b}) = \boldsymbol{\mu} + \mathbf{P}_C\mathbf{b}, \tag{6}$$

where $\mathbf{P}_C$ is the matrix with the principal components of the residues $\tilde{\mathbf{x}}_C^n = \mathbf{x}^n - \boldsymbol{\mu}$, and $\mathbf{b}$ is a parameter vector. However, the spherical topology of the shape space was disregarded. Shape model in (6), labeled as PCA_C in this work, has two drawbacks: firstly, shape instances $\mathbf{y}_C$ don't have unity norm. Secondly, the shape Procrustes distance between an instance and $\boldsymbol{\mu}$ is neither proportional to $\|\mathbf{b}\|$, nor symmetric under a sign change of $\mathbf{b}$, i.e. $d(\boldsymbol{\mu}, \mathbf{y}_C(\mathbf{b})) \neq d(\boldsymbol{\mu}, \mathbf{y}_C(-\mathbf{b}))$. The reason is that $\|\mathbf{P}_C^{\mathrm{T}}\boldsymbol{\mu}\| \neq 0$ because the residues are not orthogonal to the mean. Accordingly the maximum number of degrees of freedom is $m(k-1-(m-1)/2)$ (i.e. one dimension larger than actual Procrustes shape space). Splitting $\mathbf{b}$ into two orthogonal parts, $\mathbf{b} = \mathbf{b}_{\boldsymbol{\mu}} + \mathbf{b}_{\perp\boldsymbol{\mu}}$, such that $\boldsymbol{\mu}^{\mathrm{T}}(\mathbf{P}_C\mathbf{b}_{\perp\boldsymbol{\mu}}) = 0$, the asymmetry in norm can be shown as

$$\begin{aligned}\|\mathbf{y}_C(\mathbf{b})\| = \|\boldsymbol{\mu}(1+b_{\boldsymbol{\mu}}) + \mathbf{P}_C\mathbf{b}_{\perp\boldsymbol{\mu}}\| = \sqrt{(1+b_{\boldsymbol{\mu}})^2 + b_{\perp\boldsymbol{\mu}}^2} \neq \\ \neq \sqrt{(1-b_{\boldsymbol{\mu}})^2 + b_{\perp\boldsymbol{\mu}}^2} = \|\boldsymbol{\mu}(1-b_{\boldsymbol{\mu}}) - \mathbf{P}_C\mathbf{b}_{\perp\boldsymbol{\mu}}\| = \|\mathbf{y}_C(-\mathbf{b})\|,\end{aligned} \tag{7}$$

with $b_{\boldsymbol{\mu}} = \|\mathbf{b}_{\boldsymbol{\mu}}\|$ and $b_{\perp\boldsymbol{\mu}} = \|\mathbf{b}_{\perp\boldsymbol{\mu}}\|$. Similarly the Procrustes distance accomplishes

$$\begin{aligned}d(\boldsymbol{\mu}, \mathbf{y}_C(\mathbf{b})) = 2\arcsin\left(\frac{1}{2}\left\|\boldsymbol{\mu}\left(1 - \frac{1+b_{\boldsymbol{\mu}}}{\|\mathbf{y}_C(\mathbf{b})\|}\right) - \frac{\mathbf{P}_C\mathbf{b}_{\perp\boldsymbol{\mu}}}{\|\mathbf{y}_C(\mathbf{b})\|}\right\|\right) \neq \\ \neq 2\arcsin\left(\frac{1}{2}\left\|\boldsymbol{\mu}\left(1 - \frac{1-b_{\boldsymbol{\mu}}}{\|\mathbf{y}_C(-\mathbf{b})\|}\right) + \frac{\mathbf{P}_C\mathbf{b}_{\perp\boldsymbol{\mu}}}{\|\mathbf{y}_C(-\mathbf{b})\|}\right\|\right) = d(\boldsymbol{\mu}, \mathbf{y}_C(-\mathbf{b})).\end{aligned} \tag{8}$$

Later works, even realizing the spherical topology of shape space, made use of a linear projection instead of using the log mapping to project data on the tangent space.Therefore a distortion in the metric was introduced.

In [3] data is projected into tangent space by scaling $\mathbf{x}^n$ in such a way that the residue is perpendicular to $\boldsymbol{\mu}$, $\tilde{\mathbf{x}}_A^n = \mathbf{x}^n(\mathbf{x}^{n\mathrm{T}}\boldsymbol{\mu})^{-1} - \boldsymbol{\mu}$. In [2] the tangent coordinates were defined as the linear projection on the plane perpendicular to the mean, $\tilde{\mathbf{x}}_B^n = \mathbf{x}^n - \boldsymbol{\mu}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}^n)$. The approximation done is $\theta \approx \tan\theta$ in the first case, while in the second is $\theta \approx \sin\theta$, being θ the Procrustes distance from $\boldsymbol{\mu}$ to $\mathbf{x}$. The corresponding shape models are

$$\mathbf{y}_A(\mathbf{b}) = \frac{\boldsymbol{\mu}}{\sqrt{1+b^2}} + \frac{\mathbf{P}_A\mathbf{b}}{\sqrt{1+b^2}} \tag{9}$$

$$\mathbf{y}_B(\mathbf{b}) = \boldsymbol{\mu}\sqrt{1-b^2} + \mathbf{P}_B\mathbf{b} \tag{10}$$

labeled as PCA_A and PCA_B respectively.

The space spanned by the principal vector is one dimension smaller than $\mathbf{P}_C$ as $\mathbf{P}_{\{A,B\}}^{\mathrm{T}}\boldsymbol{\mu} = \mathbf{0}$ and so $b_{\boldsymbol{\mu}} = 0$. Therefore, unlike the model in (6), these approaches are symmetric under a change of sign of the parameter $\mathbf{b}$. The Procrustes distance from the mean is

$$d(\boldsymbol{\mu}, \mathbf{y}_A(\mathbf{b})) = \| \arctan(b) \| = d(\boldsymbol{\mu}, \mathbf{y}_A(-\mathbf{b})) \tag{11}$$

$$d(\boldsymbol{\mu}, \mathbf{y}_B(\mathbf{b})) = \| \arcsin(b) \| = d(\boldsymbol{\mu}, \mathbf{y}_B(-\mathbf{b})). \tag{12}$$

Left panel of Fig. 3 illustrates the Procrustes distance from the mean shape versus signed b. A value of $b_{\boldsymbol{\mu}}/b$ of 0.5 was used in the PCA_C model. PGA is the only model for which a linear change of the model parameters implies a linear variation of the Procrustes distance. On one hand PCA_A model generates instances closer to the mean than PGA, especially for large values of b, which means that PCA_A is more sensitive to outliers. PCA_C model has a similar behaviour for $b_{\boldsymbol{\mu}}/b$ small. On the other hand, PCA_B generates instances further away of the mean than PGA in the valid running interval $b \in [-1, 1]$.

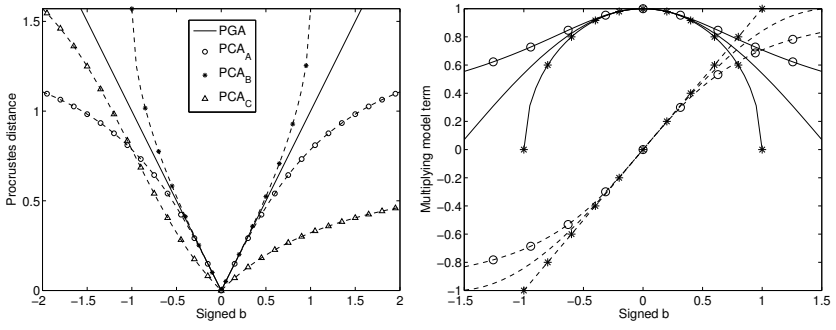

Fig. 3. Left: Procrustes distance versus signed b. Right: Model scaling terms versus b. Solid lines for mean terms and dashed lines for modes of variation terms.

There is a similarity between PGA, PCA_A and PCA_B model equations (5), (9) and (10). Mean and variation modes are orthogonal vectors multiplied by nonlinear scalar functions. The corresponding Taylor expansions of the mode of variation term differ at the 3rd order on b, being $b - 1/6b^3$ for PGA, $b - 1/2b^3$ for PCA_A and b for PCA_B. The model PCA_A underestimates the variation term compared to PGA, while the model PCA_B overestimates it, which is in agreement with the behaviour explained before on the Procrustes distance. The Taylor expansions of the mean term are different at the 4th order: $1 - 1/2x^2 + 1/24x^4$ for PGA, $1 - 1/2x^2 + 3/8x^4$ for PCA_A and $1 - 1/2x^2 - 1/8x^4$ for PGA_B. Right panel in Fig. 3 illustrates both terms for these models.

The Procrustes distance is always smaller than or equal to $\pi/2$, but it is much smaller in most of the real cases. Differences between PGA, PCA_A and PGA_B are then very small for this range of distances.

4 Illustrative example

In order to provide a simple illustrative example of the differences between previous approaches, 100 instances were randomly generated by adding noise to a reference shape (a cube in this simulation with $k = 8$). The standard deviation of the noise was set to 30% of the edge length.

Statistical shape models considered before were computed from the simulated instances. Mean shape $\boldsymbol{\mu}$ as well as the first two variation modes are illustrated in Fig. 4. There were no visible differences in the mean shape, but significant differences were found in the variation modes. The second mode of variation of PCA_C (upper right panel in Fig. 4) has a very different direction compared to PGA because this mode includes a significant part of the mean shape, introducing a relevant radial component. In contrast, PGA and $PCA_{\{A,B\}}$ do not suffer this drawback.

It can be seen in Fig. 4 that the length of PCA_A trajectories as well as PCA_C after normalization , are shorter than PGA, while PCA_B trajectories are longer. This is in agreement with the results shown in the left panel of Fig. 3. It is worthy to note that a large value of b was required to obtain visible differences.

The Procrustes distance between instances and mean shape $\boldsymbol{\mu}$ were computed for several values of the model parameters and are shown in Fig. 5. The profile of the curves in Fig. 5 resembles the theoretical ones in Fig. 3. The Procrustes distance in the second mode of variation of PCA_C is highly asymmetric due to the presence of the mean shape in that mode.

PGA is the only approach where a linear change of the model parameters involves a linear variation of the Procrustes distance from the mean. This property is really crucial for any statistical analysis performed on the linear space of the parameters, such as hypothesis testing, clustering, classification.

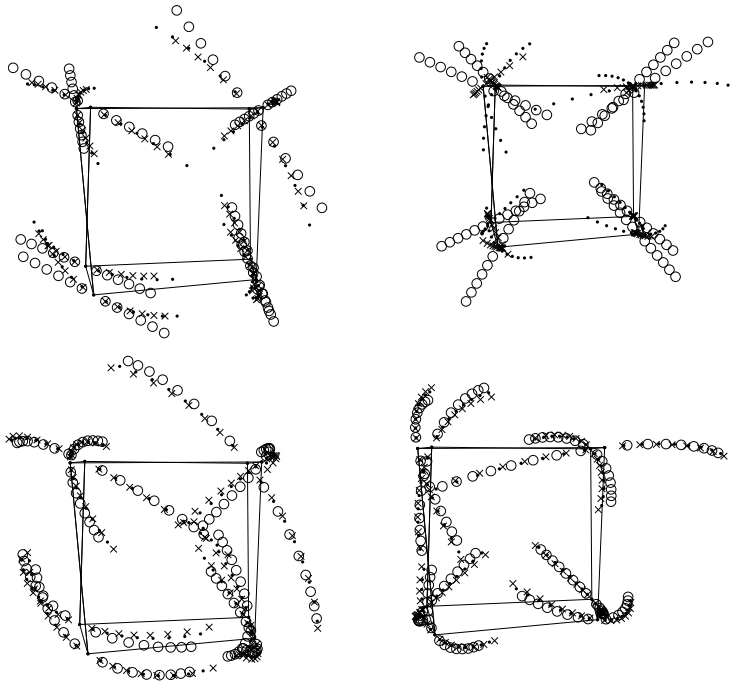

Fig. 4. Mean shape (solid line) and the first/second (left/right), modes of variation. PGA model is illustrated with dotted trajectories of geodesics. Top: Model parameter $b_{\{1,2\}}$ runs in the interval $(-3\sigma_{\{1,2\}}, 3\sigma_{\{1,2\}})$. Circles denote PCA_C straight line trajectories and crosses their corresponding normalized instances. Bottom: Model parameter $b_{\{1,2\}}$ runs in the interval $(-4\sigma_{\{1,2\}}, 4\sigma_{\{1,2\}})$. Circles denote PCA_A and crosses PCA_B trajectories.

5 Conclusions

The problem of shape description by means of a set of landmarks was revisited in this work. A local tangent shape characterization was obtained in earlier works by removing the degrees of freedom corresponding to scaling, translation and rotation. The main added value of this work is twofold: firstly, to propose a simple and intuitive way to understand this shape characterization; secondly, to apply classical multivariate statistics, such as PCA, on the shape space. The obtained shape representation is linear, non-redundant and preserves the Procrustes distance between instances and the reference shape, and also preserves angles between geodesics. Some of these properties do not hold in alternative models previously proposed in the literature. Additionally, these benefits are obtained with a minimum extra modelling effort and the formulation of the model resembles the classical ones.

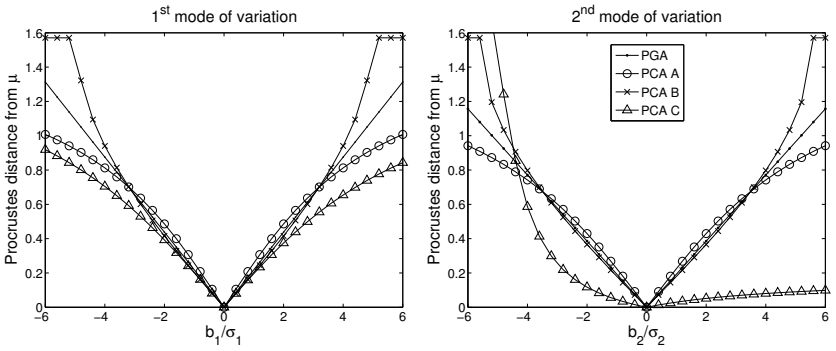

Fig. 5. Procrustes distance between generated instances $\mathbf{y}$ and reference shape .

References

1. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Computer Vision and Image Understanding **61**(1) (1995) 38–59
2. Dryden, I., Mardia, K.: Statistical shape analysis. John Wiley (1998)
3. Fletcher, P.: Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI. PhD thesis, Department of Computer Science, University of North Carolina (2004)
4. Heo, G., Small, C.G.: Form represention and means for landmarks: A survey and comparative study. Computer Vision and Image Understanding **102**(2) (2006) 188–203
5. Le, H., Kendall, D.: The Riemannian Structure of Euclidean Shape Spaces: A Novel Environment for Statistics. Annals of Statistics **21**(3) (1993) 1225–1271
6. Kent, J.T., Mardia, K.V.: Shape, Procrustes tangent projections and bilateral symmetry. Biometrika **88**(2) (2001) 469–485
7. Kendall, D.G.: The diffusion of shape. Adv. in Appl Prob **9** (1977) 428–430
8. Le, H.: On geodesics in euclidean shape space. J. London Math. Soc. **44**(2) (1991) 360–372
9. Le, H.: Unrolling Shape Curves. Journal of the London Mathematical Society **68**(2) (2003) 511–526
10. Pennec, X.: Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements. In Cetin, A., Akarun, L., Ertuzun, A., Gurcan, M., Yardimci, Y., eds.: Proc. of NSIP 1999. Volume 1., June 20-23, Antalya, Turkey, IEEE-EURASIP (1999) 194–198
11. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans. Med. Imaging **23**(8) (2004) 995–1005

Hippocampus-Specific fMRI Group Activation Analysis with Continuous M-Reps

Paul A. Yushkevich[1], John A. Detre[2], Kathy Z. Tang[2], Angela Hoang2, Dawn Mechanic-Hamilton[2], María A. Fernández-Seara[2], Marc Korczykowski[2], Hui Zhang[1], and James C. Gee[1]

[1] Penn Image Computing and Science Laboratory, Department of Radiology, University of Pennsylvania
[2] Center for Functional Neuroimaging, Department of Radiology, University of Pennsylvania

Extended Abstract

Continuous M-Reps. In earlier work [4], we developed a deformable modeling framework called the *continuous medial representation (cm-rep)*, which, in simplest terms, is a continuous analog of the *m-rep* approach by Pizer *et al.* [3]. Cm-reps are deformable templates that have a special property that the skeleton (medial axis) of the template preserves its branching topology (number and configuration of its branches) during deformation. Furthermore, the skeleton of the deforming template is modeled explicitly as a continuous manifold and the exact geometric relationship between the skeleton and the template's boundary is captured. The skeleton-boundary relationship makes it possible to define a special shape-based coordinate system on the interior of the template. One of the coordinate axes connects points on the skeleton to the corresponding (nearest) points on the boundary, while the other two axes parameterize the skeleton as well as the boundary. As argued in [5], this coordinate system is a natural way to parameterize the interiors of anatomical structures to which cm-rep templates have been fitted. It allows us to analyze the distribution of appearance features (intensity, fMRI, DTI) within the structures more or less independently of the differences in the shape of the structures. In this paper, we illustrate the application of such analysis to functional MRI data associated with the hippocampus.

fMRI Study. We designed an fMRI study in which hippocampal activation was highly anticipated. Healthy volunteers (n=18) performed a complex scene encoding task in the scanner. The study followed the block design paradigm. In addition to EPI functional imaging, T_1 anatomical scans were acquired. Standard techniques (SPM, [2]) were used to align functional images to the anatomy, to correct for motion artifacts, to smooth the data temporally and spatially and to compute subject-level statistical maps using the general linear model.

Structure-Specific Group Analysis of fMRI. Whole-brain fMRI group analysis methods such as SPM [2] rely on deformable image registration to normalize

data between subjects. Registration does not always align the hippocampus accurately [1] and such misalignment will, in theory, reduce the statistical significance of intersubject statistical maps. To improve intersubject alignment in the hippocampus, we fit cm-rep templates to manual segmentations of the hippocampus (Fig. 1) and use the cm-rep coordinate system to project subject-level fMRI statistics to a common reference space[3]. We then perform random effects analysis to compute a hippocampus-specific groupwise statistical map.

Results: Improved Sensitivity and Specificity. We compared whole-brain and hippocampus-specific maps generated by SPM and our technique, respectively (Fig. 2). The peaks in the hippocampus-specific map had higher raw t-values and lower corrected p-values than the peaks located near the right hippocampus in the whole-brain map (Table 1). This indicates that our approach is more sensitive and that, in general, the accuracy of intersubject normalization does weigh in on the sensitivity of group analysis, despite the influence of many other sources of error. Furthermore, in most studies, peaks and clusters in the whole-brain activation map are assigned anatomical labels by looking up their coordinates in the Talairach space. Due to normalization errors, peaks resulting from consistent hippocampal activation across subjects may be mislabeled as activation in nearby structures. By restricting its analysis to the hippocampus, our approach aims to increase the likelihood that activation labeled as hippocampal is indeed originating in the hippocampus of each subject. However, specificity is confounded by the fact that spatial smoothing and EPI/T_1 alignment are performed for each subject. In part, our future work will focus on developing structure-focused EPI smoothing algorithms.

References

1. O. T. Carmichael, H. A. Aizenstein, S. W. Davis, J. T. Becker, P. M. Thompson, C. C. Meltzer, and Y. Liu. Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. *Neuroimage*, 27(4):979–90, Oct 2005.
2. R. Frackowiak, K. Friston, C. Frith, R. Dolan, and J. Mazziotta, editors. *Human brain function*. Academic Press USA, 1997.
3. S. M. Pizer, P. T. Fletcher, S. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. G. Gash, J. M. Glotzer, M. R. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. Yushkevich, and E. L. Chaney. Deformable m-reps for 3D medical image segmentation. *International Journal of Computer Vision*, 55(2):85–106, Nov 2003.
4. P. A. Yushkevich, H. Zhang, and J. Gee. Continuous medial representation for anatomical structures. *IEEE Trans Med Imaging*, 2006. to appear.
5. P. A. Yushkevich, H. Zhang, and J. C. Gee. Statistical modeling of shape and appearance using the continuous medial representation. In *Medical Image Computing and Computer-Assisted Intervention, MICCAI*, volume 2, pages 725–732, 2005.

[3] Manual segmentation was performed only for the right hippocampus so our results are currently limited to that structure

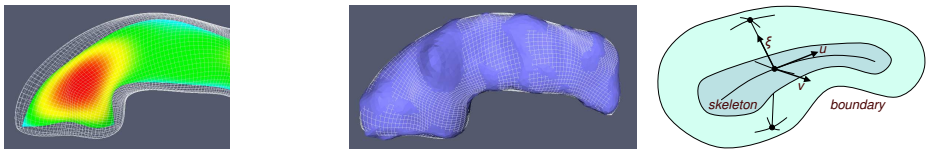

Fig. 1. A cm-rep template fitted to the hippocampus. On the left, the thickness function is plotted in color over the template's skeleton. In the middle, the target segmentation is shown in blue. The shape-based coordinate system is illustrated on the right.

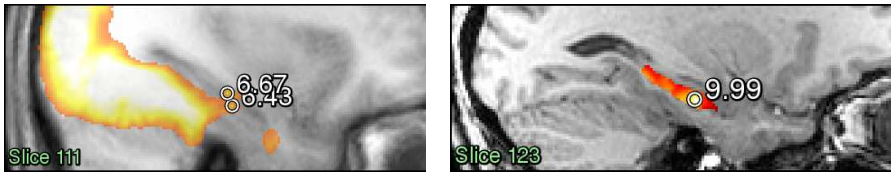

Fig. 2. Examples of sagittal slices through whole-brain (left) and hippocampus-specific (right) group analysis maps. Peaks are identified by circles.

Table 1. Peaks near the right hippocampus in the whole-brain group activation map compared to the peaks in the right hippocampus-specific group analysis map. Abbr.: right cerebrum (RC), parahippocampal gyrus (PHG), right hippocampus (RH). The column P_{hippo} indicates the fraction of hippocampus segmentations, warped into the space of the whole-brain template, that overlap each peak's location.

Peaks in Whole-Brain RFX Map

Tal. Coord.	Talairach Labels	t-Stat	p_{corr}	P_{hippo}
16, -31, -2	RC, limbic lobe, sub-gyral, *, *	7.69	0.033	0
20, -29, 8	RC, sub-lobar, thalamus, gray matter, pulvinar	6.89	0.14	0
31, -22, -9	RC, temporal lobe, sub-gyral, gray matter, hippocampus	6.67	0.21	0.94
33, -6, 27	RC, limbic lobe, uncus, white matter, *	6.47	0.30	0
31, -20, -13	RC, limbic lobe, PHG, gray matter, hippocampus	6.43	0.33	0.82
27, -21, -13	RC, limbic lobe, PHG, gray matter, Brodmann area 6	6.12	0.60	0.94
24, -17, -17	RC, limbic lobe, PHG, white matter, *	6.01	0.74	0.94

Peaks in the Hippocampus-Specific RFX Map

Loc. in RH	t-Stat	p_{corr}	Loc. in RH	t-Stat	p_{corr}
Head, Lateral	9.99	0.00016	Tail, Lateral	8.04	0.0035
Tail, Medial	7.55	0.0082	Body, Lateral	6.45	0.062

An Intrinsic Geometric Framework for Simultaneous Non-Rigid Registration and Segmentation of Surfaces *

Nicholas Lord, Jeffrey Ho, Baba C. Vemuri, Stephan Eisenschenk
University of Florida
Gainesville, FL 32611, USA
{nlord, jho, vemuri}@cise.ufl.edu,
{stephan.eisenschenk}@neurology.ufl.edu

Abstract. In clinical applications where structural asymmetries between homologous shapes have been correlated with pathology, the questions of definition and quantification of 'asymmetry' arise naturally. When not only the degree but the position of deformity is thought relevant, asymmetry localization must also be addressed. Asymmetries between paired shapes can and have been formulated in the literature in terms of (non-rigid) diffeomorphisms between the shapes. For the infinity of such maps possible for a given pair, we define optimality as the minimization of total distortion, where 'distortion' is in turn defined as deviation from isometry. We thus propose a novel variational formulation for segmenting asymmetric regions from surface pairs based on the minimization of a functional of both the deformation map and the segmentation boundary, which controls gradient discontinuity of the map. This minimization is achieved via a quasi-simultaneous evolution of the map and curve. Our formulation is inherently intrinsic and parameterization-independent. We present examples using both synthetic data and pairs of left and right hippocampal structures, hippocampus malformation being linked with such neurological disorders as epilepsy and schizophrenia.

1 Introduction

Many neurological disorders have been clinically correlated with shape abnormalities in specific brain structures. For example, hippocampal shrinkage is associated with epilepsy and schizophrenia, among other conditions. A precise analysis of such abnormalities, which are usually considered to be failures of symmetry of the left and right halves of the structure in question, could unlock the ability not only to track the progression of existent disease, but to identify at-risk individuals for preventative treatment. However, devising a framework in which to conduct the necessary shape comparison is a nontrivial matter. The shortcomings of raw volumetric comparison are obvious. Landmark-based methods

* This research was supported in part by the grant NIH R01-NS046812 to BCV and the UF Alumni Fellowship to NAL. Data sets were partially provided by Dr. Christiana Leonard of the UF McKnight Brain Institute.

are inapplicable to problems involving many anatomical structures (including the hippocampus), because many biological structures do not exhibit the sorts of readily and consistently identifiable local features on which such methods depend. The medial surface representation of Gerig et al. [1][2] represents a more sophisticated take on volumetric analysis that includes some localization of volume disparity, but the economy of the method leads both to noise sensitivity and an acknowledged inability to distinguish between different sorts of deformation (e.g. certain forms of elongation vs. bending). This is due to the expression of complicated behaviors in terms of simple distances which discard some amount of directionality.

More recently, Wang et al.[3][4][5] presented a technique involving the alignment of surfaces based on the formation of a 2D (parametric) diffeomorphic map from Riemannian surface structure information, with this map representing the non-rigid registration between the surfaces. This approach arguably makes use of what surface feature information is there to be had from relatively nondescript shapes, without imposing an ill-defined demand for landmark identification. However, in forming the correspondence by maximizing the mutual information of the chosen surface characteristics over the entirety of both surfaces, these methods ignore key considerations. For one, while the conceptual division of the body into anatomical structures is hardly arbitrary in a general sense, for the purpose of deformation analysis, it may very well be. That is to say, we may wish to identify a selected portion of the hippocampus as structurally abnormal relative to its correspondent, rather than regarding the entire hippocampus as deformed. Furthermore, the MI of the domain and target surface features does not in and of itself restrict the differential properties of the map, only what properties should be found at given locations on the target surface. Therefore two regions of the target surface that exhibit the same Riemannian characteristics cannot be differentiated between, despite the fact that one choice may be physically unreasonable as a map target. Thus this framework demands a separate regularizing constraint on the map evolution.

To address these concerns, we present a novel approach to the non-rigid registration of anatomically correspondent surfaces which rests on the fundamental assumptions that (a) the relative deformity of anatomical structures for which symmetry is expected can be intuitively and precisely quantified as the deviation from isometry of the deformation map between their surfaces, and (b) since deformities may well be locally confined to certain regions of the given surfaces, the evolution of the global correspondence must allow for a partial disconnect between normal and abnormal regions, as these are by definition not expected to exhibit the same deformation patterns. As such, surface segmentation rests at the heart of our approach, which can be viewed not only as a registration tool but also as an asymmetry localizer and quantifier. Our definition of system energy in terms of isometry inherently balances the consideration that the map should accurately match Riemannian surface characteristics with the idea that the map should incur as little deformation as possible in doing so. This gives way to a simple, elegant formulation relatively free of heuristics. Since we here

restrict our consideration to surfaces that are topologically cylindrical (appropriate for hippocampi among many other structures, once we exclude a pair of "poles" from the topology), we are able to parameterize with a single patch (periodic at a pair of ends) for each surface in the pair, and as the formulation is intrinsic, we can confine evolution to these 2D parametric domains, retaining elegance. The evolution is driven quasi-simultaneously by the energies of both the segmenting curve and the deformation field, in the spirit of the works of Yezzi et al. and Unal et al.[6][7], where the former group restricted its work to rigid image registration and the latter generalized the idea to the non-rigid deformation case. In [8], Vemuri et al. presented a registration-assisted image segmentation technique. Of the many proposals in the literature which followed, however, none addressed the issue of joint manifold segmentation and registration, which is our focus. Our work differs notably in that we evolve the segmenting curve and the registration map on a general 2D manifold (rather than a flat image plane), and rather than registering and segmenting image data, we register and segment the shape characteristics of the very surface on which the processes are operating.

The rest of the paper is organized as follows: Section 2 contains the mathematical formulation and numerical algorithm used to solve it. Section 3 contains experimental results on both synthetic and real data sets. In section 4, we draw conclusions.

2 Simultaneous Segmentation and Registration Algorithm

Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be two surfaces in $\mathbb{R}^3$. The Euclidean metric in $\mathbb{R}^3$ induces Riemannian metrics g_1 and g_2 on $\mathbf{S}_1$ and $\mathbf{S}_2$, respectively. The goal of the algorithm is to segment regions in $\mathbf{S}_1$ and their corresponding regions in $\mathbf{S}_2$ such that the metric structures of the corresponding regions match optimally. More specifically, the algorithm computes a homeomorphism f between $\mathbf{S}_1$ and $\mathbf{S}_2$ **and** a set of closed curves γ_1 on $\mathbf{S}_1$. Let γ_2 denote the closed curve in $\mathbf{S}_2$ that is the image of γ_1 under f. The restriction of f to the complement $\mathbf{S}_1 \backslash \gamma_1$ is a diffeomorphism between $\mathbf{S}_1 \backslash \gamma_1$ and $\mathbf{S}_2 \backslash \gamma_2$. If $U_1, \cdots, U_n$ and $V_1, \cdots, V_n$ denote the collections of open components of $\mathbf{S}_1 \backslash \gamma_1$ and $\mathbf{S}_2 \backslash \gamma_2$, respectively, then f maps U_i diffeomorphically onto V_i for each i and f matches the Riemannian structures between that of U_i and of $V_i = f(U_i)$. See Figure 1. We solve the simultaneous segmentation and registration problem outlined above using a variational framework. The energy functional $\mathcal{E}$ is defined as a functional of a pair (f, γ_1), where γ_1 is a set of closed curves on $\mathbf{S}_1$ and $f : \mathbf{S}_1 \rightarrow \mathbf{S}_2$ is a homeomorphism which is C^∞ on $\mathbf{S}_1 \backslash \gamma_1$. Let $f^* g_2$ denote the pull-back metric (the first fundamental form) on g_1, and $\mathcal{E}(f, \gamma_1)$ is defined as

$$\mathcal{E}(f, \gamma_1) = \int_{\mathbf{S}_1 \backslash \gamma_1} |f^* g_2 - g_1|^2 dA + \int |df(\gamma_1(t))/dt|_{g_2} dt. \tag{1}$$

The second integral computes the length of the segmenting curve γ_2 on $\mathbf{S}_2$, while the first integral computes the matching (registration) cost between the two Rie-

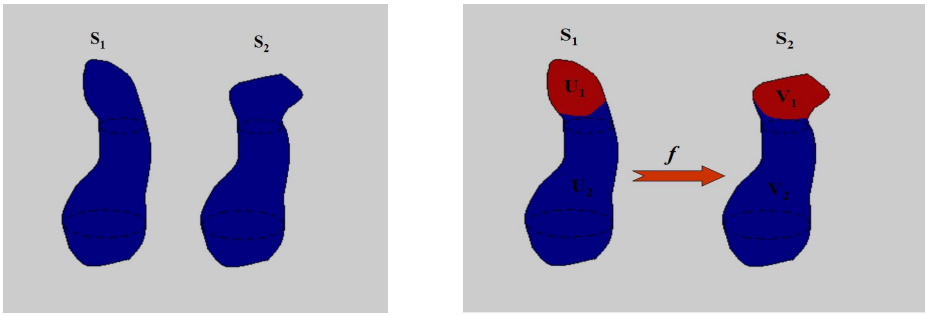

Fig. 1. Left: The two given surfaces, $\mathbf{S}_1$ and $\mathbf{S}_2$. **Right:** The desired result of simultaneous segmentation and registration of these two surfaces. The map f is a homeomorphism between $\mathbf{S}_1$ and $\mathbf{S}_2$. f maps the segmenting curve γ_1 on $\mathbf{S}_1$ to a curve γ_2 on $\mathbf{S}_2$, and f establishes correspondences between connected components of $\mathbf{S}_1 \backslash \gamma_1$ and $\mathbf{S}_2 \backslash \gamma_2$. U_1 and U_2 are two disjoint open sets in $\mathbf{S}_1 \backslash \gamma_1$, and V_1 and V_2 are two disjoint open sets in $\mathbf{S}_2 \backslash \gamma_2$. f maps U_i diffeomorphically onto V_i for each i. The restriction of f to U_i optimally matches the Riemannian structures of U_i and V_i.

mannian structures on the complement, $\mathbf{S}_1 \backslash \gamma_1$. By computing the length of the segmenting curve on $\mathbf{S}_2$ instead of $\mathbf{S}_1$, the energy functional tightly couples the two somewhat disparate processes, segmentation and registration. The integrand of the first term above, $|f^* g_2 - g_1|^2$, is the usual L^2-norm between the two tensors g_1 and $f^* g_2$. In local coordinates, it is given as the usual L^2-norm between the two matrices G_1 and $J^t G_2 J$:

$$|J^t G_2 J - G_1|^2_{g_1}, \tag{2}$$

where J is the Jacobian of f expressed in the local coordinates system, and G_1 and G_2 are the usual 2×2 positive definite matrices expressing the metric tensors g_1 and g_2 in the given local coordinates on $\mathbf{S}_1$ and the induced one (using f) on $\mathbf{S}_2$.

At this point, a comparison between our approach and the usual Mumford-Shah framework [9] is called for. Given an image I, the energy functional for Mumford-Shah is

$$\mathcal{MS}(I'\ \gamma) = \int_{\mathcal{D}} |I - I'|^2 dA + \int_{\mathcal{D} \backslash \gamma} |\nabla I'|^2 dA + \int |\gamma'(t)| dt.$$

In the above, $\mathcal{D}$ denotes the image and γ the segmenting curve on $\mathcal{D}$. The third term, which measures the length of γ, corresponds directly to the second term of $\mathcal{E}$ in 1. The first two terms of Mumford-Shah, representing data fidelity and smoothness, respectively, may seem to lack analogous entries in $\mathcal{E}$. However, we argue below that the first integral in the definition of $\mathcal{E}$ serves both as the fidelity and smoothing term for $\mathcal{E}$, and we can thus consider $\mathcal{E}$ as a generalization of the usual Mumford-Shah functional to the problem of simultaneous shape registration and segmentation.

For our problem, we want to *compare* two surfaces $\mathbf{B}_1$ and $\mathbf{B}_2$ via their respective Riemannian structures g_1 and g_2, and to extract a meaningful segmentation from the comparison. Instead of an image I' as in the usual Mumford-Shah framework, our variable is the map f between $\mathbf{S}_1$ and $\mathbf{S}_2$. In our case, the data fidelity requirement appears as the penalization of f for failure to match the Riemannian structures of the surfaces, while the smoothing term appears as the penalization of f for large first-order derivative magnitudes. A close examination of Equation 1 reveals that the first integral indeed contains both requirements. First, the norm between the two tensors f^*g_2 and g_1 as defined in Equation 2 clearly measures the quality of match between the two metric tensors. Second, the Jacobian J contains all of the first-order derivatives of f, and the squared-norm in Equation 2 indirectly measures the magnitudes of the first-order derivatives of f using both metrics g_1 and g_2. In sum, we can interpret the first term in Equation 1 as the integration of the first two terms in the usual Mumford-Shah functional $\mathcal{MS}$. In $\mathcal{MS}$, these two terms are necessarily separate because there is no natural connection between matching pixel intensities and restricting the gradient ∇I. In our problem, the connection between matching and smoothing is natural: since we are matching Riemannian structures, which are tensors defined on tangent spaces, any matching between these two structures has to involve some kind of (implicit) matching between tangent spaces, and the appearance of derivatives in defining the matching is then unavoidable.

2.1 Minimizing the Energy Functional $\mathcal{E}$

In implementing the energy minimization as defined above, we rely on alternating iterations between a Chan-Vese level set curve segmenter [10], extended to the general 2D manifold domain, which produces a step towards the best estimate of the segmentation of the energy field given the current map, and a sparse Hessian-exploiting quasi-Newton optimization scheme acting on the non-rigid map between the parametric domains, which provides a step towards the minimum-energy map between the surfaces as constrained by the current position of the segmenting curve.

The manner in which the curve separates the evolution occurring on its exterior from that occurring on its interior relates to a Neumann condition imposed across the boundary, akin to what is suggested in the Mumford-Shah curve evolution implementation in [11]. The curve length minimizing term implementation is straightforward, though curvature computation must take nonuniform surface length elements into account.

Hippocampus surface data used in this investigation was brought into rigid alignment by ICP in a preprocessing step, then 2D sliced and manually segmented by a trained neuro-anatomist into a 40x21 grid periodic in one direction. This grid in effect forms a parametric domain for the corresponding surface, and the Riemannian surface characteristics are thus calculated by numerical methods and stored in said domain. Note that it is precisely the gridded structure of the domain that makes the Chan-Vese implementation as straightforward as it is (save adjustment for the surface length elements). Also, it is appropriate, if one

wishes, to view Chan-Vese as a regularized thresholder: it is for the regularization that we use it.

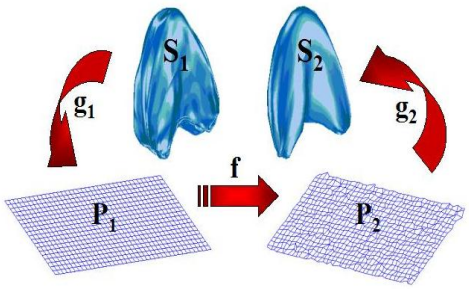

Fig. 2. Illustration of framework. Note that the map f between the parametric domains can be visualized as the deformation of a regular grid representing the left parametric domain P_1.

3 Results

We here present examples of the algorithm in action, validating its behavior on synthetic cases and demonstrating results obtained on real hippocampus pairs. As a starting point, we consider the trivial case of a pair of cylinders, where one member of the pair has had its surface distorted according to an outward normal vector field of magnitude dictated by a Gaussian distribution. That is to say, its surface has been bumped. In Fig. 3, we see a comparative case involving a cylinder with two such distortions versus its unaltered counterpart. The intuition behind our method asserts that we should "blame" the applied bumps for observed differences between the Riemannian surface characteristics of the two shapes. Obviously, then, the desired result is the segmentation of the two clearly visible bumps on the left cylinder, which is what we observe. (Red "hot spots", the curve interiors, represent regions of highest deformation energy under the current map and segmentation. The curve itself is found at the border between the red and blue regions.) Note the textured quality of the surfaces, which occurs as a result of the addition of a SD=0.02 zero-mean Gaussian noise field to the (X,Y,Z) coordinates of the (unit radius) cylinders prior to output, for demonstration of robustness in the face of noise in surface scanning.

As indicated previously, the case of the cylinder is a trivial one in this context in that there exists a natural parameterization of the surface which results in a constant metric tensor field over the parametric domain. Thus, we further validate using synthetic alterations to real data exhibiting arbitrary shape, within the confines of cylindrical topology. Since the hippocampus meets these criteria and forms part of our clinical motivation, it seems a natural choice. Selecting

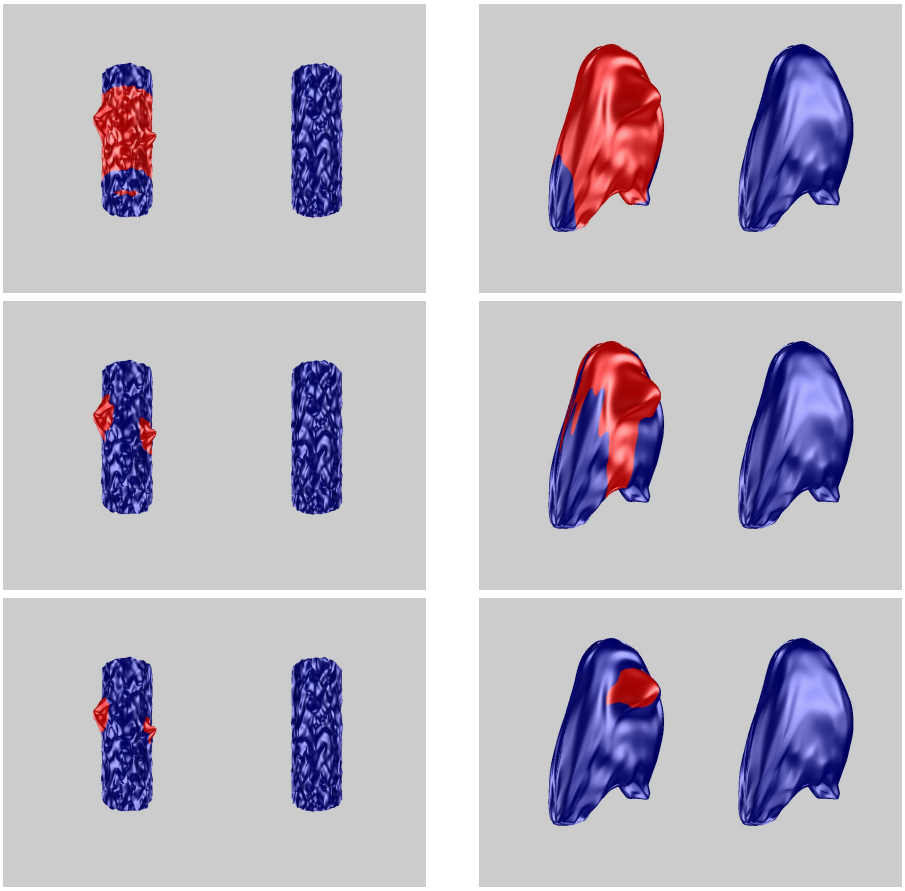

Fig. 3. Left column: Segmentation of prominent distortions (bumps) from cylindrical surface, based on comparison to homologous cylinder lacking bumps. Surfaces are both distorted with distinct random noise fields (seen as textured effect). In this and all following examples, we show the segmented region only on the left surface of the pair. A corresponding region exists on each right surface, with the correspondence maintained through the (unseen) deformation map. Top: naive initialization; Middle: 3 iterations; Bottom: 10 iterations. **Right column:** Segmentation of synthetic bump from surface of real hippocampus, based on comparison to same hippocampus without bump. The two hippocampi are under different parameterizations. Top: naive initialization; Middle: 5 iterations; Bottom: 20 iterations.

a single hippocampal horn, we again apply a small outward normal Gaussian deformation to its surface, whose location we hope to see the segmenter recover. However, rather than illustrating invariance to surface point cloud noise, we now seek to additionally validate invariance of the solution to reparameterizations of the given surfaces. To pause for conceptual refreshment, recall that in comparing surfaces we are in fact comparing two parametric domains based on the two tensor-valued functions corresponding to each of those domains, those functions being the Riemannian surface structure information of each parameterized surface. But just as two surfaces (or their parametric domains, as we do in our implementation) can be mapped to one another nonrigidly in infinitely many ways, so too can each parametric domain map to its topologically-equivalent parameterized surface in infinitely many ways. The Riemannian structure tensor fields themselves depend upon which of these infinitely many parameterizations are chosen. When synthetically producing one surface from another, we possess a trivial alignment between the domains under which the two tensor fields are identical, excepting the patch corresponding to the synthetic bump. In real cases, of course, data will not have this sort of property. The validation is thus incomplete until we can show that under a suitably deformed reparameterization of one of the surfaces being compared, a result comparable to the above is achieved. We can reparameterize simply by randomly generating a distortion of the parametric domain's regular grid, then interpolating a new point cloud from the given data based on the new locations of the gridpoints. It is completely appropriate to confine said distortion to the order of a pixel in each direction: after all, we have assumed close rigid alignment and sensibly spaced data points from our preprocessing, and if these constraints are violated, they can always be reinstated by repeating the algorithms that enforced them in the first place. The purpose of the nonrigid registration evolution is to further refine the surface correspondence following bulk rigid alignment of the shapes based on the surface characteristics observed locally, not to register raw point clouds. We cannot show the parameterization warp used due to space constraints: it is of 0.1 SD in each coordinate.

It is visually evident that the segmentation succeeds despite the change of parameterization of one of the surfaces being compared. The deformation map between the parametric domains (not shown) undergoes considerable evolution through this process. This evolution can be best understood as the method's attempt to invert the reparameterization, to achieve the optimal alignment of the identical surface patches external to the segmented bump, while at the same time trying to find the minimum energy description of the acknowledged deformation of the patch on the curve's interior.

In presenting the following real example, we issue the caveat that the method's success lies in the eye of the beholder, because there is of course no ground truth associated with this sort of problem. For this reason we have selected a case with an answer which can easily be agreed upon. Fig. 4 shows the left and right hippocampi clinically segmented from an MRI scan of an epilepsy patient. We take it as self-evident that there exists a localizable region of high distortion between

the left and right shapes, and assume that one will agree that our method has in fact converged to a segmentation of that region. Illustrations of the segmented energy field and parametric warp evolutions have been withheld due to space limitations.

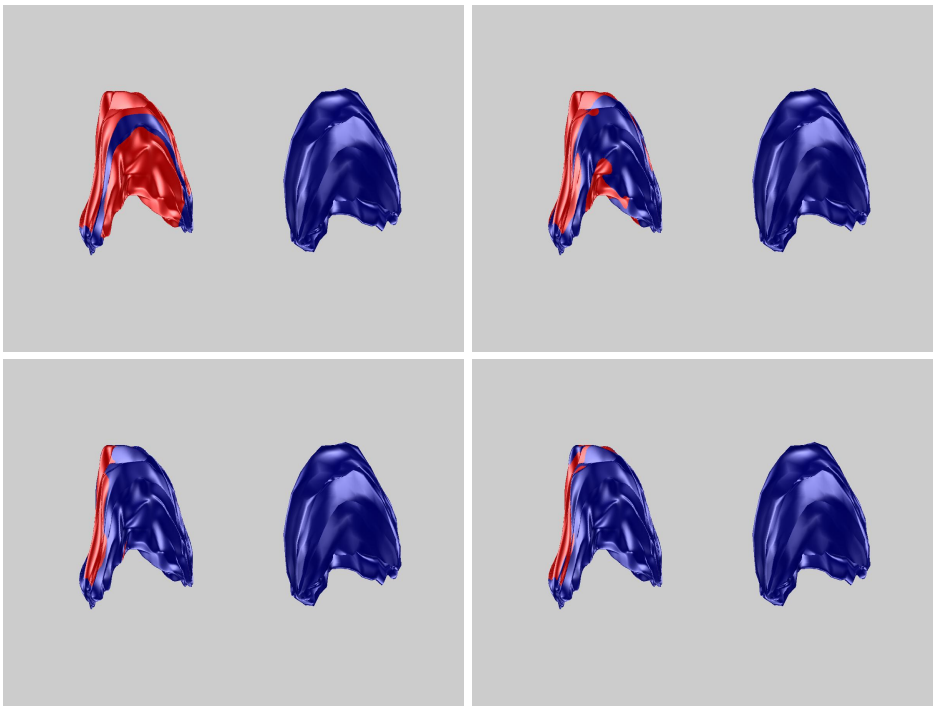

Fig. 4. Segmentation of distortion between real hippocampi of epileptic. Top left: naive initialization; Top right: 1 iteration; Bottom left: 5 iterations; Bottom right: 20 iterations.

4 Conclusion

We have presented a novel scheme for simultaneous nonrigid registration and segmentation of homologous shapes, wherein the interdependent registration and segmentation processes are driven by intrinsic geometric characteristics of the shapes themselves. As such, we are able not only to identify surface pairs representing large deformations, but also to specify which subregions of the shapes appear most likely to be involved in the deformation, to produce an estimate of the implicit deformation field, and to quantify the deformation energy of the segmented subregions separately from that of the remaining surface patches.

Given the number of considerations interplaying, the variational principle driving the evolution is notably compact. We have seen that the method's results are reasonably invariant to local parameterization changes and added surface noise, and results on real data (while inherently lacking in ground truth for comparison) appear promising. Future efforts could include the application of the method to extensive real datasets that have been clinically preclassified according to the presence or absence of pathology, as well as an attempt to correlate the locations and intensities of the segmented deformed regions with the given pathology classifications. It should also be noted that the two-region Chan-Vese segmentation method (which is often used for its strong balance of effectiveness and relative simplicity of implementation) is merely a good "first draft" of a segmentation scheme for this sort of problem. A more robust segmentation method, such as the general Mumford-Shah functional extended to arbitrary surfaces, would doubtless prove beneficial, as our energy fields will not always neatly obey the two-mean framework. Finally, it goes without saying that all evolution/optimization-based methods (apart from the most trivial cases) suffer from some amount of initialization-dependence. Intelligent initialization schemes could be expected to further boost performance.

References

1. Gerig, G., Muller, K.E., Kistner, E.O., Chi, Y.Y., Chakos, M., Styner, M., Lieberman, J.A.: Age and treatment related local hippocampal changes in schizophrenia explained by a novel shape analysis method. In: MICCAI. (2003) 653–660
2. Styner, M., Lieberman, J.A., Gerig, G.: Boundary and medial shape analysis of the hippocampus in schizophrenia. In: MICCAI. (2003) 464–471
3. Wang, Y., Gu, X., Hayashi, K.M., Chan, T.F., Thompson, P.M., Yau, S.T.: Surface parameterization using riemann surface structure. In: Proc. Int. Conf. on Computer Vision. (2005) 1061–1066
4. Wang, Y., Chiang, M.C., Thompson, P.M.: Mutual information-based 3d surface matching with applications to face recognition and brain mapping. In: Proc. Int. Conf. on Computer Vision. (2005) 527–534
5. Wang, Y., Chiang, M.C., Thompson, P.M.: Automated surface matching using mutual information applied to riemann surface structures. In: MICCAI. (2005) 777–674
6. A.Yezzi, Zollei, L., Kapur, T.: A variational framework for joint segmentation and registration. In: CVPR - MMBIA. (2001)
7. Unal, G.B., Slabaugh, G.G.: Coupled pdes for non-rigid registration and segmentation. In: CVPR. (2005) 168–175
8. Wang, F., Vemuri, B.C.: Simultaneous registration and segmentation of anatomical structures from brain mri. In: MICCAI. (2005) 17–25
9. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Commun. Pure Appl. Math **42**(4) (1989)
10. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on Image Processing **10**(2) (2001) 266–277
11. Tsai, A., Yezzi, A.J., Willsky, A.S.: Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification. IEEE Transactions on Image Processing **10**(8) (2001) 1169–1186

Geodesic Image Normalization in the Space of Diffeomorphisms

Brian B. Avants and C. L. Epstein and J. C. Gee

Depts. of Radiology and Mathematics,
University of Pennsylvania
Philadelphia, PA 19101

Abstract. Quantitative, deformable mappings between images are increasingly important as measurement tools in biology and medicine. The theory of diffeomorphisms (smooth differentiable bijections with differentiable inverse) provides a mathematical foundation for the computation and interpretation of these maps. Miller, in particular, has used this theory to develop image normalization techniques that rely on a distance measurement as a regularizer. Here, we use this metric within a general transformation framework that explicitly parameterizes the image to image mapping as a symmetric geodesic path in the space of diffeomorphisms. The key difference between our approach and Miller's is that ours does not choose a single end-point (or template image) from which to measure the map. Rather, we search for the shortest diffeomorphism (smallest deformation) between images by optimizing the transformation with respect to both of its end-points. This algorithm, geodesic normalization, provides solutions that are invariant to which of the input coordinate systems (images) is chosen as a reference. This allows the method to compute metric distances, have truly symmetric performance and to give full space-time solutions that are invertible and diffeomorphic in the discrete domain. Finally, our algorithm guarantees that our solutions and their inverses are consistent to a sub-pixel level.

1 Introduction

Computational anatomy (CA) uses imaging to make quantitative measurements of the natural world. One may view CA as the science of biological shape and its variation, with roots in the work of Charles Darwin and D'Arcy Thompson [1, 2]. The wide availability of high resolution *in vivo* functional and structural imaging has caused a rapid increase in CA's relevance and prominence. Currently, this developing science's primary tools are the topology preserving diffeomorphic transformations. These transformations are used to map an individual image J into the space of a template image, $\bar{I}$, which serves as a common coordinate system. When a population of images is mapped together by these transformations, each voxel in individual space corresponds smoothly with a single voxel in the reference space. This process creates a continuous spatial map of population

information detailing, for example, the relative volume, functional activation or diffusion at a given anatomical position, such as the anterior hippocampus or occipital lobe gray matter. Topology preserving transformation (TPT) are of special interest for this technology as they will introduce neither folds nor tears in this map and will preserve the continuity of curves and surfaces.

Topology preserving normalization permits comparisons to be made across time points in an individual's disease process or to study development patterns across a large population [3–6]. Miller, et al. showed that Large Deformation Diffeomorphic Metric Matching (LDDMM) is able to localize hippocampal activation to provide increased statistical significance in functional imaging studies [7]. Avants et al used a Lagrangian diffeomorphic normalization technique to map and statistically differentiate functionally homologous structures between species [8]. Furthermore, large deformation mappings are better able to separate structural and signal (intensity) differences in population studies, particularly in the presence of atrophy or high shape variation. Diffeomorphic methods may also be extended to normalize vector or tensor images [9].

The CA studies cited above were enabled by recently developed, theoretically well-founded methods for studying topology preserving variation. The major advancement in this aspect of CA technology is to base the work in the space of diffeomorphisms. The diffeomorphic space is the broadest smooth, topology preserving space and allows one to very accurately capture both large and small deformation differences in shape. The collection of these transformations forms a mathematical group. Grenander [10], Mumford [11], Miller [12], Trouve [13] and Younes [14] have studied this group space in the context of computer vision and deformable image transformation and have derived Euler-Lagrange equations for CA [15].

Our approach to image normalization is based upon Arnol'd's definition of a symmetric, time-parameterized *shortest path* (geodesic) between two diffeomorphic configurations of a domain [16]. We argue that this view is fundamental to the theoretical foundation of diffeomorphic normalization, essential for computing true metric distances and desirable for its symmetry properties. Our algorithm will satisfy desirable continuity, anonymity and unanimity conditions [17] as well as the metric measurement properties needed for geodesics. This yields a new algorithm, geodesic normalization (GN), that parameterizes the deformation between an image pair with respect both ends of a geodesic path.

2 Mathematical Background on Diffeomorphisms

We now discuss some basic facts from the theory of diffeomorphisms, the mathematical underpinnings of GN. This section is derived from Arnol'd [18] and Marsden and Ratiu [19]. In this section, we will refer to ϕ as a geodesic in the space of diffeomorphisms and ϕ_1 and ϕ_2 as the components of ϕ, as described below. We also assume that the images and velocity fields referred to below are sufficiently differentiable and that we are only interested in transformations that visibly change the image. For example, a diffeomorphism of a constant image will

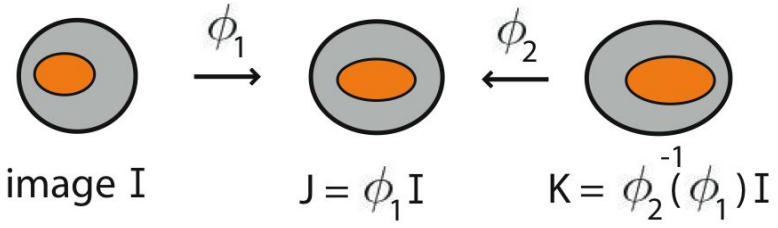

Fig. 1. A diffeomorphism is used to map image I into image J (left to center) via coordinate transformation. Composing two diffeomorphisms ϕ_1 and the inverse of ϕ_2 enables us to use the maps from K to J and from I to J to make a coordinate transformation from I to K.

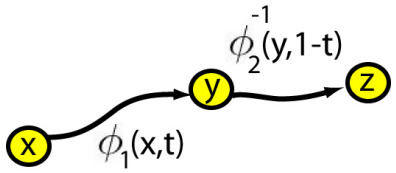

Fig. 2. An illustration of the geodesic path taken, in time, by the particle at position $\mathbf{x}$ in domain Ω. The path is geodesic if the diffeomorphism associated with the domain minimizes the distance metric in equation 2. The geodesic, ϕ is the whole path. Its points are traversed via ϕ_1 and ϕ_2.

be reduced to the identity. Image differentiability is required for the derivative computations necessary in the normalization method. Velocity field regularity guarantees the integrability necessary for generating diffeomorphisms. Typically, convolution with a Gaussian ensures image differentiability while a linear operator, L, induces sufficient smoothness on the velocity field. See Dupuis [20] for a discussion of regularity requirements on images and velocity field and the well-posedness given by diffeomorphic regularization.

Recall that a diffeomorphism is a smooth one-to-one and onto map with a smooth inverse. We always index diffeomorphisms with a spatial coordinate ($\mathbf{x}$, $\mathbf{y}$ or $\mathbf{z}$) and, if necessary, a time variable, t. However, we drop the spatial and temporal indexes for brevity where the meaning is clear. We define the diffeomorphic operations that we need as,

1. Coordinate transformation: this operation changes the coordinate system in which an object (image, vector field) is represented. The operation ϕI transforms the image $I(\mathbf{x})$ into the deformed image $\tilde{I}(\mathbf{y}) = I(\phi(\mathbf{x}, t))$ where the intensity $I(\mathbf{x})$ is equivalent to the intensity $\tilde{I}(\mathbf{y})$. We may also apply this operation to vector (or velocity) fields.
2. Transformation composition: this operation links diffeomorphisms together, generating a new map via $\phi_2(\phi_1(\mathbf{x}, t_1), t_2)$.

Understanding these operations are essential in image normalization. We illustrate them in figure 1.

Families of diffeomorphisms can be generated by integrating time-dependent velocity (vector) fields through an ordinary differential equation [18],

$$\frac{d\phi(\mathbf{x},t)}{dt} = \boldsymbol{V}(\mathbf{x},t). \tag{1}$$

This differential equation defines the change of the map, ϕ, by the value of a velocity field which is a smooth vector field. Velocities tell us how particles are moving through space: $\boldsymbol{V}$ assigns motion to a specific material point while $\boldsymbol{v}$ defines motion in a fixed coordinate system. The $\boldsymbol{V}$ in equation 1 is a *material velocity* in the Lagrangian frame. The *spatial velocity* is computed at the tangent space to the diffeomorphism at time, t, such that $\boldsymbol{v}(\mathbf{y},t) = \boldsymbol{V}(\mathbf{x},t)$ where the change in coordinates is indicated by the use of $\mathbf{x}$ in $\boldsymbol{V}$ and $\mathbf{y}$ in $\boldsymbol{v}$. We may also explicitly indicate the change in coordinates by using the map between the two coordinate systems, denoted by $\phi(\mathbf{x},t) = \mathbf{y}$, with inverse $\phi^{-1}(\mathbf{y},t) = \mathbf{x}$. Then, $\boldsymbol{V}(\mathbf{x},t) = \boldsymbol{v}(\phi(\mathbf{x}),t)$.

Deformable diffeomorphisms commonly used in image registration map domain Ω to itself. The map at the boundary, $\partial\Omega$, may also be defined as the identity, $\phi(\partial\Omega) = \partial\Omega$. This boundary constraint assumes that rigid motion has been factored out of the transformation between images. It also guarantees the transformation is everywhere one-to-one and onto and restricts the solution space to the diffeomorphic subgroup, $\mathcal{G}_0$.

$\mathcal{G}_0$ is a Frechet Lie group [21] when $\mathcal{G}_0$ is C^∞. The length of a diffeomorphic path between elements in this space is similar to the length of a curve, $\mathcal{C}$, connecting two points in Euclidean space, $l(\mathcal{C}) = \int_0^1 \|d\mathcal{C}/dt\|dt$, where the Euclidean length of the curve's tangent vector is integrated over its parameterization. Distances in the space of diffeomorphisms are infinite-dimensional analogies of curve length, where the infinitesimal increment in distance is given by a Sobolev norm, $\|\cdot\|$, operating on the tangent to the diffeomorphism (the spatial velocity) [20]. A *geodesic* between ψ_1 and ψ_2, two elements of $\mathcal{G}_0$, is defined by taking the infimum over all such paths [15],

$$D(\phi(0),\phi(1)) = \inf_{\phi} \int_0^1 \|\boldsymbol{v}(\phi(\mathbf{x},t))\|_L dt, \tag{2}$$
$$\phi(0) = \psi_1 \text{ and } \phi(1) = \psi_2,$$

where $\|\cdot\|_L$ is the Sobolev norm with respect to linear operator, L. Taking the infimum guarantees that we have a geodesic between the elements in $\mathcal{G}_0$. The length of the geodesic gives a metric distance, does not depend on the origin of its measurement (it has right invariance) and is the basis for GN as well as Miller, Trouve and Younes's work.

A geodesic in the space of diffeomorphisms thus defines the shortest route between two diffeomorphic transformations. Each transformation defines a single, unique configuration of the coordinate system. The length of the path itself is (trivially) symmetric, that is, $D(\phi(0),\phi(1)) = D(\phi(1),\phi(0)$ and satisfies metric properties. Furthermore, for all time $t \in [0,1]$, we have $\phi_2^{-1}(\phi_1(\mathbf{x},t),1-t) = \phi_1(\mathbf{x},1) = \mathbf{z}$. Rearranging this equation, we gain intermediate points along the

geodesic from $\phi_2(\mathbf{z}, 1-t) = \phi_2(\phi_1(\mathbf{x}, 1), 1-t) = \phi_1(\mathbf{x}, t)$. In this way, we see that *points along the geodesic are parameterized equivalently from coordinates at either end-point*. We will now introduce this coordinate system invariant parameterization into our normalization technology.

3 Geodesic Image Normalization

The goal of image registration, in general, is to find, for each $\mathbf{x}$ in I, the $\mathbf{z}$ in J that gives $I(\mathbf{x}) = J(\mathbf{z})$ or, alternatively, $f(J(\mathbf{z}))$ where f is an intensity-space transformation. If f is the identity, then the intensity at $\mathbf{x}$ in I should be equivalent to the intensity from coordinate $\mathbf{z}$ in J. The mapping from $\mathbf{x}$ to $\mathbf{z}$ may be written $\phi(\mathbf{x}) = \mathbf{z}$, from image I to image J such that points in I are in one-to-one correspondence with points in J.

When such maps are diffeomorphisms, we include a time parameter, t, that indexes the temporal evolution of ϕ. We refer to coordinates $\mathbf{x}$ in the time zero I domain, $\mathbf{z}$ in the time zero J domain and $\mathbf{y}$ in a common coordinate system that moves along the curve connecting $I(\mathbf{x})$ and $J(\mathbf{z})$. If this curve is a geodesic, then it is symmetric and will follow the same path whether starting from I or J. This symmetry means our time parameterized maps may be viewed from either endpoint at I or J such that $\phi_1(\mathbf{x}, t) = \mathbf{y} = \phi_2(\mathbf{z}, 1-t)$. This formulation allows us to deform I and J such that, for any $t \in [0,1]$ $I(\phi_1(\mathbf{x}, t)) = J(\phi_2(\mathbf{z}, 1-t))$. Such a motion gives a dense map in both space and time and is shown in figure 2 for one point in the image domain. The total mapping between the images is gained through the composition of these two components, $\phi(\mathbf{x}, 1) = \phi_2^{-1}(\phi_1(\mathbf{x}, t), 1-t)$. GN will exploit this geodesic view to gain symmetry. The algorithm will thus be able to compute the distance between two images, whereas previous algorithms computed asymmetric distance (not symmetric, therefore not a metric distance), due to a biased gradient descent approach that originates in a parameterization of the geodesic with respect to only one endpoint.

Let us now consider the case when we are given two images, I and J, of the same class, known to be (approximately) diffeomorphic. Here, we know neither the path in time nor which image should be considered as the template or reference image. We make the identification of $I(0)$ with I and $I(1)$ with J. We now seek to find the shortest diffeomorphism between these images such that $\phi_1(t)I = \phi_2(1-t)J$.

We now translate this example into a variational optimization problem. The variational energy for geodesic normalization therefore seeks ϕ_1 and ϕ_2 in order to locate the geodesic connecting I and J. Then, $\phi_1(\mathbf{x}, t)I = \phi_2(\mathbf{z}, 1-t)J$, gives the similarity term, $|\phi_1(t)I - \phi_2(1-t)J|^2$. The forward and backward energy is

then, using t as a parameter and solving to time $t = \bar{t}$,

$$E_{GN}(I,J) = \inf_{\phi_1} \inf_{\phi_2} \int_{t=0}^{\bar{t}} \omega\{ \|\boldsymbol{v}_1\|_L^2 + \|\boldsymbol{v}_2\|_L^2\}dt + \int_\Omega |\phi_1(\bar{t})I - \phi_2(1-\bar{t})J|^2 d\Omega. \quad (3)$$

Subject to:

each $\phi_i \in \mathcal{G}_0$ the solution of:

$$d\phi_i/dt = \boldsymbol{v}_i(\phi_i(t)) \text{ with } \phi_i(0) = \mathbf{Id}. \quad (4)$$

Minimization with respect to ϕ_1 and ϕ_2, upholding the arc length constraint, provides the *geodesic normalization*. Landmarks may also be included in this energy, as in our previous work [8], by dividing the similarity term, as done with the image match terms above. A similar image matching equation appeared in [22] and [23] as part of a derivation for template generation.

Once this problem is solved, the total symmetric normalization transformation from I to J is $\phi_1(\mathbf{x},1) = \phi_2^{-1}(\phi_1(\mathbf{x},0.5),0.5)$ and from J to I, $\phi_2(\mathbf{z},1) = \phi_1^{-1}(\phi_2(\mathbf{z},0.5),0.5)$. This is distinct from inverse consistent image registration [24] in which a variational term is used to estimate "inverse consistency" and symmetry and invertibility are not guaranteed. The inverse consistency is inherent to our method and is shown for synthetic data in figure 3 and for real data in figure 4. The algorithm is useful for generating shape means as well as symmetric geodesic image interpolation, formulated in [25].

We will show below that the maps computed by this algorithm satisfy Eckmann's continuity, symmetry (anonymity) and unanimity conditions. Eckmann and Weinberger discuss the existence of such maps [17, 26] and note the connection with the *generalized mean*. Denote a symmetric map, ϕ, connecting I and J. Eckmann's properties, adapted for image normalization, are then

1. Continuity: the map should vary continuously with the inputs I, J.
2. Symmetry / Anonymity: ϕ does not depend on permutations of I, J. This is verified if $\mathcal{A}(I,J) = \phi$ then $\mathcal{A}(J,I) = \phi^{-1}$.
3. Unanimity [17]: the map should output $\phi = \mathbf{Id}$ if $I = J$.

A map that violates anonymity (2 above) is labeled asymmetric. We now argue that GN satisfies the three conditions given above.

Theorem 1. *The solutions, $\phi_1(\mathbf{x},t)$ and $\phi_2(\mathbf{y},s)$, found by GN satisfy the three generalized mean axioms above.*

Continuity. Continuity was shown in Dupuis's proof of well-posedness for the diffeomorphic variational image matching problem [20]. As our problem is, in a global sense, identical, continuity is inherited.

Unanimity. The unanimity condition is also satisfied as all velocities will be $\mathbf{0}$ if all images are identical.

Anonymity. We prove anonymity by simply checking that permuting the labels in the variational energy produces an identical optimization problem, which

visual inspection confirms. Furthermore, the Euler-Lagrange equations depend symmetrically on gradients of both I and J. ■.

Geodesic Normalization Implementation: Algorithm 1 states the locally optimal GN algorithm without landmarks, as described in previous sections. A similar approach can be used for landmark matching. The input to the algorithm is a pair of images and a user set number of time interpolation points, $\{t_i\}$, where $t_0 = 0$ and $t_n = 0.5$ and n is even. The output is ϕ_1, ϕ_2 defined on $\Omega \times [0, 0.5]$ and their inverses. These define the full ϕ_i mappings and their inverses at (deformation) time 1.

Algorithm 1 : Geodesic Normalization (GN)

The algorithm notation is the same as in the body of the paper. We compute ϕ_1^{-1} and ϕ_2^{-1} with the inversion method used in our Lagrangian Push Forward (LPF) method [8].

1: $\forall t_i$ Initialize $\phi_1(t_i) = \phi_2(t_i) = \mathbf{Id}$.
2: **while** $\Delta E_{GN} > \epsilon_1$ **do**
3: Set $E_{old} = E_{new}$.
4: Compute $\phi_1(\mathbf{x}, 0.5)I$ and $\phi_2(\mathbf{x}, 0.5)J$.
5: Compute velocities from the symmetric Euler-Lagrange equations of 3.
6: Set $\|\boldsymbol{v}_1\| = \|\boldsymbol{v}_2\| = \min(\|\boldsymbol{v}_1\|, \|\boldsymbol{v}_2\|)$.
7: Update $\phi_1(0.5)$ and $\phi_2(0.5)$ by gradient descent such that, for $i = 1, 2$, $\phi_i(\mathbf{x}, 0.5) = \phi_i(\mathbf{x}, 0.5) + \lambda \boldsymbol{V}_i(\mathbf{x}, 0.5)$ where $\boldsymbol{V}(\mathbf{x}, 0.5) = \boldsymbol{v}(\phi_i(\mathbf{x}, 0.5))$ and λ is a gradient step length.
8: Starting at time 0.5 and going backwards toward 0, for all time points, t_i where $0 < t_i < 0.5$, update ϕ_1, ϕ_2 by gradient descent on the length of ϕ_1 and ϕ_2. Note that $\phi_i(0) = \mathbf{Id}$ and $\phi_i(0.5)$ does not change in this step.
9: Use the inversion method (below) to find ϕ_1^{-1} and ϕ_2^{-1} over all time.
10: Compute E_{new} from equation 3.
11: $\Delta E_{GN} = E_{old} - E_{new}$.
12: **end while**

Symmetry (guaranteed sub-pixel invertibility and algorithmic independence to input permutations) is built into GN and allows us to symmetrically match images to the degree that discrete diffeomorphisms are invertible. An example of a symmetric image registration when large deformations are present is shown in figure 3. Here, we use GN to find spatial correspondences by deforming a half C to a full C. We also illustrate GN for normalizing severely atrophied brains, some of which also suffer from the presence of lesions, in figure 4.

4 Discussion

A common problem with image normalization algorithms, in general, is asymmetry. Registering image I to J may not produce the same correspondences as

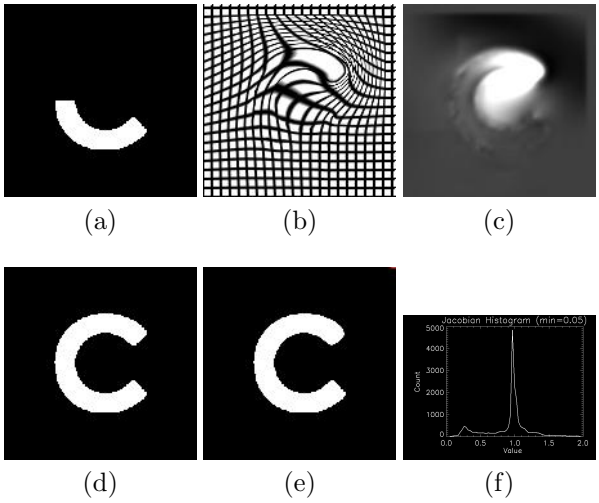

Fig. 3. The $\frac{1}{2}C$ (a) is registered to the full C (d). The grid of the deformation from (a) to (d) is shown in figure (b). The Jacobian of the transformation from C to $\frac{1}{2}C$ is in (c) while its histogram is in (f). The $\frac{1}{2}C$ to C result is in (e).

registering J to I. This problem has been addressed by methods that compute mappings in both directions. Thirion developed such an approach for Demons image registration [27]. Christensen's inverse consistent image registration (ICIR) [28] uses a similar idea. Both algorithms rely upon estimating a measure of "consistency", defined as the difference between the mapping from I to J and the mapping from J to I. Define $\mathbf{x}$ in image I, its displacement $\mathbf{u}(\mathbf{x})$ and their sum as $\mathbf{y} = \mathbf{x} + \mathbf{u}(\mathbf{x})$. Similarly, define $\mathbf{z}$ in image J and its displacement $\mathbf{w}(\mathbf{z})$. Consistency at $\mathbf{x}$ is $C(\mathbf{x}) = \|\mathbf{y} + \mathbf{w}(\mathbf{y}) - \mathbf{x}\| = \|\mathbf{u}(\mathbf{x}) + \mathbf{w}(\mathbf{y})\|$. Both algorithms attempt to minimize $C(\mathbf{x})$ and, similarly, $C(\mathbf{z})$. However, neither method guarantees the inverse's existence nor gives a well-defined numerical method for its computation.

Geodesic image normalization subsumes the above approaches by formulating the normalization process in space and time with a diffeomorphic parameterization. We do not have to "check" the consistency as (via Theorem 1) the consistency is guaranteed by the sub-pixel invertibility of ϕ_1 and ϕ_2 and the fact that these transformations may be composed together. Furthermore, our methods explicitly optimize these transformations in the large deformation space.

5 Conclusion

We explicitly optimize the length of ϕ, parameterized symmetrically by its components, ϕ_1, ϕ_2. Optimizing the length of a diffeomorphism is an alternative to finding a geodesic by solving the Euler equations [29]. Our method is shown

to satisfy the axioms of continuity, symmetry and unanimity. Satisfying these properties eliminates problems of algorithmic asymmetry. Finally, GN gives a robust estimate to geodesic distance that is invariant to which image, of a pair, is selected as template. Future work will focus on empirically demonstrating improved performance due to symmetrically parameterizing the large deformation image normalization problem.

References

1. C. Darwin, *Origin of Species*, John Murray: London, UK, 1856.
2. D. W. Thompson, *On Growth and Form*, Cambridge University Press, England, 1917.
3. A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis i: Segmentation and surface reconstruction," *Neuroimage*, vol. 9, no. 2, pp. 179–194, 1999.
4. P. Thompson and A. Toga, "A surface-based technique for warping 3-dimensional images of the brain," *IEEE Trans. Medical Imaging*, vol. 15, no. 4, pp. 402–417, 1996.
5. J. G. Csernansky, S. Joshi, L. Wang, J. W. Haller, M. Gado, J. P. Miller, U. Grenander, and M. I. Miller, "Hippocampal morphometry in schizophrenia by high dimensional brain mapping," *Proc. Natl. Acad. Sci. (USA)*, vol. 95, no. 19, pp. 11406–11411, 1998.
6. C. Studholme, V. Cardenas, R. Blumenfeld, N. Schuff, H. J. Rosen, B. Miller, and M. Weiner, "Deformation tensor morphometry of semantic dementia with quantitative validation," *Neuroimage*, vol. 21, no. 4, pp. 1387–1398, 2004.
7. M. I. Miller, M. F. Beg, C. Ceritoglu, and C. Stark, "Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping," *PNAS*, vol. 102, no. 27, pp. 9685–9690, 2005.
8. B. Avants, P. T. Schoenemann, and J. C. Gee, "Landmark and intensity-driven lagrangian frame diffeomorphic image registration: Application to structurally and functionally based inter-species comparison," *Medical Image Analysis*, vol. 10, pp. 397–412, 2006.
9. C. Yan, M. I. Miller, R. L. Winslow, and L. Younes, "Large deformation diffeomorphic metric mapping of vector fields," *tmi*, vol. 24, no. 9, pp. 1216–1230, 2005.
10. U. Grenander, *General Pattern Theory*, Oxford University Press, New York, 1993.
11. D. Mumford, "Pattern theory and vision," *Questions Matheematiques En Traitement Du Signal et de L'Image*, vol. 3, pp. 7–13, 1998.
12. F. Beg, M. Miller, A. Trouve, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *Int. J. Comp. Vision*, vol. 61, pp. 139–157, 2005.
13. A. Trouve, "Diffeomorphism groups and pattern matching in image analysis," *Intl. J. Comp. Vis.*, vol. 28, no. 3, pp. 213–221, 1998.
14. A. Trouve and L. Younes, "On a class of diffeomorphic matching problems in one dimension," *SIAM Journal on Control and Optimization*, vol. 39, no. 4, pp. 1112–1135, 2000.
15. M. Miller, A. Trouve, and L. Younes, "On the metrics and Euler-Lagrange equations of computational anatomy," *Annu. Rev. Biomed. Eng.*, vol. 4, pp. 375–405, 2002.

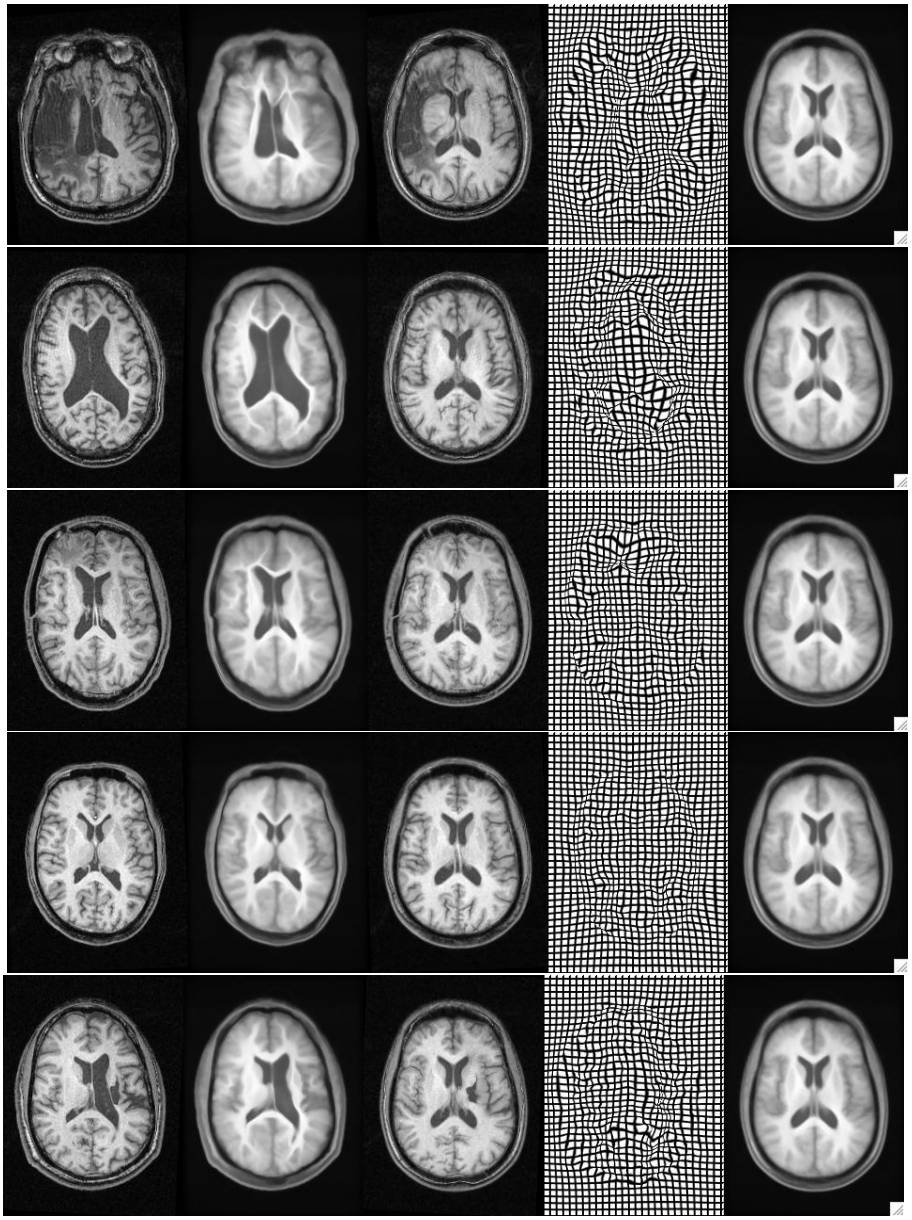

Fig. 4. We volumetrically map an average image template to a set of lesioned and/or atrophied brains using GN. The left column shows a slice of the original subject image. The second column shows the template mapped to the subject space. The center column shows the subject mapped to template space. The second to last column shows the grid deformation from template to subject. The final column shows the original template slice. Asymmetric methods do not perform as well for normalizing this dataset, particularly in the presence of difficult and unpredictable lesions.

16. V. I. Arnold, "Sur la gomtrie direntielle des groupes de lie de dimension innie et ses applications l'hydrodynamique des uides parfaits," *Ann. Inst. Fourier (Grenoble)*, vol. 16, no. 1, pp. 319–361, 1966.
17. S. Weinberger, "On the topological social choice model," *J. Econom. Theory*, vol. 115, no. 2, pp. 377–384, 2004.
18. V. I. Arnold, *Ordinary Differential Equations*, Springer-Verlag: Berlin, 1991.
19. J. Marsden and T. Ratiu, *Introduction to Mechanics and Symmetry*, Springer-Verlag: New York., 1999.
20. P. Dupuis, U. Grenander, and M. I. Miller, "Variational problems on flows of diffeomorphisms for image matching," *Quarterly of Applied Mathematics*, vol. 56, no. 3, pp. 587–600, 1998.
21. R. Schmid, "Infinite dimensional Lie groups with applications in mathematical physics," *Journal of Geometry and Symmetry in Physics*, vol. 1, pp. 1–67, 2004.
22. S. Joshi, B. Davis, M. Jomier, and G. Gerig, "Unbiased diffeomorphic atlas construction for computational anatomy," *Neuroimage*, vol. Suppl. 1, pp. S151–S160, September 2004.
23. B. Avants and J.C. Gee, "Geodesic estimation for large deformation anatomical shape and intensity averaging," *Neuroimage*, vol. Suppl. 1, pp. S139–150, 2004.
24. H. J. Johnson and G. E. Christensen, "Consistent landmark and intensity-based image registration," *IEEE Trans. Med. Imaging*, vol. 21, no. 5, pp. 450–461, 2002.
25. B. Avants, C. L. Epstein, and J. C. Gee, "Geodesic image interpolation: Parameterizing and interpolating spatiotemporal images," in *ICCV Workshop on Variational and Level Set Methods*, 2005, pp. 247–258.
26. B. Eckmann, "Social choice and topology," lecture notes, 2003.
27. J. Thirion, "Image matching as a diffusion process: an analogy with maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.
28. G. Christensen and H. Johnson, "Consistent image registration," *IEEE Transactions on Medical Imaging*, vol. 20, no. 7, pp. 568–582, 2001.
29. V. I. Arnold and B. A. Khesin, "Topological methods in hydrodynamics," *Ann. Rev. Fluid Mech.*, vol. 24, pp. 145–166, 1992.

Statistics on Anatomic Objects Reflecting Inter-Object Relations

Ja-Yeon Jeong, Stephen M.Pizer, and Surajit Ray

Medical Image Display & Analysis Group (MIDAG)
University of North Carolina, Chapel Hill NC 27599, USA

Abstract. Describing the probability densities of multi-object complexes by describing individual objects and their inter-object relationships leads to desirable locality without ignoring the context of an object. We describe a means of decomposing object variations into self effects and neighbor effects. We describe an approach for estimating the self and neighbor effect probability densities for each object in the complex using augmentation and prediction, supported by PGA on m-reps. We apply this method to the inter-day variation of m-reps of male pelvic organs within an individual patient.

1 Introduction

Statistical shape models have been proven to be very effective in a number of applications, including image segmentation [1] and characterization of the anatomic differences between the classes of normal and diseased patients [2]. In segmentation, prior shape statistics restrict the deformation of the shape model within the variations learned from training data in this optimization process. In characterization of anatomic differences these statistics provide the basis for a test of a null hypothesis that the probability densities for the two classes are the same. Obtaining accurate shape statistics is thus essential.

Shape statistics are likely to be more sensitive measures when multiple objects in a given anatomic region are considered since frequently the intensity information does not fully provide the boundary of a target object without considering the neighbors that provide its context.

Following the approach that others have taken, we have calculated both global statistics on combined shape models of the multiple objects and statistics on each object separately. We found that the former approach fails to capture the local variation of an object itself and sometimes gives misleading information about inter-relation between objects. The latter approach gives too local information while ignoring the interaction between objects. The weakness of both approaches has led us to borrow the idea of the mixed model [3] approach to handle the inter-object relation. In this work we decompose the shape variation of each of the objects into three components: the mean of the variation from some base state, self effects, and neighbor effects. The neighbor effects term are described as a function of neighbors' geometric descriptors since shape variations of an object closely surrounded by other objects are caused not only by

internal changes within the object but also by its neighboring objects. The self term describes the variation of the object itself, not affected by its neighboring objects. We present in this paper an approach to estimate probability densities on each of these components individually.

We use a single-figure m-rep model [1] for each object as the shape representation. M-reps consist of sheets of medial atoms. Medial atoms capture not only local position but also the local orientation and local magnification of a section of an object. M-reps also provide correspondences across cases of the object derived from the local coordinate system for object interiors that m-rep provides. Therefore, m-reps are powerful representation to characterize the neighbor relationship, allowing medial atoms in neighboring objects to be understood in terms of atom transformations of each other.

For both the self and neighbor terms we estimate probability densities on the nonlinear manifold by a method called principal geodesic analysis (PGA) developed in our previous work [4]. The approach for the neighbor term is an extension of our earlier work [5], [6] in which an augmentation and prediction method is introduced to estimate the relation among multiple objects.

Section 2 describes other approaches in estimating multi-object geometric statistics and basic ideas of residues, augmentation, and prediction. The difference and addition operations through which residues are obtained are also explained in that section. Section 3 presents our new iterative method to estimate self and neighbor effects of multi-objects with mathematical detail. Section 4 explains the process to estimate these probability densities for the male-pelvis data to which we applied our new method, and it gives the results of the estimation. Section 5 discusses the results as well as the work yet to be done.

2 Background

Multi-object shape statistics in deformable template models has been mostly dealt with by doing global statistics on all objects taken together. This approach has been applied on a variety of shape representations: point distribution models [7], diffeomorphisms from atlases [8], distance functions or their levels sets [9], and m-reps [1]. This global statistics lacks appropriate locality of the objects.

Other approaches in doing shape statistics have handled locality, addressing the issue of scale. They compute shape statistics hierarchically from the object complex to the individual objects [10], [11], and one of these even analyzes residues from a larger scale [12]. However, few attempts have been made to describe the inter-object relation statistically. Pohl et al. [13] describes inter-relationships by representing via the distance function to objects' boundaries. This approach however does not explicitly separate out inherent variation of object from the effects from its neighbor.

In this work we do not consider a global stage since our target problem appears to have only single object effects and inter-object effects. However, we do decompose the objects' variation into self and neighbor effects, which must sum to the overall difference of the object from its base state. Thus the self

effect must be the residue describing overall change from the base state after the neighbor effect is removed, and the neighbor effect must be the residue describing overall change from the base state after the self effect is removed. This explicit separation between self and neighbor variation is the main improvement that we have made on our earlier method [5]. We rely on the difference operation $\ominus$ on m-reps to obtain the residue. The difference operation and its complementary addition operation are described briefly in the next.

A medial atom $\underline{\mathbf{m}} = (\mathbf{x},\ r,\ \mathbf{u}, \mathbf{v})$ is defined as an element of the symmetric space $G = R^3 \times R^+ \times S^2 \times S^2$ where the hub position $\mathbf{x} \in R^3$, the spoke length $r \in R^+$, and the two unit spoke directions $\mathbf{u}, \mathbf{v} \in$ the unit sphere S^2. Let $\mathbf{R}_{\mathbf{w}}$ represent the rotation along the geodesics in S^2 that moves a point $\mathbf{w} \in S^2$ to the north pole $\mathbf{p} = (0, 0, 1) \in S^2$. For given any two medial atoms $\underline{\mathbf{m}}_1, \underline{\mathbf{m}}_2 \in G$ where $\underline{\mathbf{m}}_i = (\mathbf{x}_i, r_i, \mathbf{u}_i, \mathbf{v}_i), i = 1, 2$, the difference between them can be described as follows:

$$\underline{\mathbf{m}}_1 \ominus \underline{\mathbf{m}}_2 := (\mathbf{x}_1 - \mathbf{x}_2,\ \tfrac{r_1}{r_2},\ \mathbf{R}_{\mathbf{u}_2}(\mathbf{u}_1),\ \mathbf{R}_{\mathbf{v}_2}(\mathbf{v}_1))\ . \tag{1}$$

Its corresponding addition operator $\oplus$ is thus defined as

$$\underline{\mathbf{m}} \oplus \Delta\underline{\mathbf{m}} := (\mathbf{x} + \Delta\mathbf{x},\ r \cdot \Delta r,\ \mathbf{R}_{\mathbf{u}}^{-1}(\Delta\mathbf{u}),\ \mathbf{R}_{\mathbf{v}}^{-1}(\Delta\mathbf{v})) \tag{2}$$

for a given $\underline{\mathbf{m}} = (\mathbf{x},\ r,\ \mathbf{u}, \mathbf{v})$ and difference $\Delta\underline{\mathbf{m}} = (\Delta\mathbf{x},\ \Delta r,\ \Delta\mathbf{u}, \Delta\mathbf{v})$. For two m-reps $\underline{\mathbf{M}}_1, \underline{\mathbf{M}}_2$ that consist of medial atoms, the difference $\underline{\mathbf{M}}_1 \ominus \underline{\mathbf{M}}_2$ relative to $\underline{\mathbf{M}}_2$ coordinate is thus defined as the collection of the differences of corresponding individual atoms in two m-reps. These operations and their properties are explained in detail in [6].

We also extend the augmentation and prediction methods described in [5]. We make use of augmentation to deal with a target object's inter-relation with other objects. Based on the evidence that atoms in a target object are highly correlated with abutting atoms in the neighboring objects, we choose a set of atoms $\underline{\mathbf{A}}$ in the target object that are located near its neighboring objects and a set of atoms $\underline{\mathbf{N}}$ in neighboring objects that are close to $\underline{\mathbf{A}}$. We then produce the augmented set of atoms by putting atoms in $\underline{\mathbf{N}}$ and $\underline{\mathbf{A}}$ together into one set of atoms $\underline{\mathbf{U}}$. This augmentation allows us predict the changes brought on a target object by the change of the neighboring objects, namely as the mean of its $\underline{\mathbf{A}}$ conditioned on its neighbor atoms $\underline{\mathbf{N}}$. Then we can statistically analyze variations from this prediction. This prediction method helps to extract the deterministic effect from its neighbor on the object and concentrate on the variable part in the neighbor effects. A major improvement on our earlier method is that we consider the neighbor relation mutually, allowing each object to have any of the others as neighbors. This mutual neighbor relation is more realistic and is clearly suggested in the male-pelvis data to which we apply our new method.

3 Method in detail

Let $\Delta\underline{\mathbf{M}}^i := \underline{\mathbf{M}}^i \ominus \underline{\mathbf{M}}^b$ where $\underline{\mathbf{M}}^b$ is a base model such as the mean or some reference model, where i indexes the training cases and $\underline{\mathbf{M}}$ represents an m-rep.

All figures in $\underline{\mathbf{M}}^i$ are aligned to figures in the base model. $\{\Delta\underline{\mathbf{M}}^i\}$ are variations of m-reps from the base m-rep model. We decompose the variations into two parts, *self* and *neighbor*, as follows:

$$\Delta\underline{\mathbf{M}} := \Delta\underline{\mathbf{M}}^{self} \oplus \Delta\underline{\mathbf{M}}^{ngbr} \tag{3}$$

where $\Delta\underline{\mathbf{M}}^{self}$ measures self variation and $\Delta\underline{\mathbf{M}}^{ngbr}$ measures neighbor effects. $\Delta\underline{\mathbf{M}}^{ngbr}$ is further subdivided as the prediction from $\underline{\mathbf{N}}$ $Pred(\Delta\underline{\mathbf{M}}^{ngbr})$, and the neighbor residue, $\Delta\underline{\mathbf{M}}^{ngbr} \ominus Pred(\Delta\underline{\mathbf{M}}^{ngbr})$.

We make two assumptions in our approach. First, within each object, the self variations $\{\Delta\underline{\mathbf{M}}^{self}\}$ and the residues from the predictions $\{\Delta\underline{\mathbf{M}}^{ngbr} \ominus Pred(\Delta\underline{\mathbf{M}}^{ngbr})\}$ are considered to be statistically uncorrelated. Second, we assume that the effect of the neighboring objects is local. Based on the second assumption we define the sets $\underline{\mathbf{N}}$, $\underline{\mathbf{A}}$, and hence $\underline{\mathbf{U}}$. We currently choose atoms in the two sets based on Euclidean distance between atoms in nearby objects. $\Delta\underline{\mathbf{N}}$, $\Delta\underline{\mathbf{A}}$, and $\Delta\underline{\mathbf{U}}$ denote the m-rep variations in the sets $\underline{\mathbf{N}}$, $\underline{\mathbf{A}}$, and $\underline{\mathbf{U}}$ respectively.

3.1 Iterative steps

In calculating the geometric statistics, we begin with a simple assumption on the separation of each object's $\Delta\underline{\mathbf{M}}$ into self and neighbor components and then refine that separation by repeating the following steps over all figures in the multi-object m-reps.

In the following description of the estimation of the statistics for an object, the index over the objects and the index over the training cases is skipped.

Self step. We do PGA on $\Delta\underline{\mathbf{M}} \ominus \widehat{\Delta\underline{\mathbf{M}}}^{ngbr}$ which gives the shape space and estimate of the self part of each training case $\widehat{\Delta\underline{\mathbf{M}}}^{self}$. The hat $\widehat{(\cdot)}$ indicates the best estimate of either neighbor or self components up to the previous iteration.

Neighbor step.

1) Augmentation. We first subtract the estimate of the self part from each training case because we do not want to corrupt the effect from the neighbor by the effect to other neighboring objects from the object, i.e.,

$$\Delta\underline{\mathbf{A}}^{ngbr} := \Delta\underline{\mathbf{A}} \ominus \widehat{\Delta\underline{\mathbf{A}}}^{self}. \tag{4}$$

Then, because we need to predict $\Delta\underline{\mathbf{A}}^{ngbr}$ based on $\widehat{\Delta\underline{\mathbf{N}}}^{self}$, we form an augmented set of differences:

$$\Delta\underline{\mathbf{U}} := \Delta\underline{\mathbf{A}}^{ngbr} \cup \widehat{\Delta\underline{\mathbf{N}}}^{self}. \tag{5}$$

We use $\Delta\underline{\mathbf{N}}^{self}$ rather than $\Delta\underline{\mathbf{N}}$ because our initial assumption of local effect leads to $\Delta\underline{\mathbf{A}}^{ngbr}$ of the object and $\Delta\underline{\mathbf{N}}^{ngbr}$ of its neighbors being statistically independent.

2) Prediction. (1) Predictor function: We perform PGA on $\Delta\underline{\mathbf{U}}$ to find the shape space of the augmented atoms $\Delta\underline{\mathbf{U}}$. The shape space is used to find the deterministic effect from the neighbors as follows.

$$Proj(\Delta\underline{\mathbf{A}}^{ngbr}) = \exp_\mu\left(\sum_{l=1}\langle\log_\mu(\widehat{\Delta\underline{\mathbf{N}}}^{self}), v_l\rangle \cdot v_l\right), \tag{6}$$

$$Pred(\Delta\underline{\mathbf{A}}^{ngbr}) := Proj(\Delta\underline{\mathbf{A}}^{ngbr}) \tag{7}$$

where $\{v_l\}_{l=1}$ are principal directions in the tangent space of $\Delta\underline{\mathbf{U}}$ at its mean μ. Note that $\widehat{\Delta\underline{\mathbf{N}}}^{self}$ means $\mu|_{\underline{\mathbf{A}}} \cup \widehat{\Delta\underline{\mathbf{N}}}^{self}$ implicitly when $\log_\mu(\cdot)$ is applied.[1]

(2) Updated augmented set: We now form a newly updated augmented set after removing the prediction from the residue from the self estimate.

$$\Delta\underline{\mathbf{A}}^{rmdr} = \Delta\underline{\mathbf{A}}^{ngbr} \ominus Pred(\Delta\underline{\mathbf{A}}^{ngbr}), \tag{8}$$

$$\Delta\underline{\mathbf{U}}' := \Delta\underline{\mathbf{A}}^{rmdr} \cup \widehat{\Delta\underline{\mathbf{N}}}^{self}. \tag{9}$$

We then do PGA on the new augmented set $\Delta\underline{\mathbf{U}}'$ to obtain the shape space and the estimate $\widehat{\Delta\underline{\mathbf{A}}}^{ngbr}$ of the neighbor part of each training case. As a result, the estimate of the neighbor part comprises the two components: prediction and the estimate of the variation from the prediction.

$$\widehat{\Delta\underline{\mathbf{A}}}^{ngbr} = Pred(\Delta\underline{\mathbf{A}}^{ngbr}) \oplus \exp_\mu\left(\sum_{k=1}\langle\log_\mu(\Delta\underline{\mathbf{A}}^{rmdr}), v_k\rangle \cdot v_k\right). \tag{10}$$

3.2 Joint probability of interaction among objects

Using the joint probability on this decomposition of self and neighbor effects of multiple objects, we can interpret the prediction as the conditional mean assuming a Gaussian probability distribution. This interpretation is valid as long as we can show that the following conditions hold: the self effect and neighbor effect within each object, neighbor effects among objects, and self effects among objects are independent.

To show this, we decompose the joint probability of multiple objects using conditional probability:

$$p(\{\underline{\mathbf{M}}_k\}) = p(\underline{\mathbf{M}}_k \mid \{\underline{\mathbf{M}}_j | j \neq k\})\, p(\{\underline{\mathbf{M}}_j | j \neq k\}), \tag{11}$$

where k goes over the number of objects. Moreover,

$$p(\underline{\mathbf{M}}_k \mid \{\underline{\mathbf{M}}_j | j \neq k\}) = p(\underline{\mathbf{M}}_k^{self}, \underline{\mathbf{M}}_k^{ngbr} \mid \{\underline{\mathbf{M}}_j^{self}, \underline{\mathbf{M}}_j^{ngbr} | j \neq k\}). \tag{12}$$

[1] Refer to [4] for detailed explanation of the log map and the exponential map.

Since we have assumed the locality of the effect of the neighboring objects and we choose $\underline{\mathbf{N}}$ as a set of medial atoms in neighboring objects that has influence on the variation of the shape of an object, we can replace the set of atoms in $\{\underline{\mathbf{M}}_j^{self}, \underline{\mathbf{M}}_j^{ngbr} | j \neq k\}$ with $\underline{\mathbf{N}}$. If the independence conditions stated above hold, then we can further simplify the joint probability distribution as follows:

$$p(\underline{\mathbf{M}}_k^{self}, \underline{\mathbf{M}}_k^{ngbr} \mid \underline{\mathbf{N}}_k^{self}) = p(\underline{\mathbf{M}}_k^{self})\, p(\underline{\mathbf{M}}_k^{ngbr} \mid \underline{\mathbf{N}}_k^{self}). \tag{13}$$

4 Application on male-pelvis model and results

4.1 Materials

The training models were obtained from male-pelvis CT images of real patients taken over a series of days who underwent radiotherapy treatment. Three organs, namely the bladder, prostate, and a section of the rectum that is adjacent to the prostate are modelled. Both ends of the rectum model are arbitrary. A clinician contoured each organ slice by slice to generate binary images for all three organs. A single-figure m-rep was then fit to bladder, prostate, and rectum separately in binary images: 5×6 grids of medial atoms for the bladder, 7×4 grids of medial atoms for the prostate, and 15×3 grids of medial atoms for the rectum are used. We have software developed to fit a single figure m-rep to binary image that prevents penetration among fitted m-reps, prevents folding of the interior of the object represented, and maintains regularity of grid across the cases for correspondence of medial atoms [14]. We aligned the fitted three m-reps for bladder, prostate, and rectum of each patient by a similarity transformation that is computed from two landmarks, at the apex and base of the prostate landmark. Then those aligned m-reps for bladder, prostate, and rectum are combined into one ensemble m-rep. The total number of medial atoms are 103, and the number of parameters in the ensemble m-rep is 927.

4.2 Application of probability density estimation to male-pelvis model

We have applied our approach to male-pelvis models of five patients m-rep fits of which were obtained as described in the previous section. Patients are numbered as 3101, 3106, 3108, 3109, and B163. We have 14 m-rep fits for 3101, 17 for 3106, 18 for 3108, 18 for 3109, and 15 for B163. Models fitted to the first treatment image are used as the reference model from which the variation of the rest of models are taken.

In the first iteration, since we know that the self part of the bladder and rectum changes dominate the neighbor parts, we assume that the neighbor parts in variations of the bladder and rectum are zero. Similarly, we set the self part in the prostate to zero, for the shape of prostate changes little except as affected by the bladder and rectum.

Starting with $\Delta\underline{\mathbf{M}}_b$, $\Delta\underline{\mathbf{U}}_b$, $\Delta\underline{\mathbf{M}}_p$, $\Delta\underline{\mathbf{U}}_p$, $\Delta\underline{\mathbf{M}}_r$, $\Delta\underline{\mathbf{U}}_r$ where subscript b, p, r represent bladder, prostate, and rectum respectively, the order we compute

	total variation per object					
patient no.	bladder self	bladder neighbor	rectum self	rectum neighbor	prostate neighbor	prostate self
3101	0.379383	0.014782	0.360542	0.004983	0.179949	0.006880
3106	0.082828	0.002576	0.102897	0.001188	0.031047	0.003648
3108	0.391689	0.010346	0.181145	0.012492	0.135748	0.009021
3109	0.080293	0.006388	0.132851	0.003501	0.043438	0.005439
B163	0.067945	0.002470	0.137690	0.002819	0.053831	0.003205

Table 1. Total variations of two effects, self, and neighbor per organ that are estimated after 2nd iteration

the estimate of neighbor and self effect in each organ in the first iteration is as follows.

1. Self effect on bladder, and rectum differences $\Delta\underline{\mathbf{M}}_b$, $\Delta\underline{\mathbf{M}}_r$, assuming $\Delta\underline{\mathbf{M}}_b^{ngbr}$, $\Delta\underline{\mathbf{M}}_r^{ngbr}$ are zero
2. Neighbor effect on prostate differences $\Delta\underline{\mathbf{U}}_p$, assuming $\Delta\underline{\mathbf{A}}_p^{self}$ are zero. Note that we use for prediction the estimate of the self variations of bladder and rectum from step 1.
3. Self effect on prostate residue from the estimate of the neighbor effect. The residue is $\Delta\underline{\mathbf{M}}_p^{self} := \Delta\underline{\mathbf{M}}_p \ominus \widehat{\Delta\underline{\mathbf{M}}}_p^{ngbr}$.
4. Neighbor effect on bladder and rectum residue from the estimate of the self effect $\Delta\underline{\mathbf{U}}_b$, $\Delta\underline{\mathbf{U}}_r$ where $\Delta\underline{\mathbf{A}}_b := \Delta\underline{\mathbf{A}}_b \ominus \widehat{\Delta\underline{\mathbf{A}}}_b^{self}$ and $\Delta\underline{\mathbf{A}}_r := \Delta\underline{\mathbf{A}}_r \ominus \widehat{\Delta\underline{\mathbf{A}}}_r^{self}$. We again use the estimate of the self part of the prostate $\widehat{\Delta\underline{\mathbf{M}}}_p^{self}$ from step 3 to compute the prediction of bladder and rectum from change of their neighboring object prostate.

For the later iterations, we repeat the steps described in section 3.1 with the updated estimates of self and neighbor effects from the previous iteration.

4.3 Result

Table 1 compares the estimates of total variations of the two effects for each organ after the 2nd iteration. They are the sum of eigenvalues that are estimated in each step described in the previous section. Figure 1 shows the primary mode of self variation of the bladder and the associated prediction of the deformation of the prostate in the patient B163. We can see a dent formed in the bladder in which the prostate fits as the bladder fills.

5 Discussion and conclusion

The results are consistent with what we know about the anatomy and see in the data. The self terms for the bladder and rectum dominate the neighbor terms,

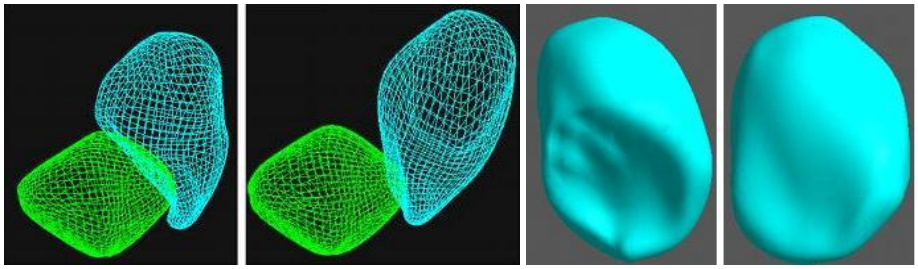

Fig. 1. These 4 figures show the prediction of the deformation on the prostate brought by the change of the patient B163. The two objects in the left 2 panels are the bladder (cyan) and the prostate (green) in wire frame and the right 2 panels show separately the bladder of the left 2 panels in solid. The bladder in the right 2 panels is in 2 standard deviations from the mean for a principal modes of the self effect estimated in the bladder. The left 2 panels show the predicted prostate corresponding to the two positions of the bladder shown in the right 2 panels.

reflecting the fact that these organs' variations are principally due to changes in their contents. On the other hand, the neighbor term for the prostate dominates the self term, reflecting the fact that the prostate is a rather hard organ that is pushed upon by the rectum and bladder.

Moreover, the predictions of the prostate from the bladder and the rectum seem realistic. Also, the prostate changes from its prediction are smaller than the self changes in the bladder and rectum, which are known to be larger.

The bladder self changes include modes corresponding to lengthening, widening, and lapping around the prostate, all anatomically observed processes. Also encouraging is that the prostate predictions of the bladder self modes have the prostate following the change in the indentations of the bladder while keeping the prostate almost entirely nonintersecting. Furthermore, the prostate self modes also make consistent predictions on the bladder indentation.

While this experiment is limited, it suggests that our new approach to separate out inherent variation of an object itself and effects from its neighboring objects is fruitful.

We still need to do further analysis to verify that our estimates truly reflects the self and neighbor effects of multi-objects. We plan to determine what are self and neighbor effects by simulating the obvious neighbor effects and self effects independent from neighbor effects on an ensemble of multiple ellipsoids m-reps and apply our approach on the the simulated ensembles.

We will also incorporate the new geometric statistics of the male pelvic organs as the prior within our segmentation method's posterior objective function and evaluate this prior by the effectiveness of the segmentation of these organs.

An open issue is how to choose augmented atoms. Clearly, that choice affects our estimate of self and neighbor effects. We believe that we should choose the most highly correlated atoms with those in the target object. In separate research

we are measuring this correlation, and the results of that research will provide us a firmer basis for our choice than the distance criterion presently used.

In the nonlinear manifold on which m-reps are situated, addition operations is not commutative. The separation of $\Delta\underline{\mathbf{M}}$ into $\Delta\underline{\mathbf{M}}^{self} \oplus \Delta\underline{\mathbf{M}}^{ngbr}$ is thus not equivalent to $\Delta\underline{\mathbf{M}}^{ngbr} \oplus \Delta\underline{\mathbf{M}}^{self}$. In our iterative algorithm, this noncommutativity is ignored and $\Delta\underline{\mathbf{M}}^{self}$, $\Delta\underline{\mathbf{M}}^{ngbr}$ are treated as if they are interchangeable. We must test whether the effect of our method's assumption of commutativity is significant.

We also need to test the four conditions stated in 3.2 to show our interpretation on joint probability holds.

6 Acknowledgements

We are grateful to Keith Muller, J. Stephen Marron and Sarang Joshi for scientific discussions, to Edward Chaney and Gregg Tracton for data, manual segmentations, and alignments, and to Delphine Bull for help in preparing this paper. The work described in this paper was done under the partial support of NIH grant P01 EB02779.

References

1. Pizer, S., Fletcher, T., Fridman, Y., Fritsch, D., Gash, A., Glotzer, J., Joshi, S., Thall, A., Tracton, G., Yushkevich, P., Chaney, E.: Deformable m-reps for 3d medical image segmentation. International Journal of Computer Vision - Special UNC-MIDAG issue, (O. Faugeras, K. Ikeuchi, and J. Ponce, eds.) **55**(2) (2003) 85–106
2. Pilgram, R., Fletcher, P., Pizer, S., Pachinger, O., Schubert, R.: Common shape model and inter-individual variations of the heart using medial representation. submitted to IEEE TMI (2004)
3. Ray, S., Simpson, S., Muller, K.: (Structural covariance for multi-object image segmentation)
4. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging **23**(8) (2004) 995–1005
5. Pizer, S., Jeong, J.Y., Lu, C., Muller, K., Joshi, S.: Estimating the statistics of multi-object anatomic geometry using inter-object relationships. In: in International Workshop on Deep Structure, Singularities and Computer Vision (DSSCV), (O.F. Olsen, L. Florack, and A. Kuijper, eds.). (2005) 60–71
6. Lu, C., Pizer, S., Joshi, S., Jeong, J.Y.: Statistical multi-object shape models. Technical report, UNC Chapel-Hill (2005)
7. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models their training and application. Computer Vision and Image Understanding **61**(1) (1995) 38–59
8. Joshi, S.: Large Deformation Diffeomorphisms and Gaussian Random Fields for Statistical Characterization of Brain SubManifolds. PhD thesis, Dept. of Electrical Engineering, Sever Institute of Technology, Washington Univ. (1997)
9. Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, E., Willsky, A.: A shape-based approach to curve evolution for segmentation of medical imagery. IEEE Transactions on Medical Imaging **22**(2) (2003) 137–153

10. Vaillant, M., Davatizikos, C.: Hierarchical matching of cortical features for deformable brain image registration. In: Information Processing in Medical Imaging (IPMI). (1999) 182–195
11. Kapur, T., Beardsley, P., Gibson, S., Grimson, W., Wells, W.: Model based segmentation of clinical knee mri. In: Model-based 3D Image Analysis workshop (in conjuction with ICCV). (1998) 156–162
12. Davatzikos, C., Xiaodong, T., Shen, D.: Hierarchical active shape models, using the wavelet transform. IEEE Transactions on Medical Imaging **22**(3) (2003) 414–423
13. Pohl, K., Fisher, J., Levitt, J., Shenton, M., Kikinis, R., Grimson, W., Wells, W.: A unifying approach to registration, segmentation, and intensity correction. In: Medical Image Computing and Computer Assisted-Intervention. (2005) 310–318
14. Merck, D., Tracton, G., Saboo, R., Chaney, E., Pizer, S., Joshi, S.: A methodology and implementation for constructing geometric priors for deformable shape models (2006) in Conference submission.

Topological Repair on Voxel-Based Quadrangular Meshes

Paulette Lieby[1], Nick Barnes[1], and Brendan D. McKay[2]

[1] Vision Science, Technology and Applications, Canberra, National ICT Australia,
Paulette.Lieby@anu.edu.au, Nick.Barnes@nicta.com.au
[2] Department of Computer Science, Australian National University,
bdm@cs.anu.edu.au

Abstract. In some neuroscience applications it is critical that the representation of anatomical structures is topologically faithful. This is especially important when the topology is known to be equivalent to a sphere. We propose a new approach to repair a voxel-based quadrangular mesh so that it becomes topologically equivalent to a sphere. The approach is graph-based and results in practice in an amount of change that is generally minimal. The algorithm was successfully applied to a database of 977 hand-segmented hippocampi from a study of ageing; spherical topology was achieved for all data sets.

1 Introduction

In medicine and neuroscience, the faithful recovery of anatomical structures from 3D voxel-based image data is an important goal. Some of these structures are known to be topologically equivalent to a sphere, that is, they can be mapped onto a sphere by continuous deformation. This is the case for the cortical surface [1, 2], the hippocampus [3, 4], and the lateral ventricle [5].

A spherical topology guarantees the existence of an invertible one-to-one map between such structures. This facilitates analysis between subjects using surface warping [6] and it simplifies registration. However, artifacts may be introduced in the recovery of these structures by hand-tracing or automatic recovery processes. In the case of both hippocampus and cortical surface, for normal neuroanatomy, artifacts such as handles and isolated voxels do not reflect the true neuroanatomical structure. This may cause problems for some applications such as warping, registration, and some forms of mathematical modelling (e.g., spherical harmonic mapping [3, 7, 4, 5]). In this case, it is important for the voxel-based data to reflect the spherical topology.

In this paper we present an algorithm for topological repair, focusing on the detection of handles and how to repair them. The traditional approach to topological repair is human intervention (e.g. [8]). Common recent approaches involve isosurface reconstruction with topological control using fast marching methods ([9, 10]), watersheds ([11]), morphological operators ([12]). Others focus on repairing the surface tessellation itself ([13]). Yet other approaches are graph-based and they also operate directly on the digital volume ([1, 2]). The aim of

all these methods (except [13]) is to repair the volume so that after applying a topologically consistent isosurface algorithm, Marching Cubes tessellation for example, one obtains a triangulation with spherical topology.

However, triangulation is less suitable if further analysis is required, such as shape analysis, or if the aim is to generate a model where we require precise correspondence with the original grey-level data (e.g. [3]). Correspondences are trivially preserved by a voxel-based mesh, but this is not obvious in the case of Marching Cubes tessellation. Further, a voxel-based mesh may be uniformly mapped on the sphere, thus enabling the computation of the shape's spherical harmonics ([7]). Finally, hand-tracings are generally performed as precisely as possible, so it is preferable to stay close to the hand-segmented data as a general principle.

Other approaches have also corrected the topology within a voxel-based representation, for example [14]. However in this case it is not clear if defects can be detected automatically. A method described by [15] applies smoothing and level-set filters on the surface. However, such operations are not guaranteed to find all handles, and may change a large number of voxels.

Our approach is graph-based, operates on the digital volume and produces a voxel-based quadrangulation as the resulting surface mesh. We apply a fundamental result from algebraic topology to detect handles directly in the surface mesh. We are then able to correct all well-behaved small handles, usually in a minimal way.

The paper is divided as follows. The next section outlines the major steps of our approach while Sections 3 to 6 explore each step in detail. In Section 7 we discuss the algorithm's performance as applied to a database of 977 hand-traced hippocampi and in Section 8 we examine the strengths and shortcomings of our approach.

2 Approach

Three-dimensional segmented digital data is a binary image that distinguishes the *interior* (foreground) voxels from the *exterior* (background) voxels. The *surface mesh* consists of the voxel corners, edges, and faces lying on the boundary between interior and exterior. A voxel corner, edge, or face on the surface mesh is called a *vertex*, an *edge* or a *face* respectively.

A vertex v is *adjacent* to vertex u if there is an edge between u and v. We say that v is a *neighbour* of u and conversely. The surface mesh is a description of the surface which, for each vertex v, lists the neighbours of v in an orderly way (see Section 4). Since every face is bordered by four edges, the surface mesh is said to be a *quadrangulation*.

Our method to produce a topologically correct surface from the hand-segmented data can be described as follows:

1. At each vertex on the surface, apply a correction filter on the original binary volume data to ensure that the surface at this vertex is non-self-intersecting (Figure 1(a)).

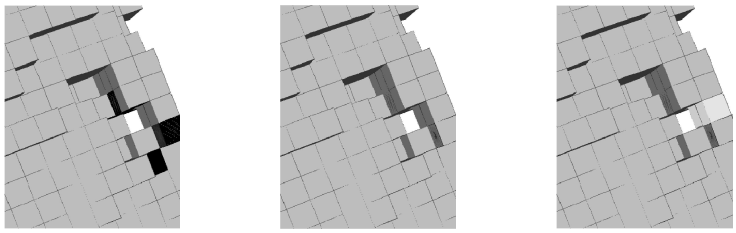

Fig. 1. A partial view of a surface mesh for a hand-segmented hippocampus. (a) After application of the correction filter, the dark voxels have been added; (b) The surface mesh; (c) Topological repair: the light voxels are marked for removal to open the handle.

2. From the non-self-intersecting surface, construct the surface mesh (Figure 1(b)).
3. Use the surface mesh to locate any handles.
4. Convert the handle into sets of voxels to be either removed or added.
5. Repair the topology by either filling the tunnel or opening the handle (Figure 1(c)).

Each of these points is addressed in the subsequent sections.

3 The Correction Filter

In order for a surface to be topologically equivalent to a sphere it must be non-self-intersecting. This is equivalent to saying that it is locally homeomorphic to a disk. Figure 2 shows the three fundamental obstructions for the surface to being locally homeomorphic to a disk. See [16] for a proof.

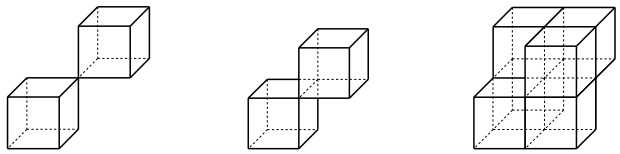

Fig. 2. The forbidden configurations

At each vertex, we determine if any configuration in Figure 2 is present; this is achieved by a simple case by case examination. The forbidden configuration is removed by either adding and removing the minimum number of voxels that renders the surface non-self-intersecting at this vertex.

This correction filter is applied to each vertex in order, where the vertices are ordered lexicographically on their coordinates. While applying the filter at some

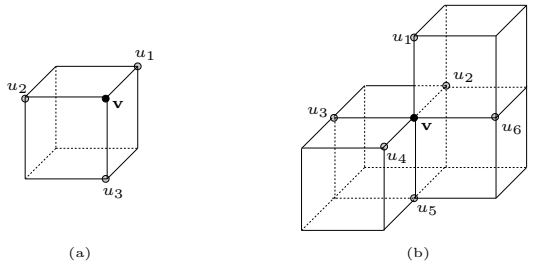

Fig. 3. Neighbours of v in anti-clockwise order. (a): u_1, u_2, u_3; (b): $u_1, u_2, u_3, u_4, u_5, u_6$

vertex v, one has to be careful not to introduce new forbidden configurations which did not originally exist. This is achieved by recursively examining all vertices whose neighbourhood has been affected and who are less than or equal to v (in the sense of the order described above). It is easy to see that this process must terminate.

It is possible that the application of the filter results in the formation of topological defects; they will be corrected as explained in Sections 5 and 6.

Note also that the application of the correction filter to all vertices of the hand-segmented data results in a *6-connected* 3D digital image. A digital image is said to be 6-connected when two voxels are neighbours if and only if they share a common face.

4 Mesh Creation

Once the surface is non-self-intersecting we construct the surface mesh consisting of the surface vertices and edges. As mentioned in Section 2 the list of neighbours of a given vertex is given in an *orderly* way.

Choose a direction, say, anti-clockwise. Given a vertex v, the ordered list of neighbours of v is $u_1, u_2, \ldots, u_n$ where the sequence of vertices is given by the direction of traversal of the neighbours of v. See Figure 3 for an example. Note that the starting vertex of the list is irrelevant.

To construct the surface mesh, for each vertex v on the surface, we perform an anti-clockwise walk starting at v and visiting all the neighbours of v. Two steps are necessary to ensure that we traverse the list of neighbours of v in a consistent fashion (e.g. always anti-clockwise). First we assign a normal to each face in a consistent manner (e.g. pointing towards the *exterior* of the object). Secondly, we ensure that the cross-products of the vectors that describe the walk have the same direction relative to the assigned face normal.

Let u_1 and u_2 be neighbours of v such that the path of vertices v, u_1, v_0, u_2 describes a anti-clockwise walk around the face f incident to v, u_1 and u_2 (see Figure 4 (a) for an example). The vertex v_0 is the vertex of this face which is not adjacent to v, i.e. there is no edge between v and v_0. Then the cross-products

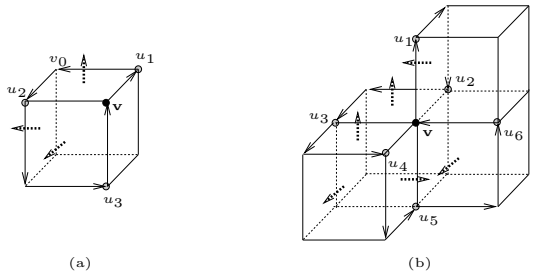

Fig. 4. Anti-clockwise walk around vertex v starting at u_1. Normals are indicated by the thick dashed arrows.

$\overrightarrow{(u_1 - v)} \times \overrightarrow{(v_0 - u_1)}$, $\overrightarrow{(v_0 - u_1)} \times \overrightarrow{(u_2 - v_0)}$, and $\overrightarrow{(u_2 - v_0)} \times \overrightarrow{(v - u_2)}$ all have the same direction as the normal of the face f. This is depicted in Figure 4.

5 Locating Handles

At this point, the surface of the object is fully described by its surface mesh, enumerating the vertices and the sequence of their neighbours. We may always assume that the mesh is connected. The aim is to obtain a mesh which is topologically equivalent to a sphere.

Given a connected surface mesh with n vertices, m edges and l faces, the *genus* g of the surface is given by Euler's formula ([17])

$$g = \frac{1}{2}(2 - n + m - l).$$

The genus measures the number of *handles* of the surface. For example, a sphere has no handles, while a torus has one handle. We name the hole surrounded by a handle a *tunnel*, see Figure 5(c).

A fundamental result of algebraic topology ([17]) states that there is a handle in the surface mesh *if and only if*

1. there are two cycles A and B in the mesh such that A and B intersect at v;
2. and, walking along B to cross A at v results in moving from the "left" to the "right" of A without encountering any other vertex of A ("left" and "right" of A are relative to an arbitrary direction assigned to A).

This is shown in Figure 5.

Let us call cycles that have the property pictured in Figure 5 (b) and (c) *interlocking cycles* or *interlocking cycles at* v. We also say that B is the *transverse* of A and conversely.

We locate the handles by finding interlocking cycles in the surface mesh. In order to achieve a minimum amount of change we must find the smallest interlocking cycles. This is easily done by performing a breath-first-search on

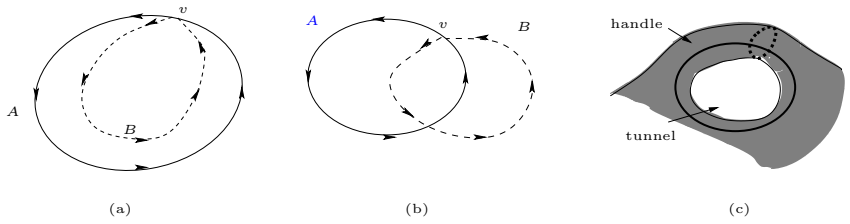

Fig. 5. In (*a*) A and B are intersecting and B is entirely on one "side" of A; this is not a handle. In (*b*) and (*c*) A and B are intersecting but B starts and ends at opposite sides of A; this is a handle.

the mesh. The fact that the neighbours of each vertex are given in an orderly fashion (Section 4) enables us to determine if cycles are interlocking.

6 Converting Handles to Sets of Voxels

Once interlocking cycles are found they are converted into sets of voxels that describe the handle to be opened or the tunnel to be filled in order to correct the topological defect. Henceforth we only discuss handles since tunnels are dealt with in a similar way, but with background voxels replacing foreground voxels. The equivalence between handles and tunnels arises from the fact that the surface is non-self-intersecting.

Let $[A, B]$ be a pair of interlocking cycles at v. Let C be one of A or B and assume that C is a cycle of vertices $v_1, v_2, \ldots, v_n$ where v_1 and v_n are actually the same vertex v. Along C construct a set P of voxels $u_1, u_2, \ldots, u_m$ such that: $u_1 = u_m$; u_j has a common face with some u_k with $1 \leq k < j$, $1 < j \leq m$; each edge (v_i, v_{i+1}) in C, $1 \leq i < n$, is adjacent to a voxel u_j in P; and, each voxel u_j in P is adjacent to an edge in C.

Now, in order for the set P to properly open a handle, it must also have genus zero. If C is small then P is likely to have genus zero (see Figure 6), but it is not difficult to see that when C is large, then P itself may possess handles. One remedy to this situation is to fill the handles of P by examining each pair of consecutive voxels u_j and u_{j+1} in P and deduce from it which voxel should be added. Once P is obtained from C and P has genus zero then one decides if removing the voxels in P (in case of a handle) or adding them (in case of a tunnel) properly corrects the topology (see Figure 6). This is easily verified by counting the vertices, edges and faces of the corresponding digital image and applying Euler's formula. After repair we need to run the correction filter on the vertices incident to these voxels to ensure that the newly obtained surface is non-self-intersecting.

Figure 7 shows the result of applying the algorithm to a segmented hippocampus. Dark voxels are those added and light voxels are those removed as a

result of applying the correction filter and the topological repair routine. Notice how the handle has been broken up.

Finding the initial set P along C is easy but may be expensive as there may be many different solutions. We always choose the smallest. Filling the handles of P, if P is not already topologically equivalent to a sphere, is harder and may not succeed in rendering P topologically correct. Similarly, we may not succeed in finding a set P such that P corrects the topology. In this case we reject P and try another path of voxels P along cycle C. If no appropriate set P can be found for either cycle A or B in the $[A, B]$ pair, we examine another pair of interlocking cycles at v. All the cycles belonging to pairs of interlocking cycles at v are considered in non-increasing order of size, thus ensuring that, at a given vertex v, we find the minimal set P that repairs the topological defect. If there is a handle in the neighbourhood of a given vertex v, there are many cycles passing through v, and given a cycle C through v, there are many transverses of C through v. Figure 6 shows how different transverses are found for the same cycle.

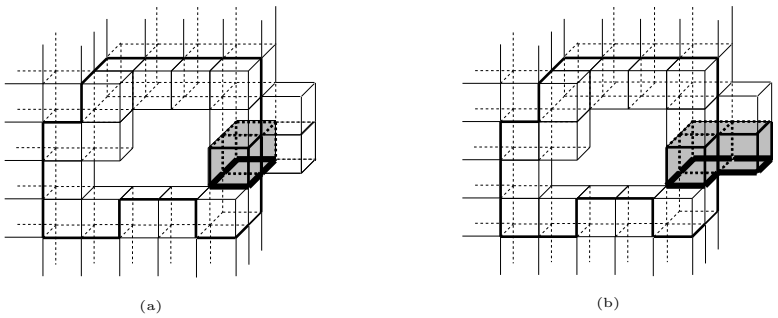

Fig. 6. The cycle in (a) does not succeed in finding an appropriate set of voxels to break the handle, while the one in (b) does.

7 Experiment

We have used the algorithm to successfully compute a topologically correct voxel-based surface mesh on all valid 977 hand-segmented hippocampi in the database resulting from a longitudinal study on ageing, The PATH Through Life Project, conducted by the Centre for Mental Health Research at the Australian National University, Canberra. One hippocampus among the data set of 977 is shown in Figure 7.

As can be seen from Section 3 and especially from Section 6 the proposed algorithm is potentially exponential in time with respect to the number of vertices on the surface. However for all practical purposes the algorithm corrects the surface in a reasonable time. Over the whole data set and on a 3.2GHz Pentium 4 processor, it spends on average 40.77 seconds processing 2977 voxels.

Of the 977 hand-segmented hippocampi, 271 present a non-zero genus surface with 378 handles in total. It takes on average 82.24 seconds to correct each data set processing an average volume of 2932 voxels, and it takes 2.57 voxels to correct each handle. When considering the total number of modified voxels (that is, including those modified as a result of the correction filter) on average 23.47 voxels are changed over all the 977 data sets.

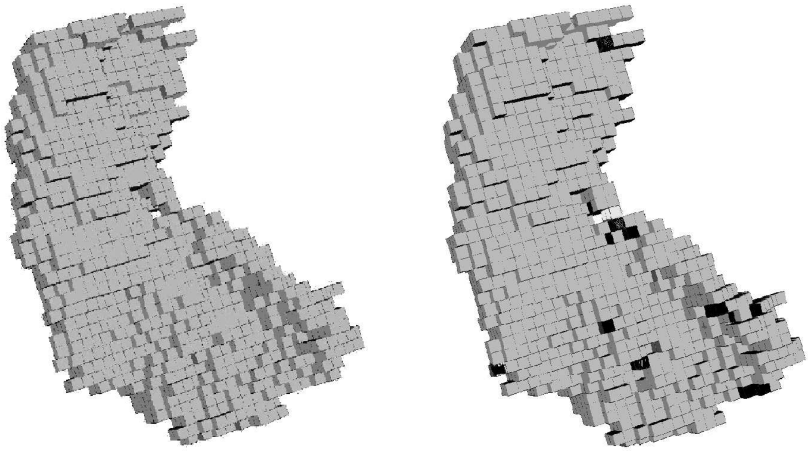

Fig. 7. Left: A segmented hippocampus: original digital data. Right: The hippocampus with spherical topology realised; Dark voxels are those added and light voxels are those removed. Notice how the handle has been broken up.

8 Discussion

One issue raised in Section 6 is the possibility of the algorithm not being able to find a set of voxels that correctly repairs the topology. Our experiment shows that we can successfully deal with anatomical data with small defects. This is consistent with our approach where difficulties may arise only when the size of the topological defect becomes significantly large. Figure 8 lists the size of the corrections effected together with their frequencies.

Size of handle corrections	1	2	3	4	5	6	7	8	9	10	11	
Frequencies	100	131	69	37	19	12	3	1	3	2	1	378

Fig. 8. Size of handle corrections and their frequencies.

There is also the difficulty of effecting topological repair by modifying a minimum number of voxels. As can be seen from the order in which we examine interlocking cycles we may not necessarily guarantee a minimal amount of change since a handle may be described by cycles interlocking at more than one vertex. Our strategy is principally guided by considerations of time and space efficiency and by the assumption that topological defects remain small.

The strengths of the proposed algorithm are its simplicity and the fact that it makes no assumption about the digital data, except that the object input is connected. In particular, as opposed to [2], one does not require that the input data be 6-connected.

The simplicity of the algorithm lies in the way handles are located. Finding interlocking cycles at a given vertex v is linear in the size of the handle present in the vicinity of v. When no handle is present the search is quadratic in s, the size of the largest handle in the surface. In practice, one terminates the search after a preset value s has been reached.

Finally, an important feature of the proposed algorithm is that it produces a voxel-based quadrangulation, as opposed to a Marching Cubes tessellation [9–11, 13, 1, 2]. This is necessary when intending to compute the shape's spherical harmonics, as the quadrangulation's vertices can be uniformly mapped onto the sphere, ensuring a faithful reconstruction using from harmonic series.

Our algorithm could be improved in at least two ways. Firstly we need to find a more robust means of converting a cycle of edges in the surface into a minimal set of voxels that repairs the topology. In this regard it might be helpful to investigate some of the ideas in [14]. These ideas may also help in reducing the algorithm's complexity, which principally arises from the attempt at repairing the discovered defects. Secondly and lastly, we may think of a different strategy for choosing interlocking cycles to ensure minimum change.

In the future we wish to test our algorithm on cortical surfaces; they are known to be topologically equivalent to a sphere and their reconstructed surface usually shows many topological defects ([1, 2, 18]). The defects are typically small so one would expect our algorithm to perform well. Of interest will be to investigate how the changes in volume resulting from the application of the correction filter and the repair routine compare with techniques described in [1, 2].

Acknowledgments

National ICT Australia is funded by the Australian Department of Communications, Information Technology and the Arts and the Australian Research Council through *Backing Australia's ability* and the ICT Centre of Excellence Program.

The authors thank K. Anstey, C. Meslin, J. Maller, and the PATH research team at the Centre for Mental Health Research, Australian National University, Canberra, and P. Sachdev, Neuropsychiatric Institute, University of New South Wales, Sydney, for providing the original MR and segmented data sets.

The authors would also like to thank M. Styner for a helpful exchange.

References

1. Han, X., Xu, C., Braga-Neto, U., Prince, J.: Topology Correction in Brain Cortex Segmentation Using a Multiscale, Graph-Based Algorithm. IEEE Transactions on Medical Imaging **21**(2) (2002) 109–121
2. Shattuck, D., Leahy, R.: Automated Graph-Based Analysis and Correction of Cortical Volume Topology. IEEE Transactions on Medical Imaging **20**(11) (2001) 1167–1177
3. Keleman, A., Szekely, G., Gerig, G.: Elastic Model-Based Segmentation of 3D Neuroradiological Data Sets. IEEE Trans. Medical Imaging **18**(10) (1999) 828–839
4. Styner, M., Lieberman, J., Pantazis, D., Gerig, G.: Boundary and Medial Shape Analysis of the Hippocampus in Schizophrenia. Medical Image Analysis **8** (2004) 197–203
5. Styner, M., Lieberman, J., McClure, R., Weingberger, D., Jones, D., Gerig, G.: Morphometric Analysis of Lateral Ventricles in Schizophrenia and Healthy Controls Regarding Genetic and Disease-Specific Factors. In: Proceedings of the National Academy of Science. Volume 102. (2005) 4872–4877
6. Thompson, P., Moussai, J., Zohoori, S., Goldkorn, A., Khan, A., Mega, M., Small, G., Cummings, J., Toga, A.: Cortical Variability and Asymmetry in Normal Aging and Alzheimer's disease. Cereb. Cortex **8**(6) (1998) 492–509
7. Brechbühler, C., Gerig, G., Kübler, O.: Parametrization of Closed Surfaces for 3-D Shape Description. Computer Vision and Image Understanding **61**(2) (1995) 154–170
8. Fischl, B., Sereno, M., Dale, A.: Cortical Surface-Based Analysis II: Inflation, Flattening, and a Surface-Based Coordinate System. NeuroImage **9** (1999) 195–207
9. Bazin, P., Pham, D.: Topology correction using Fast Marching Methods and its Application to Brain Segmentation. In: MICCAI. (2005)
10. Jaume, S., Rondao, P., Macq, B.: Open Topology: A Toolkit for Brain Isosurface Correction. In: MICCAI. (2005)
11. Couprie, M., Najman, L., Bertrand, G.: Quasi-Linear Algorithms for the Topological Watershed. Journal of Mathematical Imaging and Vision **22**(2-3) (2005) 231–249
12. Bischoff, S., Kobbelt, L.: Isosurface Reconstruction with Topology Control. In: 10th Pacific Conference on Computer Graphics and Applications. (2002) 246
13. Segonne, F., Grimson, E., Fischl, B.: A Genetic Algorithm for the Topology Correction of Cortical Surfaces. In: In Proceedings of Information Processing in Medical Imaging, LNCS. Volume 3565. (2005) 393–405
14. Aktouf, Z., Bertrand, G., Perroton, L.: A Three-Dimensional Holes Closing Algorithm. Pattern Recognition Letters **23**(5) (2002) 523–531
15. (Styner, M.) Personal communication.
16. Abrams, L., Fishkind, D., Priebe, C.: The Generalized Spherical Homeomorphism for Digital Images. IEEE Transactions on Medical Imaging **23**(5) (2004) 655–657
17. Lawson, T.: Topology: A Geometric Approach. Oxford University Press (2003)
18. Fischl, B., Sereno, M., Dale, A.: Cortical Surface-Based Analysis I: Segmentation and Surface Reconstruction. NeuroImage **9** (1999) 179–194

Non-parametric Image Registration Using Generalized Elastic Nets

Andriy Myronenko, Xubo Song, and Miguel Á. Carreira-Perpiñán

Dept. of Computer Science & Electrical Engineering
OGI School of Science & Engineering, Oregon Health & Science University
20000 NW Walker Road, Beaverton, OR 97006, USA
{myron,xubosong,miguel}@csee.ogi.edu

Abstract. We introduce a novel approach for non-parametric non-rigid image registration using generalized elastic nets. The concept behind the algorithm is to adapt an elastic net in spatial-intensity space of one image to fit the second image. The resulting configuration of the net, when it achieves its minimum energy state, directly represents correspondence between images in a probabilistic sense and recovers underlying image deformation, which can be arbitrary. Representation of elastic net in the spatial-intensity space with specific priors that enforce natural elastic deformation is introduced. Efficient algorithm for optimization of elastic net energy is developed. The accuracy and effectiveness of the method is demonstrated on different medical image registration examples with locally non-linear underlying deformations.

1 Introduction

Image registration is an important component in medical image analysis. It is a process of determining a geometric transform that relates the contents of two images in a meaningful way and establishes the correspondence between them. Applications of image registration include combining images of the same subject from different modalities, aligning temporal sequences of images to compensate for motion of the subject between scans, image guidance during interventions and aligning images from multiple subjects in cohort studies [1].

Non-rigid image registration is the most interesting and challenging work in registration today. Many non-rigid registration techniques have been proposed during last 20 years [2]. Most of them build a parameterized model that constrains the form of allowed deformations and then optimize a similarity function to find an approximation of a real underlying deformation. Study of non-parametric registration has focused on variational approaches [3]. In this paper we introduce a non-parametric registration method that can deal with non-rigid deformation of arbitrary complexity, using a probabilistic model known as the elastic net (EN). The elastic net is a net of connected points which jointly and smoothly move in a high-dimensional space to model a data set. An energy function can be defined to trade off accuracy of the net fitting the data (fitness term) vs net continuity (tension term). The elastic net was originally introduced as a

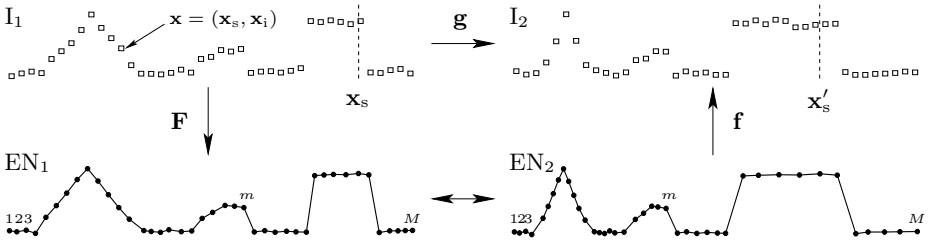

Fig. 1. Illustration of the alignment method (for 1D images). I_1 represents a 1D intensity image in spatial (x-axis) and intensity (y-axis) space (each pixel is marked as a small □). I_2 represents the same image with local, nonlinear spatial distortion and intensity noise. EN_1 is an elastic net fitted to I_1 (in spatial-intensity space, with centroids marked •) and EN_2 is adapted from EN_1 to fit I_2. Since the centroids in EN_1 and EN_2 correspond one-to-one ($1 \leftrightarrow 1$, etc.), and the elastic net allows to define mappings between image points and centroids (see section 2), we can map any spatial location $\mathbf{x}_s$ in I_1 to a spatial location $\mathbf{x}'_s$ in I_2 through the elastic nets, thus aligning I_1 to I_2.

continuous optimization method for the traveling salesman problem [4, 5] and has also been successfully applied to modeling maps of primary visual cortex. However it has had a limited use in computer vision. A generalization of elastic nets to arbitrary quadratic tension terms was investigated in [6]. Here we adapt the generalized elastic net to represent image deformations. The intuition is to position a net according to the first image and then deform it to align with the second image. The deformation produced by elastic net, when its energy is minimized, directly represents the deformation field between the images.

This is illustrated more specifically in Fig. 1. We consider an image as a noisy 2D manifold in the spatial-intensity space, i.e., each pixel is represented by a point $\mathbf{x} = (\mathbf{x}_s, \mathbf{x}_i) \in \mathbb{R}^3$ of spatial location $\mathbf{x}_s \in \mathbb{R}^2$ and intensity $\mathbf{x}_i \in \mathbb{R}$. We model this manifold in a probabilistic way with an elastic net EN_1, which allows to map any image point onto the net, and vice versa. We then adapt EN_1 for a given image I_1 to a new image I_2 in the spatial-intensity space; again this allows to map a net point onto image space and vice versa. The alignment mapping which maps a spatial location in I_1 to another spatial location in I_2 is obtained through the deformed elastic net. We describe the method of generalized elastic net and its adaptation to image registration problem in detail in section 2, give experimental results in section 3 and discuss them in section 4.

2 Image Registration with Generalized Elastic Nets

Generalized elastic nets (GEN) The *elastic net* is a Gaussian mixture with a quadratic prior on its centroids [4–6]. The centroids implicitly represent a nonlinear, low-dimensional manifold that probabilistically models a high-dimensional data set $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ (expressed as a $D \times N$ matrix). Specifically, given a collection of M D-dimensional centroids $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_M)$ (ex-

pressed as a $D \times M$ matrix) and a scale parameter $\sigma \in \mathbb{R}^+$, consider a Gaussian-mixture density $p(\mathbf{x}) = \sum_{m=1}^{M} \frac{1}{M} p(\mathbf{x}|m)$ with $\mathbf{x}|m \sim \mathcal{N}(\mathbf{y}_m, \sigma^2 \mathbf{I}_D)$. A smoothing or neighborhood-preserving prior on the centroids is defined as $p(\mathbf{Y}; \beta) \propto \exp\left(-\frac{\beta}{2} \sum_m \|\mathbf{y}_{m+1} - \mathbf{y}_m\|^2\right)$ where β is a regularization hyperparameter. Without the prior, the centroids could be permuted at will with no change in the model, since the variable m is just an index. The elastic net minimizes the energy function

$$E(\mathbf{Y}, \sigma) = -\sum_{n=1}^{N} \log \sum_{m=1}^{M} e^{-\frac{1}{2}\left\|\frac{\mathbf{x}_n - \mathbf{y}_m}{\sigma}\right\|^2} + \frac{\beta}{2} \sum_m \|\mathbf{y}_{m+1} - \mathbf{y}_m\|^2 \qquad (1)$$

which is derived from the log posterior $\log p(\mathbf{Y}|\mathbf{X}, \sigma)$ of the full model (i.e., maximum-a-posteriori estimation). We call the first term the *fitness term*, arising from the Gaussian mixture $p(\mathbf{X}|\mathbf{Y}, \sigma)$, and the second term the *tension term*, arising from the prior $p(\mathbf{Y})$. The elastic net was generalized in [6, 7] to accommodate general quadratic priors. The prior can be used to convey the topological (dimension and boundary conditions) and geometric (e.g. curvature) structure of a manifold implicitly defined by the centroids. The *generalized elastic net (GEN)* minimizes the energy function

$$E(\mathbf{Y}, \sigma) = -\sum_{n=1}^{N} \log \sum_{m=1}^{M} e^{-\frac{1}{2}\left\|\frac{\mathbf{x}_n - \mathbf{y}_m}{\sigma}\right\|^2} + \frac{\beta}{2} \operatorname{tr}\left(\mathbf{Y}^T \mathbf{Y} \mathbf{S}\right). \qquad (2)$$

Quadratic priors are considered of the form $\mathbf{S} = \mathbf{D}^T \mathbf{D}$, so that $\operatorname{tr}\left(\mathbf{Y}^T \mathbf{Y} \mathbf{S}\right) = \left\|\mathbf{D}\mathbf{Y}^T\right\|^2$ in terms of the Frobenius norm. The matrix $\mathbf{D}$ represents a discretized differential operator. For example (for a 1D net for simplicity, and using forward differences [6]), a first-order derivative results in a sum of squared lengths $\left\|\mathbf{D}\mathbf{Y}^T\right\|^2 = \sum_m \|\mathbf{y}_{m+1} - \mathbf{y}_m\|^2$ and approximates a penalty $\int \|\nabla \mathbf{y}\|^2$ over a continuous net $\mathbf{y}$ (with an infinite number of centroids). This corresponds to a matrix $\mathbf{D}$ where each row is a shifted version of $(-1\ 1\ 0\ 0 \ldots 0)$, and it was the tension term used in the original elastic net (Eq.(1)), penalizing stretching of the net. A second-order derivative results in $\sum_m \|\mathbf{y}_{m+2} - 2\mathbf{y}_{m+1} + \mathbf{y}_m\|^2$, etc. By choosing $\mathbf{S}$ as an appropriate combination of differential operators we can impose a desired type of smoothness on the GEN (see [7] for a discussion of the effect of different derivatives on the maps of primary visual cortex). The resulting $\mathbf{S}$ has a sparse, banded structure. We consider open boundary conditions at the image boundaries. Fig. 1 schematically shows a 1D elastic net.

Adaptation of the GEN Although it is possible to derive an EM algorithm to estimate $\mathbf{Y}$ and σ jointly, the GEN is usually trained with a deterministic annealing algorithm in order to obtain good local minima. This minimises E over $\mathbf{Y}$ for fixed σ, starting with a large σ and tracking the minimum to a small value of σ. For constant σ, [6] used a fixed-point iteration to find stationary points of E:

$$\frac{\partial E}{\partial \mathbf{Y}} = -\frac{1}{\sigma^2}(\mathbf{X}\mathbf{W} - \mathbf{Y}\mathbf{G}) + \beta \mathbf{Y}\left(\frac{\mathbf{S} + \mathbf{S}^T}{2}\right) = \mathbf{0} \quad \Longrightarrow \quad \mathbf{Y}\mathbf{A} = \mathbf{X}\mathbf{W} \qquad (3)$$

with weight matrix $\mathbf{W} = (w_{nm})$ and invertible diagonal matrix $\mathbf{G} = \mathrm{diag}\,(g_m)$

$$w_{nm} = \frac{e^{-\frac{1}{2}\left\|\frac{\mathbf{x}_n - \mathbf{y}_m}{\sigma}\right\|^2}}{\sum_{m'=1}^{M} e^{-\frac{1}{2}\left\|\frac{\mathbf{x}_n - \mathbf{y}_{m'}}{\sigma}\right\|^2}} \qquad g_m = \sum_{n=1}^{N} w_{nm} \qquad \mathbf{A} = \mathbf{G} + \sigma^2\beta\left(\frac{\mathbf{S} + \mathbf{S}^T}{2}\right).$$

The weight w_{nm} is the responsibility $p(m|\mathbf{x}_n)$ of centroid y_m for generating point $\mathbf{x}_n$, g_m is the total responsibility of centroid $\mathbf{y}_m$, and the matrix $\mathbf{XW}$ is a list of average centroids. We solve for $\mathbf{Y}$ in the system of eq. (3) and iterate, since $\mathbf{W}$ and $\mathbf{G}$ depend on $\mathbf{Y}$. In [6], the system (3) was solved using Cholesky factorisation. While this is robust and efficient (since it takes advantage of the sparsity structure of $\mathbf{S}$), here we use a different method based on linear conjugate gradients (CG) [8]. Linear CG solves an $M \times M$ positive definite linear system in at most M steps, each costing $\mathcal{O}(M^2)$ (actually less since $\mathbf{A}$ is sparse), and has two important advantages: (1) we can initialize the linear CG from the previous $\mathbf{Y}$ value (which will be close to the solution) rather than solving each system anew, as Cholesky does; (2) we can run only a few linear CG steps and obtain an approximate but good enough solution rather than an exact, costly one. This considerably accelerates the overall annealing algorithm without sacrificing accuracy. A further acceleration is obtained by truncating the Gaussian kernel so that most weights w_{mn} are zero and can be ignored.

Registration We now show how the framework of elastic net can be adapted for the problem of image registration. First, we represent two images I_1 and I_2 in the spatial-intensity space. Then we construct an elastic net with as many centroids as pixels in image I_1. This net $\mathbf{Y}$ is initialized with each centroid representing the spatial-intensity value of one pixel in I_1 (i.e., $\mathbf{Y} = \mathbf{X}_1$). The net is adapted by adjusting the centroids to fit data $\mathbf{X}_2$, representing image I_2 in spatial-intensity space. This is done by minimizing the energy in Eq. (2). The intuition for using the same number of centroids as there are pixels in I_1 is that the final centroid locations, when the energy function is optimized, directly shows the displacement of each pixel in I_1 when it is deformed into I_2. As a result, no interpolation is needed. It also provides the maximum level of deformation complexity. In general, we can choose to have more or fewer centroids than pixels. In this case the displacement of a pixel $\mathbf{x}$ in I_1 can be found by interpolation using the probabilities $p(m|\mathbf{x})$ and $p(\mathbf{x}|m)$ provided by the GEN.

We assume that the deformation between two images is only spatial, not in intensity. This translates to constraining the intensity components in the centroid vectors to be constant. In other words, the free parameters for centroid $\mathbf{y}_m = (\mathbf{y}_{ms}, \mathbf{y}_{mi})$ are $\mathbf{y}_{ms}$ only, and the optimization updates only apply to $\mathbf{y}_{ms}$. Doing so is important to produce only spatial deformation for I_1 when fitting it to I_2. In general, intensity variations across images can be accommodated by updating the complete $\mathbf{y}_m = (\mathbf{y}_{ms}, \mathbf{y}_{mi})$. We use the following penalty matrix: $\mathbf{S} = \beta_1\mathbf{D}_1^T\mathbf{D}_1 + \beta_2\mathbf{D}_2^T\mathbf{D}_2$, where $\mathbf{D}_1$ and $\mathbf{D}_2$ are first- and second-order derivatives, and their relative strengths are controlled by hyper-parameters β_1 and β_2; practically useful values for them can be obtained manually for a given type of images.

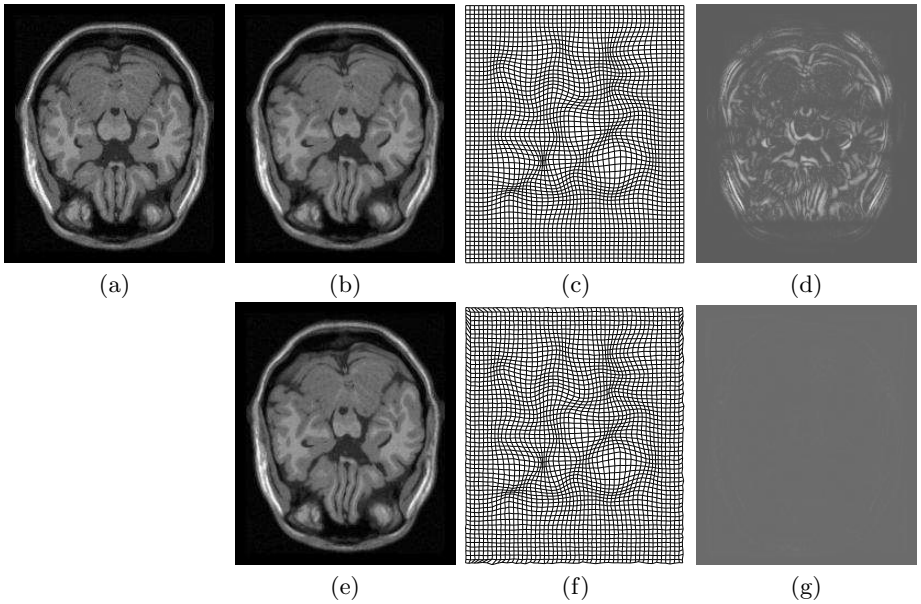

Fig. 2. (a) Original MR slice; (b) deformed original image according to control points using thin plate splines; (c) deformation field represented by control point; (d) absolute value of intensity difference between original and deformed images; (e) registration result of the algorithm (original image is registered onto the deformed one); (f) deformation field found by the algorithm (almost equal to the true one); (g) absolute value of intensity difference between registered and original deformed images (almost zero).

3 Experimental Results

In all experiments, the image intensities are first re-scaled to allow the use of a single σ for all dimensions, and the images are coarsely aligned using cross-correlation to eliminate rigid translation. The resulting data sets $\mathbf{X}_1, \mathbf{X}_2 \subset \mathbb{R}^3$ were used to adapt the elastic net. The resulting, aligned dataset $\mathbf{X}_1'$ (obtained from the spatial deformation given by the GEN and the original intensity values) was post-processed with bi-cubic interpolation to produce the aligned image. The prior parameters β_1 and β_2 were set manually for each type of image. We ran 10 annealing iterations from $\sigma = 3$ to $\sigma = 0.5$ pixels.

We show the performance of the algorithm on artificial data with known non-linear deformations and on two real-life examples. The algorithm was implemented in Matlab with subroutines coded in C, and tested on Pentium4 CPU 3.5GHz with 4Gb RAM. The test images are gray-scale images of size 250 $\times$ 250, and the registration process takes about 20 minutes for each image pair.

Brain MRI 2D images with and without known deformation A slice of MRI brain image was artificially deformed using known deformation field

Table 1. Experimental results for different deformation levels.

Deformation STD	Transformation RMSE	Intensity RMSE
1.0	0.3135	0.0044
1.5	0.5124	0.0047
2.0	0.9753	0.0053
2.5	1.1152	0.0060
3.0	1.0962	0.0059

(Fig. 3). We define a uniform grid of control points in the original image, randomly move them and use the thin plate spline technique to create a locally nonlinear deformed image. Our algorithm is applied to align original image (a) onto deformed one (b). The final absolute image difference (g) is so small it is hardly visible, demonstrating the high accuracy of the method. Table 1 shows the value of root mean square error (RMSE) between true and estimated deformation as well as the intensity RMSE between original and registered images, as a function of spatial distortion level controlled by the standard deviation (STD) of control points perturbation measured in pixels. The transformation error is at most of the order of one pixel.

Figure 3 shows the registration of images (a) and (b) from two patients. Image (b) is registered onto (a) resulting in (d). Panels (c) and (e) are color composite views of the two images before (c) and after (e) registration, where image (a) is coded with green and (b) with red color. Visual inspection clearly reveals much improved alignment, even when the two original images have significantly different intensity ranges.

Microscopic iris images We stabilize a video sequence of microscopic iris images through frame-by-frame registration. This is necessary to remove the severe jitter and deformation across frames in order to be able to track the leukocyte motion. The deformation between frames is highly nonlinear. Our algorithm proves to be accurate and effective for these images, as demonstrated in Fig. 4. Ideal registration in this case should lead to an absolute difference image after alignment with background intensity close to zero and bright blobs corresponding to the moving leukocytes, which is exactly the case in Fig.4(e).

4 Conclusion

We have developed a probabilistic approach for non-parametric non-rigid image registration based on the generalized elastic net (using first- and second-order differential priors). The method is able to accurately register images even with highly non-linear local deformations, and we have designed a new, more efficient optimization algorithm based on linear conjugate gradients. When the elastic net is initialized with one centroid for each pixel in image I_1, the resulting deformed net will provide directly the displacement for each pixel. When the number

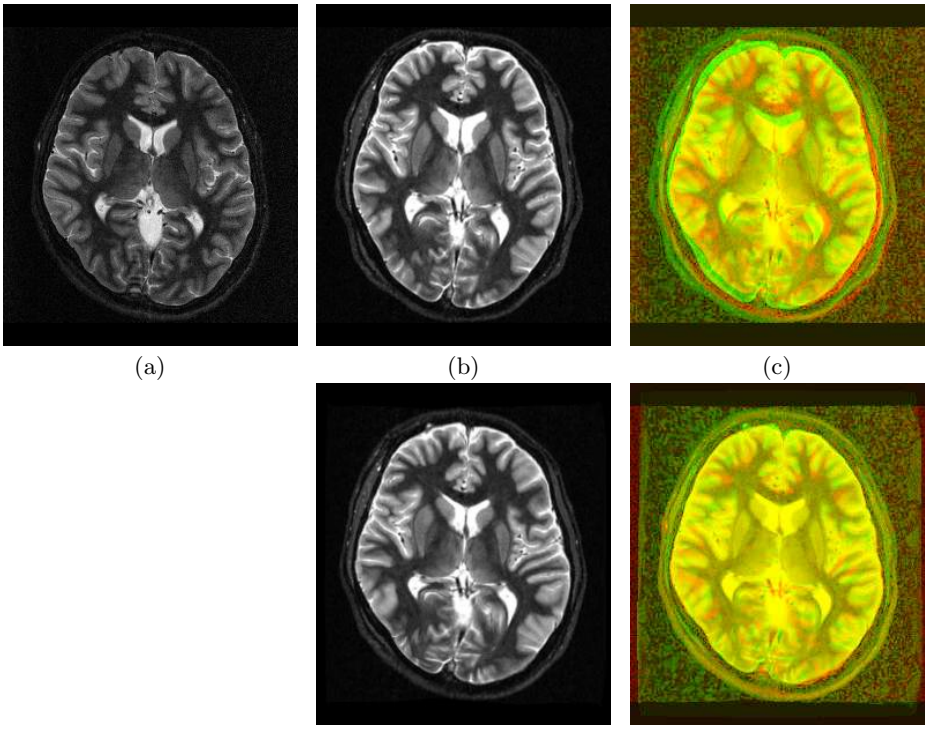

Fig. 3. (a) First person; (b) second person; (c) their composite view; (d) registration of (b) onto (a); (e) composite view of (a) and (d).

of centroids goes to infinity in the limit, the mapping approximates continuous mapping with continuous derivative, however it is still implicitly defined by finite collection of centroids. In general, the deformation complexity can be controlled by using an arbitrary number of centroids. In this case, determining the displacement of a pixel in I_1 can be interpolated by relating probabilistically the pixel and the elastic net. With the image deformation represented as the motion of net centroids, we do not need any image interpolation on each iteration, unlike most other registration methods. While we have focused on intensity features, the method easily accommodates arbitrary features (e.g. gradient information and color components) and images of different spatial resolutions. The generalization of the method for 3D images is straightforward, however the computational time is large at present. One possible way to reduce the computational complexity is to use fewer centroids. The method is also well suited for continuous tracking of the centroids over consecutive frames in an image sequence, by successively adapting the net to each image.

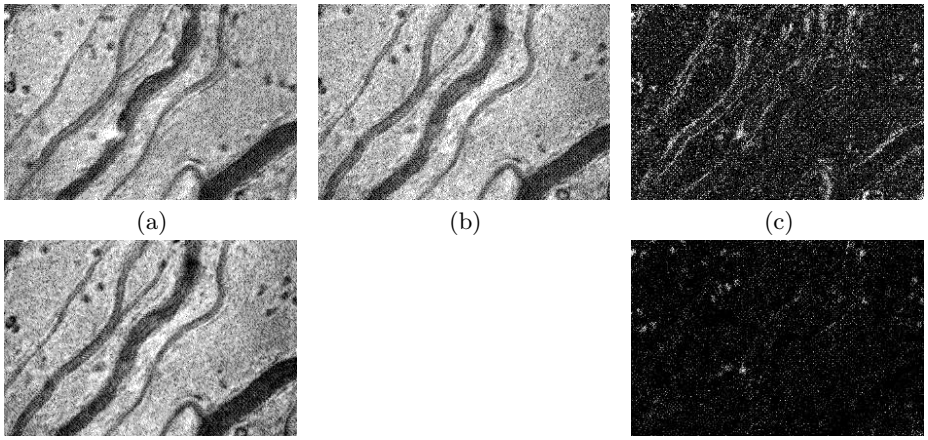

Fig. 4. Microscopic video of iris: (a) frame 1, (b) frame 37; (c) absolute intensity difference between the two frames before registration; (c) registration result of the algorithm: image (b) is aligned with image (a); (e) absolute intensity difference between the two frames after registration.

Future research includes varying the number of centroids for different deformation complexity, adaptive choice of the regularization hyper-parameter, and local adaptation of centroid variance σ for different dimensions to deal better with local image properties and intensity variations.

References

1. Hill, D.L.G., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical image registration. Physics in Medicine and Biology **46**(3) (2001) R1–R45
2. Crum, W.R., Hartkens, T., Hill, D.L.G.: Non-rigid image registration: Theory and practice. Brit. J. of Radiology **77**(special issue) (2004) S140–S153
3. Thirion, J.P.: Image matching as a diffusion process: An analogy with Maxwell's demons. Medical Image Analysis **2**(3) (1998) 243–260
4. Durbin, R., Willshaw, D.: An analogue approach to the traveling salesman problem using an elastic net method. Nature **326**(6114) (1987) 689–691
5. Durbin, R., Szeliski, R., Yuille, A.: An analysis of the elastic net approach to the traveling salesman problem. Neural Computation **1**(3) (1989) 348–358
6. Carreira-Perpiñán, M.Á., Dayan, P., Goodhill, G.J.: Differential priors for elastic nets. In Gallagher, M., Hogan, J., Maire, F., eds.: Proc. of the 6th Int. Conf. Intelligent Data Engineering and Automated Learning (IDEAL'05). Volume 3578 of Lecture Notes in Computer Science., Springer-Verlag (2005) 335–342
7. Carreira-Perpiñán, M.Á., Goodhill, G.J.: Influence of lateral connections on the structure of cortical maps. J. Neurophysiol. **92**(5) (2004) 2947–2959
8. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer-Verlag, New York (1999)

Measurment of folding in surfaces of arbitrary size in human brain development

C. Rodriguez-Carranza[1], P. Mukherjee[2], D. Vigneron[2], J. Barkovich[2], and C. Studholme[1,2]

[1] NCIRE/VAMC, San Francisco, CA
[2] Department of Radiology, University of California, San Francisco

Abstract. This paper describes a novel approach to in-vivo measurement of brain surface folding in clinically acquired neonatal MR image data. Specifically it addresses the problem of comparing folding of surfaces of arbitrary area. Most of the current measures of folding are not independent of the area of the surface from which they are derived. Therefore, applying them to whole brains or subregions of different sizes result in differences which may or may not reflect true differences in folding. In this paper we describe alternative approaches to deriving area independent measures. The measures were applied to twelve premature infants (age 28-37 weeks) from which cortical gray and white surfaces were extracted. Experimental results show that previous folding measures are sensitive to the size of the surface patch, and that the area independent measures proposed here provide significant improvements. Such a system provides a tool to allow the study of structural development in the neonatal brain within specific functional subregions, which may be critical in identifying later neurological impairment.

1 Introduction

The percent of infants born preterm, or at less than 37 completed weeks of gestation, has increased significantly in the last two decades. The latest report [6] in the USA indicates that almost one in every eight births is preterm. There is growing evidence that premature birth can result in structural and functional alterations of the brain, which are related to adverse neurodevelopmental outcome later in life [7, 8]. Some of the challenges that preterm infants face range from spastic motor deficits (cerebral palsy) [20], impaired academic achievement [3,5], and behavioral disorders [9, 12]. However, the conditions that cause the cerebral abnormalities that underlie these common and serious developmental disabilities are not entirely understood [8]. The wider availability of clinical in vivo magnetic resonance imaging of neonatal brain anatomy, provided by systems which make use of an MRI compatible incubator, creates a new opportunity to quantify brain development. In this work we are particularly interested in the study of the cortical folding, or gyrification, in preterm infants, because it may reflect underlying functional organization.

In our earlier work [16] we applied previously proposed global average measures of folding to surfaces extracted from neonatal MRI, and showed their use in tracking global age changes. The main limitation was that the previously proposed measures were heavily dependent on the size of the surface being examined. Since brain surface area increases dramatically with brain development, previous measures cannot probe whether cortical folding is following a normal pattern, independent of size. Additionally these cannot be calculated on functional sub-regions of the cortex which may have a different surface area in different individuals. To address the problem, in this paper we describe and analyze measures of global folding that are independent of the size of the surface of analysis. We apply the measures to surfaces extracted from MR images of twelve premature infants with ages from 28-37 weeks. The folding of whole surfaces, as well as their left and right hemispheres, was studied.

2 Theory

Several global 3D brain surface measures of gyrification have previously emerged. Some are based on surface principal curvatures (k_1, k_2), mean curvature $(H = \frac{1}{2}(k_1 + k_2))$, or Gaussian curvature $(K = k_1 k_2)$. Examples are folding index (FI) [4], the intrinsic curvature index (ICI) [4], L^2 norm of the mean curvature (MLN) [1], L^2 norm of Gaussian curvature (GLN) [1], and average curvature [13]; global shape index (GS) and global curvedness (GC) were defined in our earlier work [16] based on the local shape descriptors curvedness (c) [10] and shape index (s) [10]. Examples of non-curvature based measures are the gyrification index [21], calculated as the ratio of the entire cortical contour of the brain to the superficially exposed contour, and roundness[1] (Rn), based on surface area (A) and volume (V). A list of current global 3D measures of folding is shown on the left column of Table 1. These expressions are normalized to yield the unit value for a sphere [16].

2.1 Size-independent measures of surface folding

Examination of the form of the current global measures in Table 1 reveals a critical dependency on surface area, with the exception of global shape index and gyrification index. This can be better illustrated with the following example, using MLN to measure folding. Take a whole sphere of radius Ro, half a sphere of radius Ro, and a whole sphere of half that radius. The three objects have the same surface complexity, hence a measure of global folding should yield identical results. By definition, $H = 1/R$ at each point in a sphere. The whole sphere with radius Ro yields $MLN = \frac{1}{4\pi}\sum_A H^2 = \frac{1}{4\pi} \times 4\pi Ro^2 \times (1/Ro)^2 = 1$, and so does the whole sphere with radius $Ro/2$: $MLN = \frac{1}{4\pi} \times (4\pi\frac{Ro}{2})^2 \times (2/Ro)^2 = 1$. On the other hand, MLN for the half sphere of radius Ro is half that value: $MLN = \frac{1}{4\pi} \times \frac{1}{2} 4\pi Ro^2 \times (1/Ro)^2 = \frac{1}{2}$. This demonstrates the dependency of the

[1] The terms surface complexity [13] and isoperimetric ratio [1] have also been used.

measure on the size of the surface of analysis. Normalizing MLN with the surface area does not alleviate the problem: $MLN = \frac{1}{A}\sum_A H^2 = \frac{1}{4\pi Ro^2}(4\pi Ro^2)(1/Ro)^2 = \frac{1}{Ro^2}$ (whole sphere of radius Ro). We propose two normalization factors: 1) $T = 3V/A$ and 2) $|\overline{H}| = \frac{1}{A}|\sum_A H|$. The resulting T-normalized and H-normalized expressions (T-measures and H-measures, respectively) are shown in the right column of Table 1. For a sphere, $T = Ro$, $|\overline{H}| = 1/Ro$, and it can be verified that $MLN_T = MLN_H = 1$ for the three sphere cases. Similarly for the rest of the newly defined area-independent measures.

Table 1. Measures of surface folding

Current measures	**New measures**
Area-dependent	Area-independent
$MLN=\frac{1}{4\pi}\sum_A H^2$ [1]	$MLN_T=\frac{T^2}{A}\sum_A H^2$
$GLN=\frac{1}{4\pi}\sqrt{A*\sum_A K^2}$ [1]	$GLN_T=T\sqrt[4]{\frac{1}{2}\sum_A K^2}$
$ICI=\frac{1}{4\pi}\sum_A K^+$ [4]	$ICI_T=T\sqrt{\frac{1}{A}\sum_A K^+}$
$FI=\frac{1}{4\pi}\sum_A ak$ [4]	$FI_T=T\sqrt{\frac{1}{A}\sum_A ak}$
$GC=\frac{1}{\sqrt{A4\pi}}\sum_A c$ [16]	$GC_T=T\sqrt{\frac{1}{A}\sum_A c}$
$\overline{U}=\frac{1}{A}\sum_A U$ [13]	$\overline{U_T}=\frac{T}{A_U}\sum_A U$
$\overline{V}=\frac{1}{A}\sum_A V$	$\overline{V_T}=T\sqrt{\frac{1}{A_V}\sum_A V}$
$Rn=\frac{A}{\sqrt[3]{36\pi V^2}}$ [1]	$SH2SH=T\sum_A H^2/\sum_A H$
	$SK2SK=T\sqrt{\sum_A K^2/\sum_A K}$
Area-independent	
$GS=\frac{1}{A}\sum_A s$ [16]	$MLN_H=\frac{1}{\lvert\overline{H}\rvert}\sqrt{\sum_A H^2/A}$
Gyrification Index [21]	$GC_H=\frac{1}{A\lvert\overline{H}\rvert}\sum_A c$
	$AF_i=A_i/A,\ i\in\{H^+,K^+\}$

Notation: $c = \sqrt{\frac{1}{2}(k_1^2+k_2^2)}$ (curvedness [10]), $s = \frac{2}{\pi}\arctan\frac{k_2+k_1}{k_2-k_1}$ (shape index [10]), $ak = |k_1|(|k_1| - |k_2|)$, $T = 3\frac{V}{A}$, and $|\overline{H}| = \frac{1}{A}|\sum_A H|$. $U \in [H^+, H^-]$, $V \in [K^+, K^-]$, and A_{H+} (or A_{K+}) is the area of the surface with positive H (or K) curvature.

In addition, we defined three new area-independent folding measures: $SH2SH$, $SK2SK$, and $AF_{\{H^+,K^-\}}$. The first two were based on the rationale that measures composed of ratios of local curvature factors would intrinsically eliminate the dependence on area. The latter was based on the idea that a reasonable characterization of the degree of folding in a brain surface is the fraction of the surface which contains convex folds (gyri). This can be mathematically characterized by looking at the relative portion of the surface which has positive mean or Gaussian curvature. For an undeveloped brain, the fraction would be

large, and it would progressively decrease as the concave folds (sulci) appear. The definition of the three expressions is shown in Table 1.

3 Method

The sphere example was useful to identify global folding measures that have the potential to consistently evaluate surface folding, independent of the size of surface of analysis. A more formal verification of this area-independence property is in progress. In this work we present an experimental comparison of the measures on two datasets. The first dataset consisted of 15 neonatal brains. The surfaces of this brains not only increase in size with age, but they also become more complex. So, folding varies proportionally to size. The second dataset consisted of iso-surfaces for which surface folding varied inversely to size. They were generated from one neonatal brain at various percentage occupancy thresholds (see Section 3.3). A larger threshold produces a smaller brain and larger separation between sulci walls, which translates to increased folding. For both datasets, gyrification was measured on left and right hemispheres, in addition to the whole brains. The infants in this study are normal, therefore gyrification is expected to be similar on both sides. Area-independent measures are expected to yield similar values for each hemisphere and the full-brain, but not so the area-dependent measures.

3.1 Data

Human Neonatal Images. High resolution ($0.703 \times 0.703 \times [1.5 - 2.2]$ mm) 3D T1 weighted SPGR images were acquired on premature infants using a 1.5T GE MRI scanner with an MRI compatible incubator. The gestational ages (GA) of the twelve premature infants in this study ranged between 24-31 weeks. The postmenstrual ages (gestational age + postnatal age) at the time of acquisition were 28-37 weeks. A subsequent scan for three infants was available, hence a total of 15 brains were processed.

3.2 Image Segmentation

The outer gray matter and the gray-white matter interface[2] surfaces of the premature brains were extracted semi-automatically as described in [16] using the rview software package [17]. Cortical gray matter was segmented for fifteen brains and gray-white matter interface for thirteen brains. A value of 1000 was assigned to brain voxels and 0 to background voxels.

[2] In the remaining of the text we will refer to the gray-white matter interface simply as white matter.

3.3 Computation of Brain Surface Folding

Each binary segmented brain was supersampled using voxel replication to prevent loss of fine scale features ($2 \times 2 \times 4$). A voxelwize approach to surface curvature estimation from iso-surfaces, derived from that developed by Rieger [14], was then employed. It avoids the need of a parametric model. A summary of the sequence of steps (described in detail in [16]) is as follows: 1) computation of the image gradient $g = \nabla f(x, y, z)$ of the replicated binary volume; 2) computation of the gradient structure tensor (GST) [14] defined as $T = \overline{gg^t}$; 3) calculation of the eigenvectors v_1, v_2, v_3 of T and the mapping $M(v_1) = \frac{v_1 v_1^t}{||v_1||}$; 4) on the 50% occupancy iso-surface, computation of the principal curvatures from $|k_{1,2}| = \frac{1}{\sqrt{2}}||\nabla_{v_{2,3}} M(v_1)||_F$ (Fröbenius norm); the sign of k_1 and k_2 is determined from the Hessian matrix of the image and the eigenvectors v_2 and v_3. The curvatures k_1 and k_2 were then used to compute mean (H) and Gaussian (K) curvature and the folding measures of Table 1. Each step involves image smoothing or differentiation. For all images, the 3D filters used at the ith step had a full-width at half maximum (f_i) of: $f_1 = 2.1$mm, $f_2 = 1$mm, $f_3 = 2$mm, $f_4 = 1$mm. The rationale behind the choice of values was to create a smooth surface that would yield smooth variations in curvature. Normalized Gaussian derivatives [11] were used for all computations.

The input to this algorithm is the binary replicated segmented brain, for which foreground voxels have a value of 1000 and background voxels a value of 0. Near the border of the brain the kernel incorporates brain and background voxels, hence the raw effect of the convolution on the image is that the resulting volume has voxel values in the range [0,1000]. The resulting voxel values are interpreted as the partial occupancy by brain tissue. The larger the number indicates that more foreground voxels were present in the kernel. For each brain, the surface on which curvature and the global measures of folding were calculated was taken to be the iso-surface with at least 50% occupancy (i.e. voxel values of at least 500) and satisfying 6-connectivity with the background.

3.4 Cortical Partitioning using Spatial normalization

Full-brains and their left and right hemispheres were processed with the algorithm just described. The procedure used to identify the left and right hemispheres in each neonatal brain was the following. A reference neonatal anatomy was manually partitioned into left and right hemisphere. A multi-resolution B-Spline based spatial normalization, adapted from that described in [18, 19] and as used in [2], was then applied to estimate a spatial transformation mapping from this reference to each subject MRI being studied. This transformation was then numerically inverted to allow the assignment of the nearest reference label to a given subject image voxel, which brought the voxelwize partitioning of the reference anatomy to the same space as the surface extracted from each infant. The partitioning was then used to constrain the voxelwize evaluation of the folding measures.

4 Results

We applied the measures to surfaces from the 15 MRI studies to investigate their response to folding with age. The relationship between age and folding was analyzed with regression and for this work a linear model was used. Due to space limitations results from only seven measures are shown in Figure 1. Top two rows correspond to gray matter results, and bottom two rows to white matter. The first observation is that measures with large goodness-of-fit for gray matter, had very small goodness-of-fit for white matter, and viceversa. This was true for the whole brain and both hemispheres. The same was observed in our results [15] on 10 MRI studies. The measures with greater goodness-of-fit for gray matter were the T-normalized measures: ICI_T, FI_T, $\overline{H^+}$, $\overline{K^-}$, $\overline{K^+}$; for white matter, GS (global shape index) and the $|\overline{H}|$-normalized measures: MLN_H and GC_H. The goodness-of-fit for gray matter and white matter (in that order) on whole surfaces (depicted as $*$ in the figure) was: CI_T (**0.85**, 0.004), FI_T (**0.85**, 0.19), GLN_T (0.80, 0.21), ICI_T (0.83, 0.002), MLN_T (0.80, 0.07), GS (0.61, **0.95**), $SK2SK$ (0.67, 0.46), $SH2SH$ (0.77, 0.25), MLN_H (0.48, **0.96**), GC_H (0.52, **0.95**), $\overline{H^-}$ (0.73, 0.07), $\overline{H^+}$ (**0.85**, 0.12), $\overline{K^-}$ (**0.85**, 0.27), $\overline{K^+}$ (same as ICI_T), AF_{H^+}=(0.62, 0.87), AF_{K^+}=(0.45, 0.13). The measures had similar goodness-of-fit scores for the left and right hemispheres. Area, volume, and T varied linearly in all cases.

Next we analyzed the dataset for which smaller brains were more folded. The iso-surfaces were created from a single brain (age 28 weeks) at percentage occupancy thresholds in the range 5-95% (in increments of 5%). An example of one coronal slice from five iso-surfaces obtained at five different thresholds is shown in Figure 2. The results for five measures on whole surfaces and their left and right hemispheres are shown in Figure 3. In contrast with the first dataset, the folding measures had similar goodness-of-fit for both gray and white matter (either both large or both small). Another difference was that FI_T presented a small goodness-of-fit. For this iso-surface dataset the area of gray matter varied in a non-linear fashion (plot not shown). This behavior is explained by the separation of the gray matter volume into two hemispheres, for larger thresholds (see Figure 2). The increase in surface area due to this separation is dominant over the relative decrease in area due to the larger threshold.

It can be observed that in the first dataset (Figure 1) the slopes of T-normalized measures have different tendencies for gray matter (positive) and white matter (negative). This difference is explained by the slope of T itself. For gray matter the rate of change of volume, with age, was faster than that of area, therefore $T = 3V/A$ had a positive slope. The inverse occurs for white matter. For the second dataset (Figure 3), the slope of T was negative for both gray and white matter, therefore the slope of all T-normalized measures is also negative. Once this slope difference is considered, it is clear that the new measures (T- and $|\overline{H}|$-normalized) change consistently with folding, independent of the size of the surface.

Results on both datasets for previously defined measures corroborate their dependency on the size of the surface of analysis. For each hemisphere, the

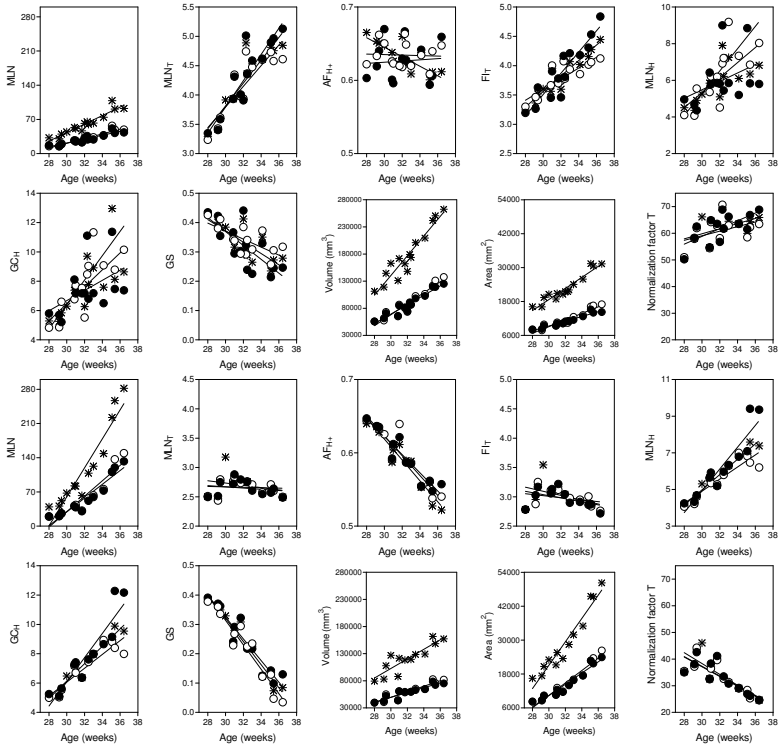

Fig. 1. Plots of seven folding measures applied to gray matter (top two rows) and white matter (bottom two rows). Filled shapes correspond to results for left hemisphere, and empty ones for right hemisphere. Results for the whole surface ($\star$) are also included.

magnitude of previously defined measures was at least half the value than for the whote-brain. In contrast, for the newly defined measures, their magnitude for whole-brain was in between that for left hemisphere and right hemisphere. As an example, results for MLN and GLN are shown in Figures 1 and 3, respectively.

Assessment of intra-observer segmentation variability (preliminary stage). The white matter of two brains, from the 15 available, was segmented and analyzed with an approach similar to the one in [1]. One brain was 29.1 weeks and was segmented four times; the other brain was 32.4 weeks and was segmented twice. Measures were computed on all the segmented surfaces. Given a set of segmentations for one brain, the segmentation variability was assessed as the maximum percentage variation of a measure from the mean value in the set. This was compared to the percentage change of the mean value of a measure between the 29.1 week brain and the 32.4 week brain. This number was taken as an indication of the extent of the variability in segmentations. An example of the results is given for MLN_H, GS, and AF_{H^+}. In that order, the maximum percentage variation for each measure on the 29.1 week old brain was: 2.81%,

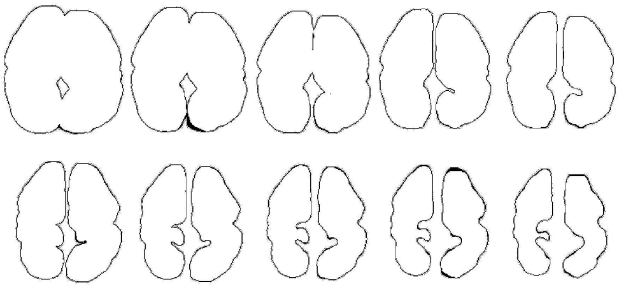

Fig. 2. A coronal slice each from five gray matter (top) and white matter (bottom) iso-surfaces extracted at percentage occupancy thresholds: 10%,30%,50%,70%,90%. All slices come from one brain.

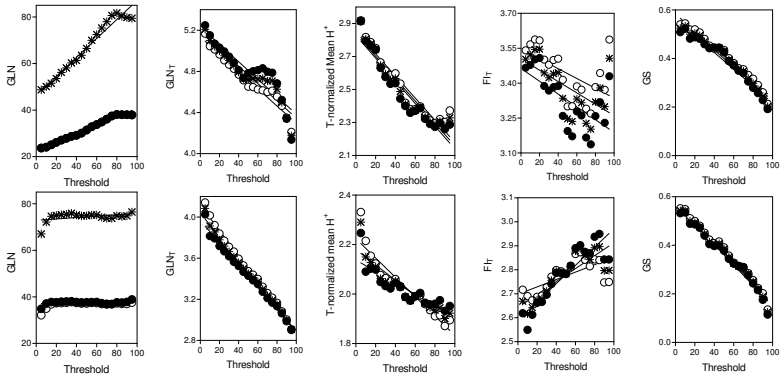

Fig. 3. Plots of five folding measures applied to iso-surfaces of a single brain at various occupancy thresholds. Top row shows results for gray matter and bottom row for white matter. Filled shapes correspond to left hemisphere and empty ones to right hemisphere. Results for the whole surface ($\star$) are also included.

2.57%, and 0.55%. The percentage difference between the 29.1 week brain and the 32.4 week brain was: 35.4%, 41.9%, 8.1%.

5 Discussion

An understanding of the cortical folding process in the development of premature infants may be important in explaining and predicting abnormal neurological outcome. The use of formal mathematical descriptions provides a more quantitative tool to study the folding process than is available with simple visual evaluation of MRI scans. The long term goal of our research is to create a model to track preterm *in vivo* neonatal brain cortical development that will help characterize normal gyrification and departures from it.

In this work we have shown that most previously proposed folding measures [1, 4] are dependent of the size of the surface patch on which they are calculated. To alleviate this problem we examined two approaches to area normalization of the measures and proposed new forms of measure, which satisfy the requirement of area-independence. These were evaluated on neonatal brain surfaces. We evaluated the measures ability to detect change in normal development by computing the linear regression of each measure with age. Finding the relationship between folding measures and age is an important part of the goal of understanding brain development. We will explore higher order models when more data is available.

In the dataset of 15 neonatal brains, folding varied proportionally to size. White matter folding tracked best with age, while no single measure rated consistently well for both gray and white matter surfaces. For the iso-surfaces dataset, where folding varied inversely to size, most measures rated well, and similarly for both gray and white matter. Despite differences in the behaviour of the normalization factor T and in the linearity of area on both datasets, most newly defined measures appeared to consistently assess folding. Recommended measures for gray matter are ICI_T, $\overline{H^+}$, $\overline{K^-}$; and for white matter GS, MLN_H, GC_H, and AF_{H^+}. The proposed normalization factors T and $|\overline{H}|$ opened the possibility to adapt previously defined measures to quantify gyrification in subregions of the brain. The slope of T appears to determine that of the T-normalized measures. Awareness of this is important for the correct interpretation of results. Alternative normalization factors need to be explored.

Preliminary results indicate that the effect of segmentation variability on the measures seems to be minimal: the differences in the measures with segmentation seems to be at least 10 times smaller than the differences with gestational age. We need to further investigate the sensitivity and validity of the measures in a larger cohort, containing both normal and pathological cases. In conclusion, the proposed new and normalized measures provide area-independent assessment of folding which provide the ability to study local gyrification. Such measures will provide a new tool in assessing global and local surface folding which is independent of the overall surface area, and are hence applicable to developing a model that tracks development in premature infants

Acknowledgments

This work was funded by the NIH grant R01 MH65392.

References

1. P. G. Batchelor, A. D. Castellano Smith, D. L. G. Hill, D. J. Hawkes, T. C. S. Cox, and A. F. Dean. Measures of folding applied to the development of the human fetal brain. *IEEE Transactions on Medical Imaging*, 21(8):953–965, August 2002.
2. V. Cardenas, L.L. Chao, R. Blumenfeld, and *et al.* Using automated morphometry to detect association between ERP latency and structural brain MRI in normal adults. *Human Brain Mapping*, 25(3):317–327, 2005.

3. R. W. I. Cooke, L. Foulder-Hughes, D. Newsham, and D. Clarke. Ophtalmic impairment at 7 years of age in children born very preterm. *Archives of Disease in Childhood Fetal and Neonatal Edition*, 89:249–253, 2004.
4. D.C. Van Essen and H. A. Drury. Structural and functional analyses of human cerebral cortex using a surface-based atlas. *The Journal of Neuroscience*, 17(18):7079–7102, 1997.
5. M. Hack, D. J. Flannery, and *et al.* Outcomes in young adulthood for very-low-birth weight infants. *New England Journal of Medicine*, 346(3):149–157, 2002.
6. B. E. Hamilton, J. A. Martin, and *et al.* Births: Preliminary data for 2004. *National Vital Statistics Reports*, 54(8), 2005.
7. P. S. Hüppi, B. Schuknecht, C. Boesch, and *et al.* Structural and neurobehavioral delay in postnatal brain development of preterm infants. *Pediatric Research*, 39(5):895–901, 1996.
8. T. E. Inder, S. K. Warfield, H. Wang, P. S. Hüppi, and J. J. Volpe. Abnormal cerebral structure is present at term in premature infants. *Pediatrics*, 115(2):286–294, 2005.
9. T. E. Inder, S. J. Wells, N. B. Mogridge, C. Spencer, and J. J. Volpe. Defining the nature of the cerebral abnormalities in the premature infant: a qualitative magnetic resonance imaging study. *Journal of Pediatry*, 143:171–179, 2003.
10. J. J. Koenderink and A. J. van Doorn. Surface shape and curvature scales. *Image and Vision Computing*, 10(8):557–564, 1992.
11. T. Lindeberg. On scale selection for differential operators. In *Proc. 8th Scandinavian Conference on Image Analysis*, pages 857–866, 1993.
12. M. Luciana. Cognitive development in children born preterm: implications for theories of brain plasticity following early injury. *Development and Psychopathology*, 15:1017–1047, 2003.
13. V. A. Magnotta, N. C. Andreasen, and S. K. Schultz *et al.* Quantitative *in vivo* measurement of gyrification in the human brain: changes associated with aging. *Cerebral Cortex*, 9:151–160, 1999.
14. B. Rieger, F. J. Timmermans, L. J. van Vilet, and P. W. Verbeek. On curvature estimation of iso surfaces in 3D gray-value images and the computation of shape descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1088–1094, August 2004.
15. C. Rodriguez-Carranza, P. Mukherjee, D. Vigneron, J. Barkovich, and C. Studholme. A system for measuring regional surface folding of the neonatal brain from MRI. In *9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2006)*, 2006.
16. C. Rodriguez-Carranza, F. Rousseau, B. Iordanova, O. Glenn, D. Vigneron, J. Barkovich, and C. Studholme. An iso-surface folding analysis method applied to premature neonatal brain development. In *Proc. SPIE Medical Imaging: Image Processing*, volume 6144, pages 529–539, 2006.
17. C. Studholme. http://rview.colin-studholme.net/.
18. C. Studholme, V. Cardenas, and *et al.* Accurate template-based correction of brain MRI intensity distortion with application to dementia and aging. *IEEE Transactions on Medical Imaging*, 23(1):99–110, 2004.
19. C. Studholme, V. Cardenas, and *et al.* Deformation tensor morphometry of semantic dementia with quantitative validation. *Neuroimage*, 21(4):1387–1398, 2004.
20. J. J. Volpe. Neurologic outcome of prematurity. *Archives of Neurology*, 55:297–300, 1998.
21. K. Zilles, G. Schlaug, M. Matelli, G. Luppino, and et al. The human pattern of gyrification in the cerebral cortex. *Anatomy and Embriology*, 179:173–179, 1988.

Realizing Unbiased Deformation: A Theoretical Consideration

A.D. Leow, M.C. Chiang, S.C. Huang, A.W. Toga, and P.M. Thompson

***Abstract*— Maps of local tissue compression or expansion are often recovered by comparing MRI scans using nonlinear registration techniques. The resulting changes can be analyzed using tensor-based morphometry (TBM) to make inferences about anatomical differences. Numerous deformation techniques have been developed, although there has not been much theoretical development examining the mathematical/statistical validity of each technique. In this paper, we propose a basic principle that any registration technique should satisfy: realizing unbiased test statistics under null distribution of the displacement. In other words, any registration technique should recover zero change in the test statistic when comparing two images differing only in noise. Based on this principle, we propose a fundamental framework for the construction and analysis of image deformation. Moreover, we argue that logarithmic transform is instrumental in the analysis of deformation maps. Combined with the proposed framework, this leads to a theoretical connection between image registration and other branches of applied mathematics including information theory and grid generation.**

***Index Terms*-Mutual information, Image registration, Computational anatomy.**

1. Introduction

Non-linear image registration is a well-established field in medical imaging with many applications in functional and anatomic brain mapping, image-guided surgery, and multimodality image fusion [1-8]. The goal of image registration is to align, or spatially normalize, one image to another. In multi-subject studies, this serves to reduce subject-specific anatomic differences by deforming individual images onto a population average brain template.

The deformations that map each anatomy onto a common template can be analyzed voxel-wise to make inferences about relative volume differences between the individuals and the template or statistical differences in anatomy between populations [9-13]. Similarly, in longitudinal studies it is possible to visualize structural brain changes that occur over time by deforming subjects' baseline scans onto their subsequent scans, using the deformation map to quantify local changes. This area of computational anatomy is known as *tensor-based morphometry* [11-15]. As there are numerous techniques for non-linear registration, one may ask whether some approaches are better than others in practice and/or in theory. In the following section, we will try to answer this fundamental question from a statistical standpoint.

2. Criteria for proper image registration

One could not study non-linear registration without closely examining the common practice of applying logarithmic transformation to Jacobian maps. Log-transformation of a Jacobian determinant field has become standard practice in most tensor-based morphometry (TBM) papers [11,12]. The Jacobian determinant of a diffeomorphic (smooth) map is bounded below by zero but unbounded above. Thus, at any voxel, its null distribution would be a better fit to a symmetric Normal distribution if the Jacobians are logged. Additional arguments that favor log-transformation of Jacobian maps come from the need to symmetrize the probabilities of expansions and shrinkages that are reciprocals of each other. We will discuss two different, yet related, concepts of symmetrizing; each supports the use of log transformation.

2.1. Two types of symmetry favor the log transformation of Jacobian maps

The first symmetry operates on a voxel level. The motivation behind this approach is the result of voxel-wise statistical tests. Consider, for example, testing for the presence of voxel-wise mean structural change in a multi-subject experiment. One might want to employ classical statistical approaches and test the null hypothesis of zero mean change using one-sample Student's *t* tests. A more general linear model typically relies on the residuals being a good approximation to a Gaussian random field. This null model assumes that at each voxel the observed rates of change over time (or relative volumes in a cross-sectional study) collected from all subjects follow a normal distribution, thus suggesting the use of log-transform – otherwise the Jacobian determinants are bounded below by zero but unbounded above.

The second type of symmetry is the symmetry of Jacobian distributions inside a region (in contrast, the symmetry previously described deals with multiple observations across subjects, at a single voxel). This second, albeit less intuitive, level of symmetry sheds light on how one might construct better non-linear image registrations.

In theory, a proper registration algorithm should produce unbiased estimators of the real anatomical change. An unbiased algorithm should detect no statistically significant change between two serial images if there is no difference other than noise, and statistics quantifying change collected from all voxels should ideally have a zero mean value; rates of change deviating from zero should be considered as errors. In a classical statistical setting, one might expect that rates of change thus follow a Gaussian distribution with zero mean, again justifying the use of the logarithmic transform.

These two types of symmetry operate at different levels: one across all voxels when comparing two images, and the other across all subjects at one single voxel. Assuming this symmetric log-normal distribution for the residuals in a statistical model for the Jacobian determinant, one symmetrizes the rate of change distribution by considering a halving or doubling of volume to be equally likely *a priori.* A related approach is taken by Pennec et al. [16], where the Cauchy-Green strain tensor [17, 18] of a deformation mapping is logged and used as a term in a penalty functional that is integrated over the image domain to regularize the deformation.

2.2. Detecting no change in the absence of real change as a necessary criterion for proper registrations: realizing unbiased test statistics

As noted previously, the logarithmic transform is fundamental in analyzing deformation maps through Jacobian maps, and an ideal registration algorithm should yield Jacobian maps and test

statistics that imply zero-change, when no difference is present between two registered images. We refer to this as the *principle of realizing unbiased test statistics under null distribution of the displacement.*

Before we show how to construct unbiased test statistics, we first define both *T* and *S*, on an image domain Ω, as the two images to be registered. Let us also assume, without loss of generality, that the volume of this domain is 1, i.e., $|\Omega| = 1$. We seek to estimate a transformation h such that *S* is non-linearly registered to *T* when deformed by *h* (i.e., $S(h(x))$). In this paper, we will restrict this mapping to be differentiable, one-to-one, and onto from the image domain onto itself [19] (in practice/implementation, the one-to-one and onto property can be approximated by extending the boundary towards infinity). Let us associate three probability density functions, defined on this image domain, to the identity mapping (id) as well as the deformation *h* and its inverse:

$$pdf_h(x) = |Dh(x)|;\ pdf_{h^{-1}}(x) = \left|Dh^{-1}(x)\right|;\ pdf_{id}(x) = 1. \quad \textbf{(1)}$$

Here, the Jacobian matrix of a transformation *h* is denoted *Dh*. As noted in [19], integrating the log of the Jacobian determinant over the image domain simply calculates the Kullback-Leibler distance between the identity map and *h*.

$$\int_\Omega \log|Dh(x)|dx = -KL(pdf_{id}, pdf_h) \le 0. \quad \textbf{(2)}$$

Here *KL*, the non-negative asymmetric Kullback-Leibler (*K-L*) distance, between two PDFs *p* and *q*, is defined as $KL(p,q) = \int_\Omega p \log\frac{p}{q} dx \ge 0;$

Given Eq. (2), one might ask if we could use the negative integral of the logged Jacobian determinant $-\int_\Omega \log|Dh(x)|dx$ as a basis for regularization. This integral only evaluates to zero when *h* is volume-preserving everywhere. Minimizing this integral gives us a volume-preserving map, and thus realizes unbiased statistics. On a global scale, this integral also evaluates the overall mean log-Jacobian with respect to the computation domain. Moreover, in the field of information theory, the KL distance is geometrically important, providing a means to compare probability density functions on general manifolds [20].
Interestingly, the integral Eq.(2) has skew-symmetry with respect to *h* and its inverse

$$\begin{aligned} KL(pdf_{id}, pdf_h) &= \int -\log|Dh|dx \\ = KL(pdf_{h^{-1}}, pdf_{id}) &= \int \left|Dh^{-1}\right| \log\left|Dh^{-1}\right| dx \end{aligned} \quad \textbf{(3)}$$

Notably, the idea of integrating the square root of the Jacobian, as proposed in [21], to remove the skew-symmetry, is equivalent to calculating the Bhattacharyya distance ***B,*** a well-known measure in information theory [21,22].

$$B(pdf_{id}, pdf_h) = \int |Dh(x)|^{1/2} dx \quad (4)$$
$$= \int |Dh^{-1}(x)|^{1/2} dx = B(pdf_{h^{-1}}, pdf_{id})$$

Surprisingly, notice that here the Bhattacharyya distance is symmetrical with respect to its two arguments, as well as inverse-consistent. To connect the KL-distance and Bhattacharyya distance, one can consider the geodesic linking of the two PDFs: $P(.,t)$, parameterized by time t:

$$P(x,t) = \frac{pdf_{id}(x)^t \, pdf_h(x)^{1-t}}{N}, \quad N = \int pdf_{id}{}^t(x) \, pdf_h(x)^{1-t} dx \quad (5)$$

The Bhattacharyya distance corresponds to the arbitrary choice of t=1/2, while a generalization of the above leads to the Chernoff distance in information theory [23].

2.3. Realizing unbiased deformation using symmetrical KL distance

To construct unbiased deformations, we first generalize eq. (2): given any diffeomorphism g mapping Region of interests S (with size a) to T (with size b), we have the following $\int_S \log|Dg(x)|\,dx \le \log b/a$, while equality holds if and only if the Jacobian of g takes the constant value b/a. The implication of this generalization is that, assuming ROI mapping from S to T, the Log operation is unbiased if and only if the corresponding Jacobian field is evenly distributed. Thus, minimizing the negative integral of logged Jacobian (treated as a cost function) as in 2.2 again leads to unbiased statistics. Given this generalization, let us now introduce the following by combining **Eq.(3)**, and its counterpart contributed by the inverse mapping:

$$KL(pdf_h, pdf_{id}) + KL(pdf_{h^{-1}}, pdf_{id}) = KL(pdf_h, pdf_{id}) + KL(pdf_{id}, pdf_h) \quad (6)$$
$$= KL(pdf_{id}, pdf_{h^{-1}}) + KL(pdf_{id}, pdf_h) = KL(pdf_{id}, pdf_{h^{-1}}) + KL(pdf_{h^{-1}}, pdf_{id})$$
$$= \int \left(|Dh(x)| - 1\right) \log |Dh(x)| dx = \int \left(|Dh^{-1}(x)| - 1\right) \log |Dh^{-1}(x)| dx$$

In this symmetric form, the integrand $\left(|Dh| - 1\right)\log|Dh|$ is always non-negative, compared to the integrand in **Eq. (3)** where locally negative numbers can be obtained (when the Jacobian is greater than zero), though globally non-negative. The unbiased property has two layers of meaning: the first one being realizing null distribution when comparing identical images differing in noise (from **Eq**. (2)); the second being realizing correct log Jacobian statistics in ROI analysis (from the generalization of **Eq**. (2)).

Under this framework, constructing deformations can be viewed as quantifying the symmetric KL distance between the identity map and the resulting deformation (or the inverse deformation due to its symmetry). Moreover, this framework embeds statistical analyses into the construction of deformations, penalizing deformations that skew the distribution of test statistics. A second interpretation of **Eq.(6)** is that it simply calculates the mean log Jacobian for h and its inverse inside the domain, thus encoding regional volume changes. Let us also point out an interesting observation by applying the square root Jacobian integrand as in **Eq.(4)** to the log-transformed Jacobian framework. A simple change of variables verifies the following:

$$\int \left(\log\left|Dh\right|\right)\left|Dh\right|^{1/2} dx + \int \left(\log\left|Dh^{-1}\right|\right)\left|Dh^{-1}\right|^{1/2} dx = 0 \qquad (7)$$

To further link **Eq.(6)** to other branches of mathematics, optimization problems involving Jacobian operator are commonly encountered in grid generation [24] and in continuum mechanics, where the Hencky tensor arises as logged tensor parameters in modeling very large deformations. However, we believe that the logarithmic transform has not been formally introduced in the grid generation literature and may also be useful there.

2.4. Logarithmic transform: a ubiquitous operation

With **Eq. (6),** realizing unbiased log-Jacobian values can be thought of as equivalent to minimizing the KL-distance. Moreover, in this case the concept of inverse-consistency translates to the symmetrization of the KL-divergence. The choice of the logarithmic operation in defining KL-divergence, though arbitrary at first sight, can now be easily justified by linking it to the statistical analyses of deformations in non-linear registration.

Finally, we comment briefly on how this framework can be implemented. Given an image matching cost function *C*, we seek, among all deformations minimizing this matching cost, the deformation with minimal symmetric KL distance as defined in Eq (6). In practice, this often means implementing a Lagrange multiplier, resulting in the following combined minimization problem

$$\arg\min_{h\in H} C(T, T\circ h^{-1}, S, S\circ h) + \lambda\left(KL(pdf_h, pdf_{id}) + KL(pdf_{id}, pdf_h)\right) \qquad (8)$$

Here λ is the multiplier, and H is the solution space, of which a common choice consists of all one-to-one, onto, and differentiable maps. Often, the solution is numerically obtained by recursive smoothing or regularization applied to the force field. In this paper, we focus on constructing a general principle that applies to different numerical approaches, and refrain from touching on the issue of regularization. The terms in **Eq.(6)** may also be viewed as regularizers, or deformation priors, as they penalize log-Jacobian values that deviate from zero. However, it is well-known in the grid generation field that an integral constraint on the Jacobian alone does not generally guarantee a globally smooth grid [24], so the smoothness of the resulting maps deserves further study.

Lastly, to complete our discussions on implementation issues, let us provide the gradient descent direction contributed by the symmetric KL distance term in eq(8), via its Euler-Lagrange equation. To this end, let us denote Co_{ij}, the matrix cofactor for the (i,j) th component of the Jacobian matrix Dh, we then obtain its Euler-Lagrange equation, using standard Calculus of variations with respect to the i-th coordinate as follows

$$\sum_j \frac{\partial}{\partial x_j}\left\{\left(1+\log|Dh(x)| - 1/|Dh|(x)\right)Co_{ij}(x)\right\} = 0, \qquad (9)$$
$$\left(Dh(x)\right)^{-1} = \left(Co_{ij}(x)\right)^T / |Dh(x)|.$$

Figure 1 shows a numerical example demonstrating the proposed approach implemented using a gradient descent projection method. Here, we matched the sequential MRI images (both of size 128 by 128 by 128) from a single subject diagnosed with semantic dementia. The figure showed three 2D slices, plotting the source, target, as well as the deformed source images. As in other TBM approaches, the technique is valuable as the pattern of atrophy is computed automatically, without interactive specification of regions of interest.

3. Conclusion

This paper is the first effort to systemically examine the relationship of image registration, information theory, and grid generation. While information-theoretic measures such as mutual information [25-28], *f*-divergence [29] and Jensen-Rényi divergence [13] have been popular measures to describe intensity correspondences in nonlinear image registration, it is much less common to appeal to statistical divergence measures in analyzing deformation fields. The proposed formulation calculates the KL-divergence between the deformation and the identity map, treating them as density functions defined on images. Unlike approaches employing conventional continuum mechanics (e.g., Lamé coefficients), our formulation is unbiased and parameter-free (the only parameter involved is the weight, which may be viewed as a Lagrange multiplier). The approach does not therefore make strong assumptions to model the deformation process. In contrast, the commonly employed elasticity theory assumes that displacements or velocity fields to conform to the law of elasticity, which most likely does not accurately describe any real brain deformation process over time.

The proposed framework helps explain the need for log-transformation of Jacobian values in TBM studies, which is ubiquitous and essential in analyzing tissue shrinkage/expansion. Symmetrization is also fundamental to securing inverse-consistency, an important property in image registration. Lastly, this framework is also consistent with a large-deformation approach [4,6,7], as any one-to-one and onto diffeomorphism remains in the solution space (since its Jacobian is positive and finite everywhere, and thus is theoretically attainable). In the future, we intend to further investigate this approach, which we believe provides a new perspective in non-linear registration and connects it to other fields of mathematics.

References

[1] P. M. Thompson and A. W. Toga, "A framework for computational anatomy," *Computing and Visualization in Science*, vol. 5, pp. 13-34, 2002.

[2] U. Grenander and M. I. Miller, "Computational anatomy: An emerging discipline," *Quarterly of Applied Mathematics*, vol. 56, pp. 617-694, 1998.

[3] M. K. Chung, K. J. Worsley, T. Paus, C. Cherif, D. L. Collins, J. N. Giedd, J. L. Rapoport, and A. C. Evans, "A unified statistical approach to deformation-based morphometry. *NeuroImage*, vol. 14, pp. 595–606, 2001.

[4] B. Avants and J. C. Gee, "Geodesic estimation for large deformation anatomical shape averaging and interpolation," *NeuroImage*, vol. 23, suppl. 1, S139-50, 2004.

[5] D. Shen, and C. Davatzikos, "Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration," *NeuroImage*, vol. 18, no. 1, pp. 28-41, 2003.

[6] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinematics," *IEEE Transactions on Image Processing*, vol. 5, no. 10, pp. 1435-1447, 1996.

[7] M. I. Miller, "Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms." *NeuroImage*, vol. 23, suppl. 1, pp. S19-S33, 2004.

[8] D. L. Collins, T. M. Peters, and A. C. Evans, "Automated 3D nonlinear deformation procedure for determination of gross morphometric variability in human brain," *Proc. SPIE* 2359, pp. 180-190, 1994.

[9] C. Studholme, V. Cardenas, N. Schuff, H. Rosen, B. Miller, and M. Weiner, "Detecting spatially consistent structural differences in Alzheimer's and fronto-temporal dementia using deformation morphometry," *Proc. MICCAI*, pp. 41-48, 2001.

[10] P.M. Thompson et al. (2004). *Mapping Cortical Change in Alzheimer's Disease, Brain Development, and Schizophrenia*, NeuroImage, 23 Suppl 1:S2-18, September 2004.

[11] P. M. Thompson, J. N. Giedd, R. P. Woods, D. MacDonald, A. C. Evans, and A. W. Toga, "Growth patterns in the developing brain detected by using continuum mechanical tensor maps," *Nature*, vol. 404, no. 6774, pp. 190-3, 2000.

[12] P. Cachier and D. Rey. "Symmetrization of the Non-Rigid Registration Problem using Inversion-Invariant Energies: Application to Multiple Sclerosis." *In MICCAI'00, LNCS* 1935:472-481, Pittsburgh, PA, Oct. 2000.

[13] A. D. Leow, J. C. Soares, K. M. Hayashi, A. D. Klunder, A. D. Lee, C. E. Bearden, E. S. Monkul, M. A. Nicoletti, A. P. Cerchiari, M. Trakhenbroit, P. Brambilla, R. B. Sassi, A. G. Mallinger , A. W. Toga, and P. M. Thompson, (2005). *Asymmetrical Effects of Lithium on Brain Structure Mapped in Healthy Individuals*, submitted for publication.

[14] A. D. Leow, A. D. Klunder, C.R. Jack, A.W. Toga, A.M. Dale, M.A. Bernstein, P.J. Britson, J.L. Gunter, C.P. Ward, J.L. Whitwell, B. Borowski, A. Fleisher, N.C. Fox, D. Harvey, J. Kornak, N. Schuff, C. Studholme, G.E. Alexander, M.W. Weiner, P.M. Thompson, For the ADNI Preparatory Phase Study (2005). "Longitudinal Stability of MRI for Mapping Brain Change using Tensor-Based Morphometry," *NeuroImage*, in press.

[15] J. Ashburner, J. Anderson, and K. Friston, "High-dimensional image registration using symmetric priors," *NeuroImage*, vol. 9, pp. 619-628, 1999.

[16] X. Pennec, R. Stefanescu, V. Arsigny, P. Fillard, and N. Ayache, "Riemannian elasticity: a statistical regularization framework for non-linear registration", *Proc. MICCAI 2005*, LNCS 3750, pp. 943-950, 2005.

[17] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Fast and simple calculus on tensors in the Log-Euclidean framework," Proc. *MICCAI 2005*, Palm Springs, CA, October 26-29, 2005.

[18] R. P. Woods, "Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation," *NeuroImage*, vol. 18 no. 3, pp. 769-88, 2003.

[19] Leow AD, Huang SC, Geng A, Becker JT, Davis SW, Toga AW, Thompson PM (2005). *Inverse Consistent Mapping in 3D Deformable Image Registration: Its Construction and Statistical Properties*, IPMI2005, Glenwood Springs, Colorado, July 11-15, 2005.

[20] N.N. Cencov. *Statistical Decision Rules and Optimal Inference*, Volume 14 of *Translations in Mathematics*. American Mathematical Society, Providence, RI, 1982

[21] M. Nielsen, P. Johansen, A Jackson, B Lautrup (2001). Statistical warps, a least committed model, MICCAI2001.

[22] A. Bhattacharyya. On a Measure of Divergence Between Two Statistical Populations Defined by their Probability Distributions. Bull. Calcutta Math. Soc., 35, 99 (1943)

[23] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[24] V.D. Liseikin, Grid Generation Methods. Springer-Verlag, Heidelberg, 1999.
[25] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens, "A viscous fluid model for multimodal non-rigid image registration using mutual information," *Medical Image Analysis*, vol. 7, pp. 565-575, 2003.
[26] B. Kim, J. L. Boes, K. A. Frey, and C. R. Meyer, "Mutual information for automated unwarping of rat brain autoradiographs," *NeuroImage*, vol 5, no. 1, pp. 31-40, 1997.
[27] P. Lorenzen, B. Davis, and S. Joshi, "Model based symmetric information theoretic large deformation multi-modal image registration," *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pp. 720-723, 2004.
[28] C. J. Twining, T. Cootes, S. Marsland, V. Petrovic, R. Schestowitz, and C. J. Taylor, "A Unified Information-Theoretic Approach to Groupwise Non-Rigid Registration and Model Building," *IPMI 2005.*
[29] J. P. W. Pluim, J. B. Antoine Maintz, and M. A. Viergever, "*f*-information measures in medical image registration," *IEEE Transactions on Medical Imaging* 23(12):1508-1516, 2004.

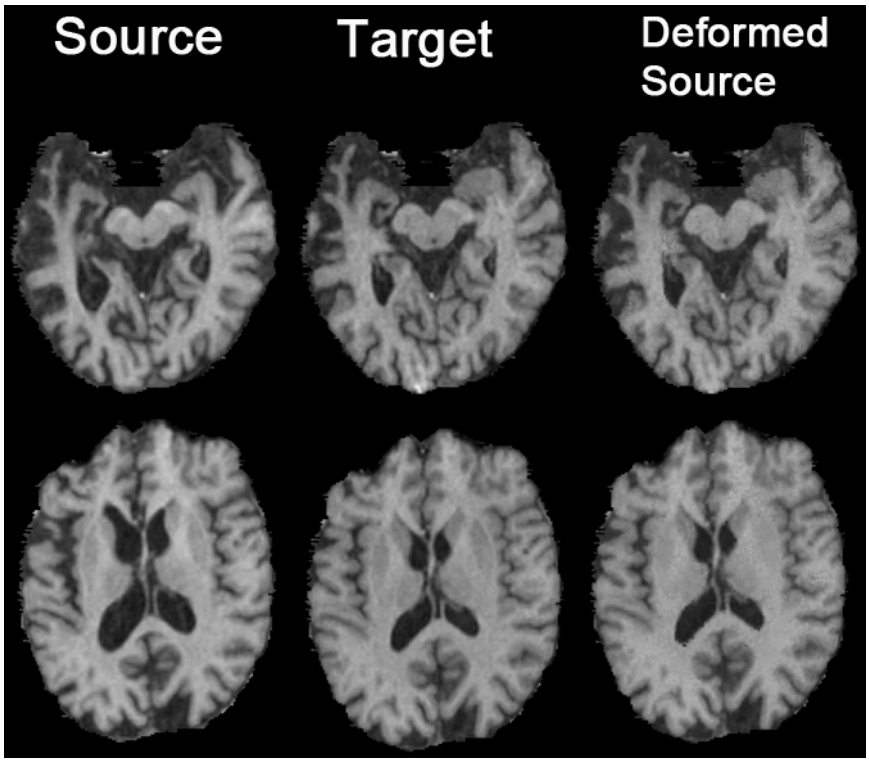

Fig. 1. Image registration using the unbiased deformation algorithm, with the sum of the squared intensity difference (SSD) as the cost function. 3D scans from a semantic dementia patient imaged at two time points were nonlinearly registered to estimate the profile of volumetric change. The later scan (first column) was chosen to be the source image. Progressive brain atrophy is observed in these axial slices.