Title
Shape Analysis Methods for 3D Brain and Skull Imaging

Author
Gutman, Boris Alexander

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Shape Analysis Methods for 3D Brain and Skull Imaging

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Biomedical Engineering

by

Boris Alexander Gutman

2013

ABSTRACT OF THE DISSERTATION

Shape Analysis Methods for 3D Brain and Skull Imaging

by

Boris Alexander Gutman

Doctor of Philosophy in Biomedical Engineering

University of California, Los Angeles, 2013

Professor Paul M. Thompson, Chair

Anatomical shape analysis problems are ubiquitous in medical imaging. In brain MRI imaging, the problem arises when comparing cortical features of functional importance, such as gray matter thickness, as well as when performing fine-grained analysis of sub-cortical structures. Two general types of approaches have been developed over the years to perform quantitative shape comparison of anatomy: the first, and more intuitive approach, attempts to bring surface models into dense point-to-point correspondence; the alternative approach avoids the need for dense registration by exploring intrinsic global properties of the shape. Each of these methods has their advantages and disadvantages, and I explore examples from each in

applications to brain imaging. The general underlying theme of my work has been spherical parametric shape registration, and other descriptions based on spherical maps. Several times throughout this work I utilize the theory of spherical harmonics, both in their scalar and vector form as a means of parameterizing and describing brain anatomy. Using this theory, I develop several algorithms for rigid and non-rigid shape registration for a variety of brain shapes, which are described in Chapters 1-4.

In Chapter 5, I present an application of the methods developed in previous chapters to developing brain imaging biomarkers of Alzheimer's disease. I show that using the new surface-based features and statistical learning, it is possible to develop biomarkers which outperform all others to date in terms of sensitivity to disease-modifying effects and disease specificity.

A more recent field of application for shape-based analysis comes from human skull models generated from conical CT imaging. The general registration problems are similar in spirit to brain imaging. However, the much lower signal-to-noise ratio and greater topological variability of the shape models require a new set of tools to deal with these shapes in practice. In Chapter 6, I develop an approach to modify the topology of a skull surface while preserving useful features, and again return to the sphere for dense registration and the creation of shape atlases. Due to the noisy nature of the data, a fast non-local non-linear correspondence search is developed for pairs of skull shapes from different subjects. I show that this search is crucial for high-quality registration based on resulting population-based averages and individualized shape analyses.

The dissertation of Boris Alexander Gutman is approved.

Won Moon

David W. Shattuck

Daniel B. Ennis

Luminita Vese

Arthur W. Toga

Paul M. Thompson, Committee Chair

University of California, Los Angeles

2013

To my uncle Yura

TABLE OF CONTENTS

LIST OF FIGURES

Chapters 1 - 4

Chapter 5

Chapter 6

LIST OF TABLES

Chapters 1 - 4

Chapter 5

ACKNOWLEDGEMENTS

I am grateful to all those who have helped me grow during my time as a doctoral student.

Funding Agency Acknowledgements

Projects were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (NIH Grant U01 AG024904). ADNI is funded by the NIA, NIBIB, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as nonprofit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. FDA. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at UCSD. ADNI data are disseminated by LONI. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation. Algorithm development and data collection were also funded by the NIH - NIA, NIBIB, NICHD, NLM, NCRR AG040060, MH097268, MH089722, EB01651, LM05639, RR019771, EB008432, EB008281, and EB007813, to Paul Thompson.

Co-author Acknowledgements

I would like to thank all of the following co-authors for their contributions. These articles are reprinted here in original or modified from with permission.

B. A. Gutman, S. K. Madsen, A. W. Toga, and P. M. Thompson, "A Family of Fast Spherical Registration Algorithms for Cortical Shapes," in *Multimodal Brain Image Analysis 2013*., accepted for **oral presentation**, Nagoya, Japan Sept. 22, 2013.

B. A. Gutman, X. Hua, P. Rajagopalan, Y.-Y. Chou, Y. Wang, I. Yanovsky, A. W. Toga, C. R. Jack Jr, M. W. Weiner, and P. M. Thompson, "Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features," *Neuroimage*, vol. 70, pp. 386-401, 2013.

B. Gutman, R. McComb, J. Sung, W. Moon, and P. Thompson, "Robust Shape Correspondence via Spherical Patch Matching for Atlases of Partial Skull Models," in *Mesh Processing in Medical Image Analysis 2012*. vol. 7599, J. Levine, R. Paulsen, and Y. Zhang, Eds., ed: Springer Berlin Heidelberg, 2012, pp. 89-100.

B. A. Gutman, W. Yalin, P. Rajagopalan, A. W. Toga, and P. M. Thompson, "Shape matching with medial curves and 1-D group-wise registration," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, 2012, pp. 716-719.

B. Gutman, Y. Wang, J. Morra, A. W. Toga, and P. M. Thompson, "Disease classification with hippocampal shape invariants," *Hippocampus*, vol. 19, pp. 572-8, Jun 2009.

B. Gutman, Y. Wang, L. M. Lui, T. F. Chan, P. M. Thompson, and A. W. Toga, "Shape Registration with Spherical Cross Correlation," *MICCAI Workshop on Mathematical Foundations in Computational Anatomy (MFCA '08),* 2008.

B. Gutman, Y. Wang, L. Lui, T. Chan, P. Thompson, "Hippocampal Surface Discrimination via Invariant Descriptors of Spherical Conformal Maps," 4th IEEE International Symposium on Biomedical Imaging (**ISBI '07**): 1316 – 1319, April 12-15, 2007

VITA

EDUCATION:

University of California, Los Angeles	B.S.	**Applied Mathematics**	2003-2006

RESEARCH AND PROFESSIONAL EXPERIENCE:

Graduate Student Researcher, UCLA Lab. of Neuro Imaging	2009-2013
Programmer analyst, UCLA Lab. of Neuro Imaging	2007-2009
UCLA Research Experience for Undergraduates program	2005-2006

HONORS/AWARDS:

Staff Achievement Award, Lab. of Neuro Imaging	2009
Graduated UCLA with honors	2006
Best visualization software, CCB All-Hands Meeting	2006
All-American Scholar	2002

FELLOWSHIPS:

UCLA Graduate Division Unrestricted Fellowship	2011
NIH Neuro Imaging Training Program Graduate Fellowship	2010
National Science Foundation Research Experience for Undergraduates	2006
National Science Foundation Research Experience for Undergraduates	2005

SELECTED PUBLICATIONS:

[1] **B. A. Gutman**, Y. Wang, I. Yanovsky, X. Hua, A. W. Toga, C. R. Jack, M. W. Weiner, and P. M. Thompson. “Empowering Imaging Biomarkers of Alzheimer’s Disease,” *Neurobiology of Aging*, in review

[2] **B. A. Gutman**, S. K. Madsen, A. W. Toga, and P. M. Thompson, "A Family of Fast Spherical Registration Algorithms for Cortical Shapes," in *Multimodal Brain Image Analysis 2013.*, accepted for **oral presentation**, Nagoya, Japan Sept. 22, 2013

[3] **B. A. Gutman**, X. Hua, P. Rajagopalan, Y.-Y. Chou, Y. Wang, I. Yanovsky, A. W. Toga, C. R. Jack Jr, M. W. Weiner, and P. M. Thompson, "Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features," *Neuroimage,* vol. 70, pp. 386-401, 2013.

[4] **B. Gutman**, R. McComb, J. Sung, W. Moon, and P. Thompson, "Robust Shape Correspondence via Spherical Patch Matching for Atlases of Partial Skull Models," in *Mesh Processing in Medical Image Analysis 2012*. vol. 7599, J. Levine, R. Paulsen, and Y. Zhang, Eds., ed: Springer Berlin Heidelberg, 2012, pp. 89-100.

[5] **B. A. Gutman**, W. Yalin, P. Rajagopalan, A. W. Toga, and P. M. Thompson, "Shape matching with medial curves and 1-D group-wise registration," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, 2012, pp. 716-719.

[6] X. Hua*, **B. Gutman***, C. P. Boyle, P. Rajagopalan, A. D. Leow, I. Yanovsky, A. R. Kumar, A. W. Toga, C. R. Jack, Jr., N. Schuff, G. E. Alexander, K. Chen, E. M. Reiman, M. W. Weiner, and P. M. Thompson, "Accurate measurement of brain changes in longitudinal MRI scans using tensor-based morphometry," *Neuroimage,* vol. 57, pp. 5-14, Jul 1 2011.

* – *these authors contributed equally*

[7] **B. Gutman**, Y. Wang, J. Morra, A. W. Toga, and P. M. Thompson, "Disease classification with hippocampal shape invariants," *Hippocampus,* vol. 19, pp. 572-8, Jun 2009.

[8] **B. Gutman**, Y. Wang, L. M. Lui, T. F. Chan, P. M. Thompson, and A. W. Toga, "Shape Registration with Spherical Cross Correlation," *MICCAI Workshop on Mathematical Foundations in Computational Anatomy (MFCA ’08),* 2008.

INTRODUCTION

As image acquisition becomes ever-more widespread in medical applications, there is an increasing need for more accurate, faster image processing methods focused on capturing and quantifying anatomy. Clinical brain imaging studies in particular now involve processing of hundreds, and sometimes thousands of images for statistical analysis, with the hope of detecting disease effects, effects of growth and aging, etc. One of the most salient aspects of anatomy is the geometry of the surfaces which form the boundaries of important, functionally distinct anatomical regions. Thus, anatomical shape analysis was born as a discipline to supplement existing standard 3D image analysis approaches [1, 2]. In addition to improved visualization, such analyses are also typically faster than dense 3D-based approaches due to the reduced dimensionality of the problem. The general underlying theme of my work is spherical parametric shape registration, and other descriptions based on spherical maps. Several times throughout this work I utilize the theory of spherical harmonics, both in their scalar and vector form as a means of parameterizing and describing brain anatomy. Using this theory, I develop several algorithms for quantifying shape, which are described in Chapters 1-4, and 6.

Shape-based approaches can be crudely categorized into two categories: the first, and more intuitive approach, attempts to bring surface models into dense point-to-point correspondence; the alternative approach avoids the need for dense registration by exploring intrinsic global properties of the shape. Each of these methods has their advantages and disadvantages, and I explore examples from each in applications to brain imaging. While the global shape signatures have the benefit of being invariant to Euclidean motion, as is desired for analyzing shape, they generally do not lend themselves to reconstruction, and are therefore

theoretically not unique to the particular shape [3-5]. Further, there is often some difficulty in visualizing the meaning of such description from an anatomical perspective. I show one possible means of visualizing an invariant description in Chapter 2, while also showing that the shape signature coupled with supervised machine learning can distinguish between a diseased subject and a control in an Alzheimer's study.

The more intuitive approach – one that involves registering surfaces for dense correspondence – is typically more computationally challenging. Further, simply having a correspondence does not immediately lead to a local description of the shape. One must design a measure of shape that is appropriate for the particular anatomy. In Chapter 1 I develop a robust non-local registration approach for bringing general shapes into rough alignment, improving on prior work [6, 7]. In Chapter 3 I return to this problem for the case of subcortical structures, and develop an improvement on existing medial core methods [8-10] for shape description and non-linear registration.

In Chapter 4, I continue the effort to improve shape registration, in this case delving into the problem of matching the complex folding patterns of the human cortex. My approach is again to use the spherical domain, this time by adopting several well-known Euclidean image registration algorithms [11, 12] to the 2-sphere. I develop a technique for smoothing spherical tangential fields, which are used to encode the non-linear deformations which match shapes. This fast spherical vector smoothing proves crucial for the registration process. The resulting algorithms are the fastest spherical adaptation-type algorithms to date [13], and are shown to accurately match folding patterns across subjects.

In Chapter 5, I present an application of the methods developed in previous chapters to developing brain imaging biomarkers of Alzheimer's disease (AD). Alzheimer's disease, a neurodegenerative disorder associated with neurofibrillary tangles in the brain, is quickly becoming one of the costliest diseases in the developed world. AD-associated costs in the U.S. approach $100 billion [14]. As a result, there is much interest in developing disease-slowing drugs to help reduce this cost, which requires well-formed cognitive as well as biological biomarkers of the disease which are sensitive to drug effects. Biomarkers of Alzheimer's disease based on brain imaging must offer relatively high power to detect longitudinal changes in subjects scanned repeatedly over time [15-17]. Much research has been conducted towards this end in recent years, focusing mainly on developing core morphometric techniques aimed at longitudinal studies. While continuing this effort, I also address the need to combine the measures into one salient "atrophy score," that is both interpretable and allows for reduced sample sizes in clinical studies. I show that using the new surface-based features and statistical learning, it is possible to develop biomarkers which outperform all others to date in terms of sensitivity to disease-modifying effects and disease specificity [18].

A more recent field of application for shape-based analysis comes from human skull models generated from conical CT imaging. The general registration problems are similar in spirit to brain imaging. However, the much lower signal-to-noise ratio and greater topological variability of the shape models require a new set of tools to deal with these shapes in practice. In Chapter 6, I develop an approach to modify the topology of a skull surface while preserving useful features, and again return to the sphere for dense registration and the creation of shape atlases [19]. Due to the noisy nature of the data, a fast non-local non-linear correspondence search is developed for pairs of skull shapes from different subjects. I show that this search is

crucial for high-quality registration based on resulting population-based averages and individualized shape analyses.

CHAPTER 1

Initial Shape Registration: Rigid Alignment on the Sphere

This chapter is adapted from:

B. Gutman, Y. Wang, L. M. Lui, T. F. Chan, P. M. Thompson, and A. W. Toga, "Shape Registration with Spherical Cross Correlation," *MICCAI Workshop on Mathematical Foundations in Computational Anatomy (MFCA '08),* 2008.

Abstract

We present a framework for shape alignment that generalizes several existing methods for parametric registration. We assume that the shape is a closed genus zero surface. Our framework requires a diffeomorphic surface mapping to the 2-sphere which preserves rotation. The registration is based on maximizing spherical cross-correlation function of surface-intrinsic scalar attributes, weighted by the cross-correlation of the parameterization distortion. The similarity measure may be customized according to the surface-intrinsic scalar functions used in the application. The optimization is global and non-iterative, leading to fast running times.

1.1 Introduction

Problems of shape alignment are ubiquitous in medical imaging. While many problem-specific solutions exist for particularly common cases (e.g. cortex, hippocampus) [20-22], high quality

general shape alignment remains an open problem. Our framework generalizes some existing methods without assuming the existence of any landmarks or data-specific features. We assume that the shape is a closed genus zero surface. Our framework requires a diffeomorphic surface mapping to the 2-sphere which preserves rotation. Our similarity measure is a global spherical cross-correlation function of surface-intrinsic scalar attributes, weighted by the cross-correlation of the parameterization distortion, sometimes known as the conformal factor. The final similarity measure may be customized according to the surface-intrinsic scalar functions used in the application. Higher order scalar functions such as mean and Gaussian curvature may be used in conjunction with low order ones, like distance to mass center, to incorporate more localized shape information in addition to global measures. In this study we have used a global conformal mapping as the spherical parameterization, and only the distance to mass center as intrinsic scalar shape measure. Using our method, we created atlases and registered shapes from a population of hippocampi.

1.2 Previous Work

There is doubtless a galaxy of existing general shape alignment methods, and for the sake of brevity we will only mention those most prevalent and those closest to the present work. Davies [23] has used an information-theoretic framework to formulate shape alignment as a minimum description length (MDL) problem. Here, each point is treated as an independent variable, while the cost function is aimed at reducing the "code length" of each shape's representation in this shape space. The beauty of this statistical approach lies in its ability to register multiple shapes simultaneously without the need to select a "target" shape. Of course, this is also its limitation, as

for example when a known atlas exists and all data shapes are to be registered to it. The method requires a spherical diffeomorphism like ours.

Variants of the well-known ICP algorithm are another variety of recent developments in rigid shape alignment. Granger [7] introduced the EM-ICP method. ICP's proclivity for terminating at suboptimal local minima is greatly reduced by treating the problem as a general expectation-maximization problem. A multi-scale approach is used: at coarser scales, the blurring factor is sufficient to give crude but correct alignment, which is improved upon in later refined stages, where the algorithm approaches the original ICP. Though quite robust, the method depends on the scale of blurring factor being set correctly to avoid local minima. Thus, avoiding them is not guaranteed. Our method, by contrast, performs a global search non-iteratively. Thus, a global maximum correlation is guaranteed irrespective of the shape's original orientation without the need to tune any parameters.

Much like our algorithm, some previous methods have used spherical harmonics for rigid shape alignment. Among them are the first order ellipsoid (FOE) method, popularized by Brechbuhler [6] and used extensively in medical imaging [10, 24] applications and SHREC, a recent variant of the ICP algorithm. Like ours, these methods make use of rotational properties of spherical harmonics.

FOE alignment uses the fact that a shape reconstructed from only the first order spherical harmonics forms an ellipsoid in object space. The method works well when the ellipsoid's three axes have distinct lengths, which largely depends on the shape itself and the degree to which the spherical parameterization pre serves area. The method gives only a crude alignment and fails when two or more axes have similar lengths. Even with a proper ellipsoid, there is a symmetry problem.

SHREC [25] is another variation of the ICP. As in our case, the correspondence search is done iteratively on the sphere via Euler's rotation formula and icosahedral subdivision. The mapping satisfies our conditions, while the similarity measure is the RMSD. Since RMSD depends on the position of the object in space, the algorithm requires an initial pre-alignment in both spaces. Rigid Quaternion Transform (RQT) is used to align shapes in $\boldsymbol{R^3}$ after each iteration of parametric alignment. The main limitation of this algorithm is that it is not guaranteed to converge to the optimal solution (i.e. a correspondence which, when applied to RQT, minimizes RMSD). This is because the optimization in parameter space depends on the object's position in native space. Thus, though the search is more global than in the original ICP, the parameter space search is still locally biased. The original ICP suffers the same problem.

The last algorithm uses a brute force correlation: it computes a cost function anew for each rotation. To mitigate the cost of this, a hierarchical approach is used. Instead, we reduce parameter space alignment to a global refined search via the FFT. Our numerical scheme separates the effects on computation time of the level of detail used for alignment and the number of rotation samples. In SHREC, these two are tied together owing to their brute-force nature. This means that we can refine rotation space tessellation while maintaining the same level of surface detail without significantly affecting computation time. Our use of orientation-invariant shape attributes in conjunction with scale invariant cross-correlation makes our approach completely independent of changes in object position and size. SHREC, by contrast, requires volume normalization as a pre-processing step.

1.3 Mathematical Preliminaries

1.3.1 Spherical Harmonics

Spherical harmonics are functions $f: \mathbb{S}^2 \to \mathbb{C}$ which are simultaneously eigenfunctions of the Laplace-Beltrami and the angular momentum operators; they are expressed explicitly as

$$Y_l^m(\theta, \phi) = (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos\theta) e^{im\phi} \tag{1.1}$$

for degree $m, l \in \mathbb{Z}, |m| \leq l$, where $P_l^m(x)$ is the associated Legendre polynomial. Spherical harmonics form a countable orthonormal basis for square-integrable functions on the sphere. A projection of a function $f \in L^2(\mathbb{S}^2)$ onto this basis yields the SPH coefficients

$$\widehat{f_{l,m}} = \langle f, Y_l^m \rangle, \tag{1.2}$$

where $\langle f, g \rangle$ is the usual L^2 inner product.

A key property of spherical harmonics is their behavior under a shift on the sphere. Given an element of the rotation group $SO(3) \ni R = R(\alpha, \beta, \gamma)$, a rotated spherical harmonic is expressed as

$$Y_l^m(\omega) = \sum_{|n| \leq l} Y_l^n(R^{-1}\omega) D_{m,n}^l(R), \tag{1.3}$$

where

$$D_{m,n}^l(R) = e^{-i(m\alpha + n\gamma)} d_{m,n}^l(\beta), \tag{1.4}$$

α, β, γ are the Euler angles of R and $d_{m,n}^l$ are irreducible representations of $SO(3)$ [26],

$$d_{m,n}^l(\beta) = \sum_t (-1)^t \times \frac{\sqrt{(l+n)!\,(l-n)!\,(l+m)!\,(l-m)!}}{(l+n-t)!\,(l-m-t)!\,(t+m-n)!\,t!} \tag{1.5}$$

$$\times (\cos\beta/2)^{2l+n-m-2t} (\sin\beta/2)^{2t-n+m}.$$

In particular, (1.3) implies that

$$f(\omega) = g(R^{-1}\omega) \Longrightarrow \widehat{f_{l,m}} = \sum_{|n|\leq l} \widehat{g_{l,n}} D^{l}_{m,n}(R). \tag{1.6}$$

1.3.2 Discrete spherical cross correlation

Given two functions $f, g \in L^2(\mathbb{S}^2)$, their spherical cross-correlation is defined as

$$C_{f,g}(R) = \int_{\mathbb{S}^2} f(\omega) g(R^{-1}\omega) d\omega. \tag{1.7}$$

In the special case where $f(\omega) = g(R'^{-1}\omega)$, $C_{f,g}(R)$ is maximized when $R' = R$, assuming that f is not spherically symmetric. For bandlimited functions, i.e. for those functions whose spherical harmonic coefficients vanish for all $l \geq B$ for some bandwidth B, the correlation becomes

$$C_{f,g}(R) = \sum_{\substack{|m|\leq l \\ l<B}} \hat{f}_{l,m} \overline{\widehat{\Lambda(R)} g_{l,m}}, \tag{1.8}$$

Here, $\Lambda(R)$ is the operator associated with the rotation matrix. The expression for shifted spherical harmonic coefficients (1.6) implies that

$$C_{f,g}(R) = \sum_{l,m,n} \hat{f}_{l,m} \overline{\hat{g}_{l,n}} D^{l}_{m,n}(R), \tag{1.9}$$

This expression forms the basis of our similarity measure.

1.3.3 Fast cross correlation via FFT

The material presented so far has been used in the prior works we mentioned. Now, we present a simple lemma which leads to a great speed up in computing the correlation (9). It suffices to make the observation that any rotation $R(\alpha,\beta,\gamma)$ may be expressed as a product of two rotations:

$$R(\alpha,\beta,\gamma) = R_1\left(\alpha+\frac{\pi}{2},\frac{\pi}{2},0\right)R_2\left(\beta+\pi,\frac{\pi}{2},\gamma+\frac{\pi}{2}\right). \tag{1.10}$$

Now using the fact that

$$D^l_{m,n}(R_1\circ R_2) = \sum_k D^l_{m,k}(R_1)D^l_{k,n}(R_2), \tag{1.11}$$

we substitute (1.10) into (1.9) to obtain

$$\begin{aligned} C_{f,g}(R) &= \sum_{l,m,n,k} \hat{f}_{l,m}\overline{\hat{g}_{l,n}} \qquad (1.12)\\ &\times D^l_{m,k}\left(\alpha+\frac{\pi}{2},\frac{\pi}{2},0\right)D^l_{k,\mathrm{n}}\left(\beta+\pi,\frac{\pi}{2},\gamma+\frac{\pi}{2}\right)\\ &= \sum_{l,m,n,k} \hat{f}_{l,m}\overline{\hat{g}_{l,n}}\; d^l_{m,k}(\frac{\pi}{2})d^l_{k,\mathrm{n}}(\frac{\pi}{2})e^{-i\left(m\left[\alpha+\frac{\pi}{2}\right]+k[\beta+\pi]+n\left[\gamma+\frac{\pi}{2}\right]\right)}\\ &= IFFT\left\{\sum_l \hat{f}_{l,m}\overline{\hat{g}_{l,n}}\; d^l_{m,k}\left(\frac{\pi}{2}\right)d^l_{k,\mathrm{n}}\left(\frac{\pi}{2}\right)\right\}(\alpha+\frac{\pi}{2},\beta+\pi,\gamma+\frac{\pi}{2}) \end{aligned}$$

This simple result has been shown elsewhere [27], but to the best of the authors' knowledge this is the first time it has been used for shape registration.

1.4 Shape registration with cross correlation

1.4.1 Similarity Measure

Given a 2-manifold M ⊂ R3, a diffeomorphic spherical parameterization $f: \mathbb{S}^2 \to \mathcal{M}$ and a family of rotation-invariant shape attributes $s_i: \mathcal{M} \to \mathbb{R}, 0 < i \leq N$, let

$$S_i = s_i \circ f. \tag{1.13}$$

Then, given two manifolds $\mathcal{M}_1, \mathcal{M}_2$ and their corresponding shape attribute map $\{S_{1,i}\}_{i=1}^{N}$, $\{S_{2,i}\}_{i=1}^{N}$, we define our shape similarity measure as

$$\mathcal{S}_{\mathcal{M}_1,\mathcal{M}_2}(R) = C_{\lambda_1,\lambda_2}(R) \sum_{i=1}^{N} \kappa_i \, C_{S_{1,i},S_{2,i}}(R), \tag{1.14}$$

where κ_i are user-defined shape attribute weights, and λ_1, λ_2 are spherical maps of the conformal factor of each manifold. These last two are used to mitigate the fact that scalar functions which appear similar on the sphere may in fact represent vastly differently-sized regions on the original surfaces due to varying area distortion of the spherical map. Because we recover shifts in object space with shifts on $\mathbb{S}^2$, we require that the spherical parameterization preserve rotation in the following sense. Suppose $\mathcal{M}_2 = R \circ \mathcal{M}_1$ and $f_1, f_2: \mathbb{S}^2 \to \mathcal{M}_1, \mathcal{M}_2$ are spherical maps. Then,

$$f(\omega) = R \circ f_1(R^{-1}\omega). \tag{1.15}$$

Many existing parameterizations satisfy this requirement, e.g. [6, 28].

1.4.2 Previous methods as special cases

SHREC and FOE are special cases of our method. FOE simply takes the conformal factor λ to be constant and uses spherical harmonics up to order one only. This is equivalent to setting the bandwidth *B* to 2. The single shape attribute used is the Euclidian distance to the surface average

value. This is roughly the same as distance to mass center, especially for area-preserving spherical maps with which FOE is typically used.

SHREC minimizes RMSD, which can be reduced to spherical cross-correlation since $4\pi \boldsymbol{RMSD}^2_{\mathcal{M}_1,\mathcal{M}_2}(R) =$

$$\sum_{i\in\{x,y,z\}} \int_{\mathbb{S}^2} \|S_{1,i}, - \Lambda(R) S_{2,i}\|^2 \tag{1.16}$$

$$= \sum_{i\in\{x,y,z\}} \int_{\mathbb{S}^2} \left\{S_{1,i}^2 + \left[\Lambda(R) S_{2,i}\right]^2 - 2 S_{1,i} \Lambda(R) S_{2,i}\right\}$$

$$= \sum_{i\in\{x,y,z\}} \|S_{1,i}\|_2^2 + \|S_{2,i}\|_2^2 - C_{S_{1,i},S_{2,i}}(R).$$

Thus, minimizing RMSD is equivalent to maximizing the correlation of the two shapes' spatial coordinates. These are, of course, not quite the scalar shape attributes we intend to use in our similarity measure. SHREC's dependence on iterative RQT refinement for correspondence optimization makes it less robust. This is the price of using orientation-dependent features. Further, it is not clear whether the correspondence which, when applied to RQT, minimizes RMSD is truly the best correspondence. One can conceive two shapes with some patches quite similar and other very different. One may then like to align the two objects to get the best correspondence between the similar patches without regard to the different ones. In such a case, cross correlation of invariant features will achieve a better alignment. Still, SHREC could be made faster with the use of FFT-based correlation rather than a brute-force approach.

SHREC requires $\mathcal{O}(B^3 N_R)$ operations, where B is again the bandwidth and NR the number of rotation samples. This is because recomputing $D^l_{m,n}(R)$ and the corresponding shifted spherical harmonic coefficients requires $\mathcal{O}(B^3)$ operations. Our method requires $\mathcal{O}(B^4 + N_R log N_R)$. N_R is roughly also of the order B^4 even with hierarchical sampling; this means we have effectively reduced the order of operations from $\mathcal{O}(B^6)$ to $\mathcal{O}(B^4)$. This allows us to sample

rotation space more finely without a significant change in execution time. It also makes using higher order coefficients and hence greater level of detail for alignment purposes feasible. We see this in experiments below.

1.5 Experimental Results

We used a population of 45 right hippocampal surfaces extracted from healthy elderly subjects. Our spherical parameterization was the global conformal map of Gu. et al. [28]. Spherical harmonics were computed with the spherical FFT of [29] and cross correlation was computed with the help of the FFTW library [30]. *SO(3)* was sampled at 200 X 100 X 200 samples, which yielded an angle frequency of ~ 1.8 degrees. A bandwidth of 64 was used throughout the testing. Only the distance to mass center was used as an invariant shape attribute for both populations.

As a preliminary experiment, we applied the cross correlation algorithm to a pair of hippocampal surfaces shown in **Figure 1A and 1B**. Here we see that the initial spherical maps do not provide a very good correspondence. **Figure 1E** shows the result of a spherical shift based on cross correlation. **Figure 2** shows two point-wise averages of the shapes, before and after cross correlation. The improvement is obvious. For B = 64, the average running time for above experiment was 44.6 $\pm$3.6 seconds on a Gateway 7426GX Laptop with a 2.41 GHz AMD processor and 1 GB RAM, tested with 45 hippocampal surfaces. This includes computation of spherical harmonics, rotation matrices $d^l_{m,n}$ and correlation. Shen [25] reports an average of 23.5 seconds running time on a common laptop for SHREC, while using B = 12. One would expect SHREC to take $(64/12)^3$ times longer for our bandwidth (see above), or on the order of 60 minutes. As a preliminary comparison, we implemented a modified version of SHREC.

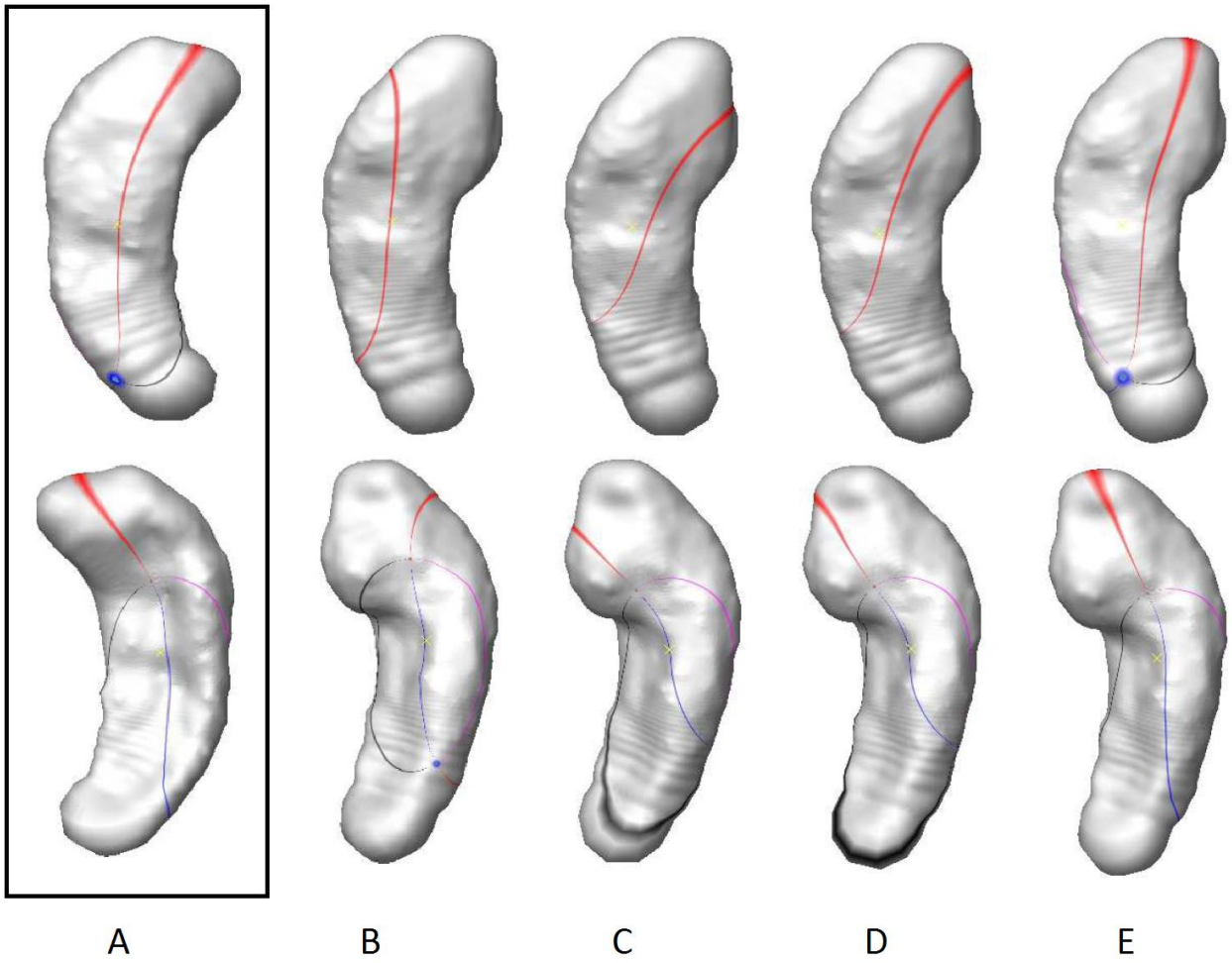

Fig. 1. (A), (B): Two randomly selected hippocampal surfaces, target shape (A) and shape to be registered (B). The red circle is homologous to the North Pole on the sphere; the blue circle, to the South Pole; and the blue line, to ϕ=0. The initial spherical mappings show significant misalignment. (C), (D): reparameterizations based FOE alignment (D) and SHREC (C). (E): reparameterization based on spherical cross correlation at $B = 64$.

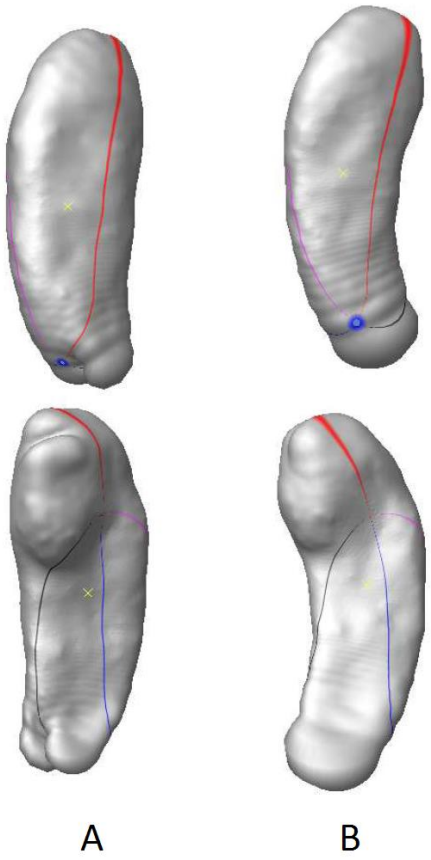

Fig. 2. Average of two shapes from **Figure 1** before (A) and after (B) correlation.

The only difference with the original is that at each iteration, our parameter alignment was initially done with cross correlation as described above, and subsequently refined according to the scheme outlined in [25]. Again, a bandwidth of 64 was used. We only present the results of one subject in **Figure 1C**. Here, we can see that the initial alignment determines the final result to a great degree. This example converged after only 4 iterations. Execution time was close to 7 minutes. Since the execution time reported in [25] was for a MATLAB implementation, while ours is in C++, and because we use a fast cross-correlation, this time is significantly lower than the 60 minute estimate.

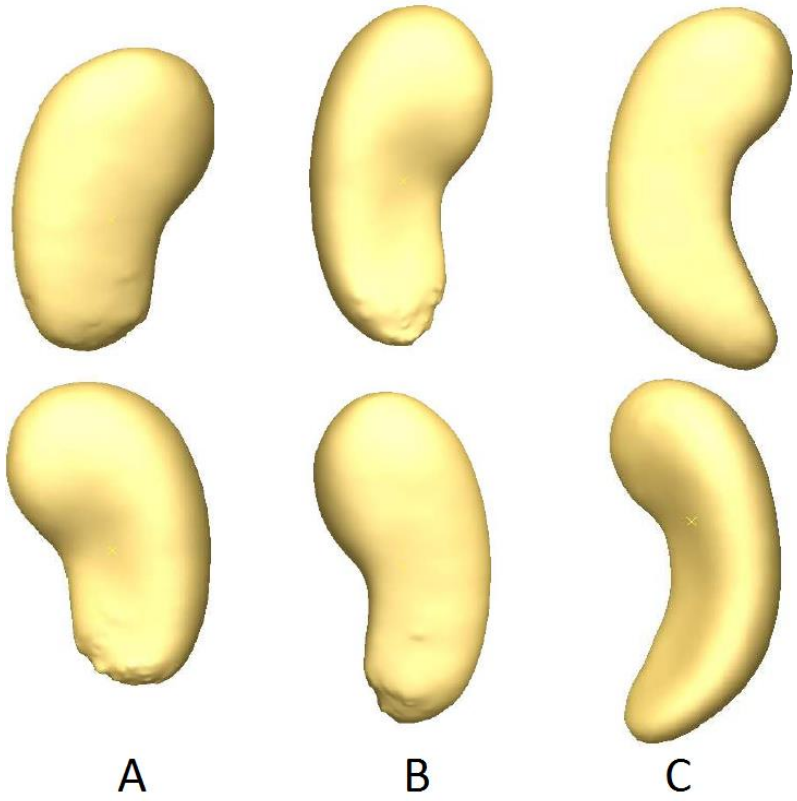

Fig. 3. Average of 45 right hippocampi with (A) FOE, (B) cross correlation, bandwidth = 16, (C) cross correlation, bandwidth = 64.

To test the effect of higher coefficients on correlation quality, we increased the bandwidth to 128, while keeping the same rotation tessellation and correlated 6 of the 45 subjects in our HP population. Running time increased to 404 $\pm$10.4 seconds, while the shape distance to the target hippocampus decreased on average only 3.9 $\pm$9.4 %.

To test for the effect of rotation sampling frequency, we also decreased N_R to 100 X 50 X 100 samples, while keeping the bandwidth at 64. Running time was 38.2 $\pm$3.5 seconds, an insignificant improvement in speed. In another experiment, we limited our bandwidth to 16 and reran cross-correlation based alignment while keeping angle resolution at 200 X 100 X 200. Execution time was reduced to about 3 seconds.

Table 1. Weighted point-wise distance of 45 (44) hippocampal surfaces by registration method.

Method	Distance to Target	Distance to Average
FOE	4.14 ±1.33	2.89 ±1.05
Cross Cor. $B = 64$	2.64 ±0.92	2.36 ±0.75
Cross Cor. $B = 16$	3.38 ±1.06	2.33 ±0.85

We constructed hippocampal averages using FOE and two versions of cross-correlation ($B = 16$ and $B = 64$) and compared results. First, a shape was selected, and all remaining shapes registered to it using each of the methods. Shapes were then averaged, normalized for volume (after registration, only for a fair distance comparison) and a rigid quaternion transform applied to each shape to align it both to the target subject and to the volume-normalized average shape. Our shape distance was defined as vertex-wise distance between the surfaces, weighted by product of the sum of areas of adjacent triangles in each mesh. The results are in **Table 1**. The table does not show a notable fact: distance to the target subject was improved by our method for every subject compared to FOE. Minimal improvement was 8%, and maximal 117%. All but one subject registered with $B = 64$ had superior alignment to the result of using $B = 16$. Compared to FOE, B = 16 reduced shape distance for all but 5 subjects. **Table 2** shows a summary of intra-subject differences by registration method. Using a bandwidth of 16 gives a significant improvement compared to FOE, but the results are still much improved by using a bandwidth of 64. **Figures 3A-3C** illustrate the hippocampal averages achieved with the three methods. Note that these averages were computed without spatially aligning the subjects to the target. Doing so would have likely given a more detailed shape.

We should note here that the poor result of the FOE average is due in part to the large area distortion of the conformal map. The tail of the hippocampus is mapped to such a small region on the sphere that it is very hard to match well. We see in **Figure 3A** that the tail suffers the most. The resulting spherical harmonic representation contains much redundant information; hence, the area distortion partially cancels out the benefits of using high order coefficients. The mapping used in [6, 25] is by contrast area-preserving, and hence better suited for alignment. The results of [6, 10] look closer to the one achieved here with cross correlation. This is, however, indicative of the potential our method has when applied to area-preserving spherical maps.

Table 2. Intra-subject improvement by registration method, in percentage of the second method's result.

Common shape	FOE vs. B = 64	FOE vs. B = 16	B = 16 vs. B = 64
target	61% ±26%	25% ±23%	30% ±18%
average	24% ±29%	26% ±22%	-6% ±29%

1.6 Conclusion

We have presented a framework for shape alignment which generalizes several existing methods. Our method is robust, fast and allows for use of greater detail in alignment than was possible before. The correspondence search is performed globally and no pre-alignment is required; thus, the result and computation time are independent of the shape's size and initial orientation. Reaching a global maximum is always guaranteed. Our method can be tailored to suit a particular application by selecting the appropriate shape features for a particular data type. We intend to experiment with various shape attributes, apply area-preserving spherical maps to our method and extend the technique to automated patch selection and matching.

CHAPTER 2

Global Shape Description: Application to Disease Classification

This chapter is adapted from:

B. Gutman, Y. Wang, J. Morra, A. Toga, P. Thompson, "Disease classification with hippocampal shape invariants," MICCAI 2008 Workshop on Computational Anatomy and Physiology of the Hippocampus (**CAPH '08**):76–86.

B. Gutman, Y. Wang, J. Morra, A. W. Toga, and P. M. Thompson, "Disease classification with hippocampal shape invariants," *Hippocampus,* vol. 19, pp. 572-8, Jun 2009.

Abstract

Weighted spherical harmonic shape descriptors are based on subspaces of $L^2(\mathbb{S}^2)$ spanned by spherical harmonics of a single degree. Such shape descriptors incorporate both shape and scaling information, while preserving invariance with respect to other non-reflexive affine transformations. Thus, their application allows for direct comparison of shapes across subjects, resolutions and within-subject components. On the other hand, spherical conformal parameterization typically leads to highly variable conformal (scaling) factors across the entire shape. While an issue for surface registration, this becomes an advantage when using invariant frequency-based descriptors: the descriptors attain a spatial interpretation, with higher-order terms corresponding to regions with a higher conformal factor. We validate our novel description in a classification study of Alzheimer's Disease vs. control subjects. With global descriptors forming our bag of features, Support Vector Machine classification of 49 Alzheimer(AD) and 63

elderly control subjects yielded 75.5% sensitivity and 87.3% specificity with 82.1% correct overall in a leave-one-out test. We show that our description contributes new information to simpler shape measures. Armed with a rigid shape registration tool, we also present a way to visualize variation in global shape description as a local displacement map, thus clarifying the descriptors' anatomical meaning.

2.1 Introduction

Shape of subcortical structures in the human brain is known to correlate with some neurodegenerative diseases. This fact has motivated several approaches to quantifying subcortical shape in anatomical studies. Since we think of an object's shape as separate from size, position and orientation, shape quantification must be made invariant to these factors. Generally, two approaches have been used.

In the first of these, shape measures are tied to particular locations on the hippocampus [9, 31]. These methods lend themselves to immediate visualization. It is easy to see which regions of the shape contain meaningful variation. The disadvantage of this approach is the need to normalize the data with respect to those factors that are not intrinsic to shape. In other words, one must first bring hippocampal models into register and align them in space. This normalization step introduces additional noise to the data. Also, in recent years a great multitude of ways to spatially register shapes has been developed and each one can lead to a slightly different result.

In the second approach, a global shape description is used instead. Examples of this include spherical moments, spherical harmonics, etc. [3, 32]. Although these measures are

entirely intrinsic to the object's shape and require no registration, it is usually not possible to reconstruct a surface from them or relate them to particular regions where the variation occurs. For this reason they have not been widely used in anatomical studies since their inception.

This work combines these two approaches. The first part presents global shape description as a viable alternative in detecting Alzheimer's disease (AD). The second shows a way to visualize meaningful variation in global description locally. We base our description on invariant properties of spherical harmonics (SPH). Such a description has already been proposed in the graphics community [3] and prior subcortical anatomy studies [33]. To validate our description's ability to discriminate between patients and controls, we use a linear Support Vector Machine (SVM) classifier with leave-one-out testing.

Finally, taking those features of our description which are most effective in separating patients from controls, we reconstruct each hippocampal model from a mixed spherical harmonic spectrum. In this mixture, only those components which are selected for SVM are kept from the original individual model, while the rest are taken from the population's average shape. By analyzing local distance between the average shape and each mixed reconstruction, we detect regions of the hippocampal shape most affected by the selected global descriptors.

2.2 Materials and Methods

2.2.1 Data

Our data set consists of 112 1.5T T1-weighted magnetic resonance imaging (MRI) scan images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, with 49 AD patients and 63 controls, age and gender-matched (mean age: 76.14, 76.76, P = 0.609). Initially, structural

MRI images are automatically converted into binary hippocampal masks with the help of the recent Auto Context Model [34]. ACM uses a few hand-traced hippocampi as a training set for AdaBoost to create a voxel-level classification function. We then convert the masks to a signed distance function and apply topology-constrained mean curvature flow following the topology-preserving geometric deformable model algorithm [35]. Following triangle mesh extraction and minor processing [36], a quick visual check is done on each mesh to ensure that the original masks correspond to a hippocampal shape.

2.2.2 Global Description

After extracting triangle mesh models of our hippocampi, we generate an invariant description of each shape. This generation consists of three steps: (1) spherical conformal parameterization following [28], (2) computing SPH coefficients of each surface with the help of a spherical FFT [29], (3) computing shape invariants from SPH coefficients.

To preserve rotational invariance of our description we require that the spherical parameterization map a rotated object onto a sphere such that the new spherical image is a rotated version of the original (these two rotations needn't be the same, although they usually are.) This is the motivation for choosing the global conformal map. Once spherically mapped, the surfaces are represented by the three inverse maps $f^x, f^y, f^z: \mathbb{S}^2 \to \mathbb{R}$. The SPH-based shape invariants are then defined as

$$s(l) = \sum_{i\in\{x,y,z\}} \sum_{|m|\le l} \left\| \widehat{f_{lm}^{i}} \right\|^2 . \qquad (2.1)$$

To see that this description is invariant under Euclidean motion, we start by revisiting equation (1.3) from the previous chapter. The immediate implication is that spherical functions which are

contained in a subspace of $L^2(\mathbb{S}^2)$ formed by harmonics of a single degree, $f \in \mathcal{S}_l = span\{Y_l^{-l},\ Y_l^{1-l}, \dots Y_l^m, \dots Y_l^{l-1}, Y_l^l\}$, remain in this subspace after a rotation of the function on the sphere. Since a rotation leaves the norm of a function unaltered, given if $h(\omega) = f(R^{-1}\omega)$, we have

$$\sum_{|m|\le l} \left\|\widehat{f_{lm}}\right\|^2 = \left\|Proj_{\mathcal{S}_l} f\right\|^2 = \left\|Proj_{\mathcal{S}_l} h\right\|^2 = \sum_{|m|\le l} \|h_{lm}\|^2 \tag{2.2}$$

Further, taking into account the linearity of $\Lambda(R)$, we can sketch a proof of $s(l)$'s invariance. Of course, we must also remember that a rotation of the shape in its native space will alter the image of the spherical scalar functions to be rotated parametrically. Letting $R_{i,j}, i,j \in \{x,y,z\}$ represent the elements of *R*, suppose $s'(l) = \sum_{i\in\{x,y,z\}} \sum_{|m|\le l} \left\|\widehat{h_{lm}^i}\right\|^2$. Then,

$$s'(l) = \sum_{j\in\{x,y,z\}} \sum_{|m|\le l} \left\| \langle \sum_{i\in\{x,y,z\}} R_{j,i} f^i(R^{-1}\omega)\, , Y_l^m \rangle \right\|^2 = \tag{2.3}$$

$$\sum_{j\in\{x,y,z\}} \sum_{|m|\le l} \left\| \sum_{i\in\{x,y,z\}} R_{j,i} \langle f^i(R^{-1}\omega), Y_l^m \rangle \right\|^2 =$$

$$\sum_{j\in\{x,y,z\}} \sum_{|m|\le l} \left\| \langle f^j(R^{-1}\omega), Y_l^m \rangle \right\|^2.$$

Note the use of the rotation-preserving property of our conformal map above. Now applying (2.2) to the last line above, we see that

$$s'(l) = \sum_{j\in\{x,y,z\}} \sum_{|m|\le l} \left\| \langle f^j(R^{-1}\omega), Y_l^m \rangle \right\|^2 = \sum_{j\in\{x,y,z\}} \sum_{|m|\le l} \left\| \widehat{f_{lm}^j} \right\|^2 = s(l). \tag{2.4}$$

By setting the zero-order coefficient to zero we achieve translational invariance. Essentially, the *l*-th shape invariant is the L^2 norm of the Euclidean distance from the surface to the average value of the spherical map, projected onto the *l*-th degree subspace. Note that the average

value will only coincide with the center of mass if the mapping is equiareal.

2.2.3 Support Vector Machines classification

Linear SVM [37] seeks an optimally separating hyperplane to distinguish two classes within a feature space. Given $\{\mathbf{x_i}, c_i\}_{i=1}^{N}$ data points and their class c_i, linear SVM minimizes $\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$ constrained by $c_i(\mathbf{w}\cdot\mathbf{x_i} - b) \geq 1 - \xi_i$, where ξ_i are the slack variables, measuring the degree of a data point's misclassification, and $\mathbf{w}$ are the weights defining the hyperplane. An unlabeled example is classified by the sign of the SVM score $\mathbf{w}\cdot\mathbf{x_i} - b$. In this study, we use Joachims' svmPerf package, described in [38]. Shape invariants form our feature space. We compute spherical harmonics, and consequently the shape description, up to a bandwidth $l < 256$. Since we have a left and a right hippocampus, we have a total of 510 features. Though far smaller than the initial sets of locally-based models, this is still too large to train a good model given our number of subjects. Feature selection is needed.

Feature selection is a problem encountered in many SVM classification studies, and a wide array of literature on the subject exists [39]. The most used selection method in AD studies seems to be optimal thresholding of SVM weights with cross-validation within the training set [40]. For now we have instead chosen simple t-statistic threshold as selection criteria. Thus, for each test subject the t-statistic is computed anew and the same threshold is used for each new test.

2.2.4 Invariant Descriptor Visualization

Although no spatial registration is required for our shape measures, it is needed for visualization. We require that the shapes are registered by a rigid rotation of their spherical maps, as that

ensures that the shape's global description does not change. We use the method of spherical cross-correlation developed in the previous chapter. The idea then is to create an average shape based on the newly found local correspondence and rigid spatial alignment. Once a point-wise correspondence between each hippocampal model and the average has been established, it is possible to reconstruct each individual shape with any desired mixture of averaged and individual SPH coefficients. Thus it is possible to visualize the local effect of each shape descriptor by keeping only those coefficients which contribute to a particular descriptor from the individual surface, while taking the rest from the average. Visualizing a point-wise distance between the average and each mixed reconstruction gives an idea of which regions of the surface contribute the most to each descriptor.

To make our visualization as objective as possible, the initial correspondence is established using a target subject that is not used in any subsequent analysis, or classification. Each hippocampal model is then spatially aligned to this subject using the rigid quaternion transform. A different shape average is computed for each test subject, leaving the subject's model out of the computation. The mixed-to-average distance of each individual shape is then itself averaged and displayed.

2.3 Experimental Results

Before using our shape description for classification it was important to ascertain whether its theoretical invariance holds in practice. For all its advantages, the conformal map has one significant shortcoming: its large area distortion. We illustrate this in **Figure 4**. Regions of extreme Gaussian curvature which protrude are mapped to very small regions on the sphere and

suffer from undersampling. This could potentially cause our shape description to lose its invariance.

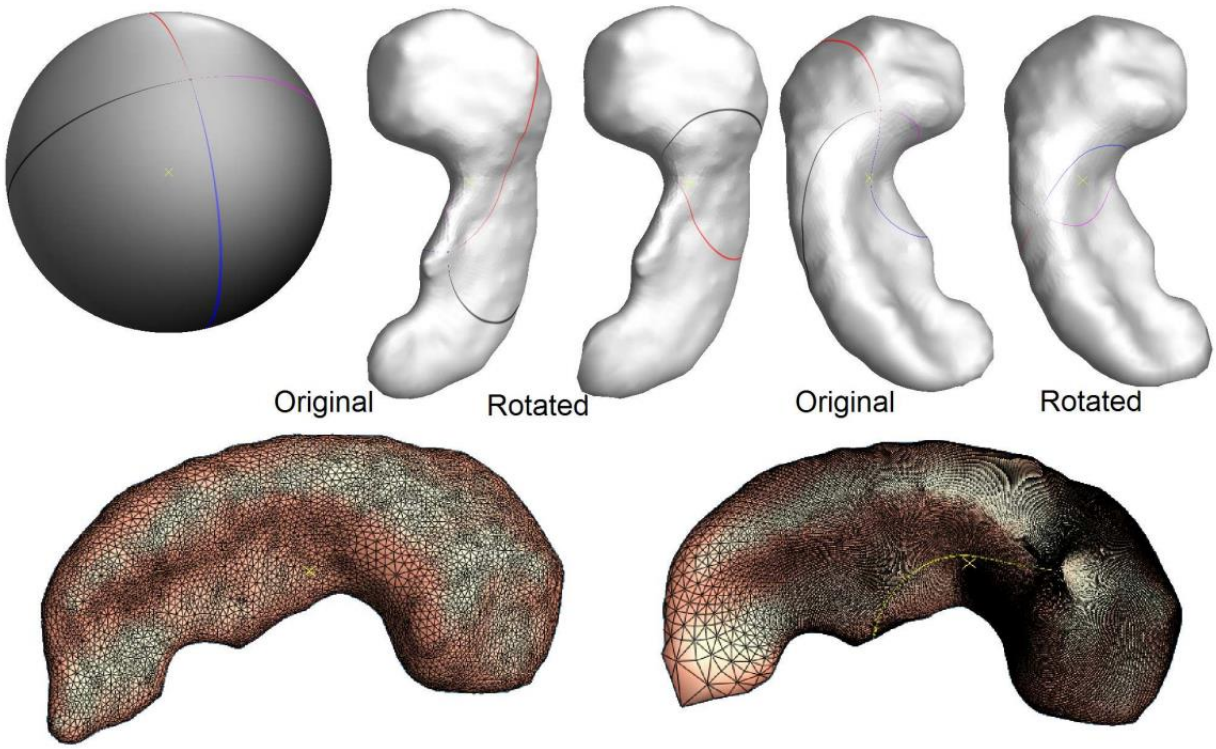

Fig. 4. Rotation and Undersampling: the bottom row shows the original 20K triangle hippocampal surface (left) and its sampled version (right). Note the undersampling. The top row shows the effect of a random rotation. The same surface was rotated by $\alpha = 151.8$ $\beta = 75.6$ $\gamma = 259.5$ reparameterized and resampled. Note the slightly different parameterization of the undersampled region.

Figure 4 and 5 illustrate the effect a random rotation of a surface has on its invariants. In the first hundred orders, error is within 2%. More importantly, the greatest error of the invariants selected for SVM classification (see below) is within 0.5%.

For each training set in the leave-one-out test, we selected a feature if its t-statistic exceeded a threshold. After testing a few subjects, we noticed that the best overall accuracy is achieved with $6.7 \leq t_{min} \leq 6.9$, and set it globally to 6.8. This yielded between 6 and 14 features, depending on which subject was left out. All selected features *s(l)* were of order

$37 \leq l \leq 58$. Our margin/error coefficient C was set to 1,000. All features were normalized with respect to standard deviation (differently for each excluded subject) and translated so that $min(x) = -max(x)$. The transformation was saved and applied to the remaining subject. The result was 75.5% sensitivity and 87.3% specificity for a total correct rate of 82.1% (AD considered positive).

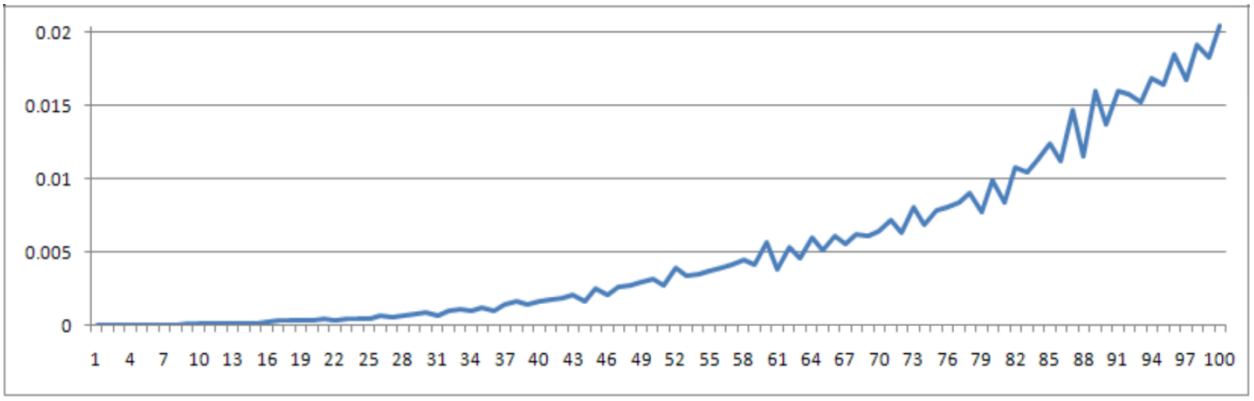

Fig. 5. Effect of undersampling on invariance. We show the relative error of invariants corresponding to the surface in **Fig. 4**: $|s(l)-s'(l)|/s(l)$ vs. l, where l is the degree of the invariant.

By comparison, hippocampal volume gave 67.3% sensitivity and 76.2% specificity in a leave-one-out test, with 72.3% correct overall. To combine our best features into one measure, we ran SVM on the entire data set with the same t_{min} and C and obtained each subject's SVM score. In regression, SVM score correlated slightly better with Mini-Mental State Examination (MMSE) and Clinical Dementia Rating Sum of Boxes (CDR) scores than volume: for MMSE, $R^2_{vol} = 0.253$, $R^2_{SVM} = 0.291$; for CDR, $R^2_{vol} = 0.276$, $R^2_{SVM} = 0.295$, $P < 0.001$ for all. Since all our selected discriminating features came from the right hippocampus, consistent with a locally-based study on this data [41], we ran the same tests using only right HP volume. We found it is a worse predictor than combined volume in all cases. To measure how much new information is contained in our shape description compared with volume, we ran a linear

regression on combined volume and SVM score. Results are shown in **Figure 6a**. **Figure 6b** shows the ROC curve of our SVM model, total volume and the combined model. SVM is superior to volume for the entire range of sensitivities.

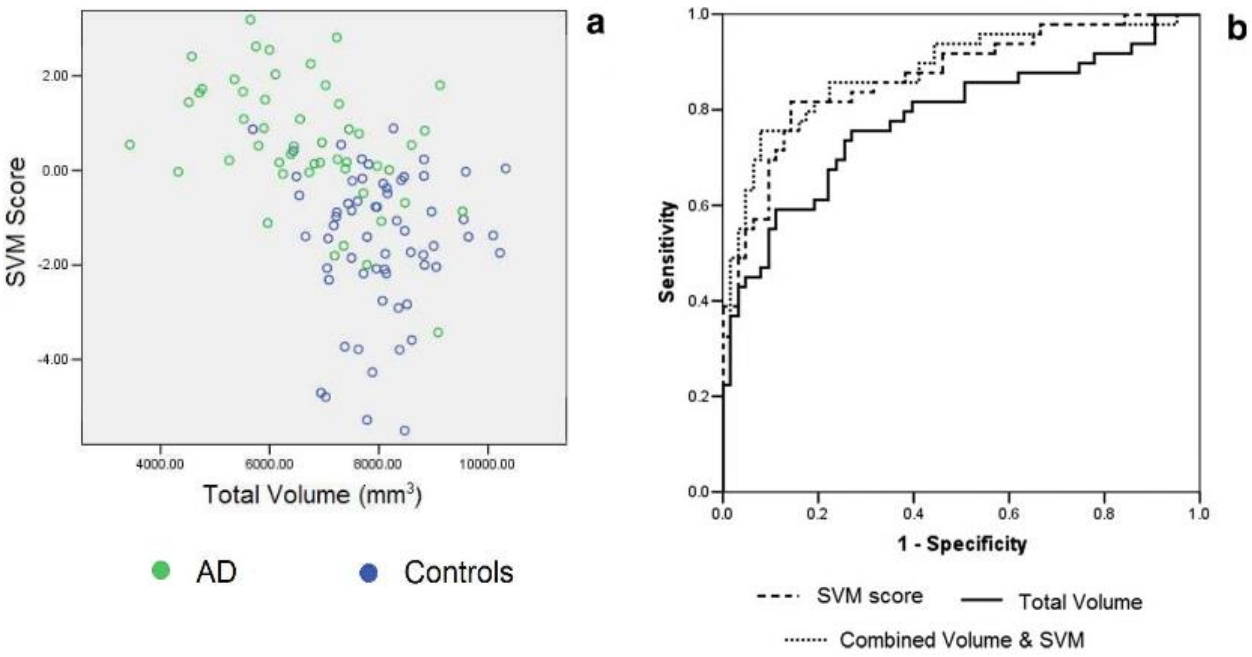

Fig 6. (a) Scatter of SVM score and total hippocampal volume. Regression R^2 =0.16. (b) ROC curves of SVM score, total volume, and the model combining volume and SVM.

The six descriptors that were selected in every case were of degree $l = 39, 41, 46, 48,$ and 58. Our mixed reconstructions were based on coefficients of these orders **(Fig. 7)**. The average mixed-to-average distance is displayed on the overall average surface in **Figure 8a**. Thresholding surface regions based on a peak histogram displacement value gives an idea of which surface parts contribute the most to these harmonics (**Fig. 8b**).

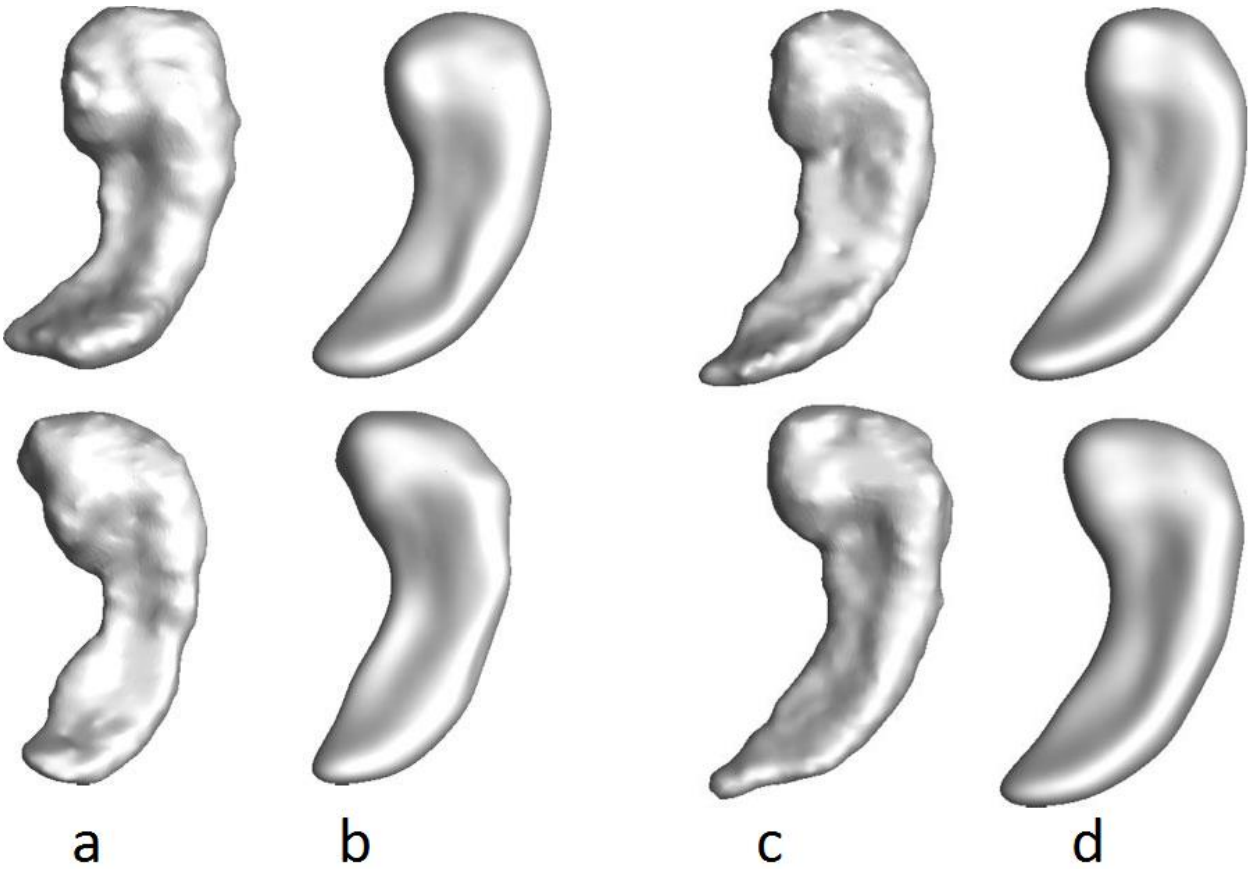

Fig. 7 Four examples of right hippocampal reconstructions (a,b) – controls, (c,d) - patients. The original shapes (a,c) are on the left and reconstructions (b,d) on the right.

2.4 Comparison to prior work

2.4.1 Methodology and results

While achieving modest classification results compared to some recent AD studies, our global description appears to contain shape information that is not captured by simpler measures like volume. This is apparent in the low correlation between our description's SVM score and hippocampal volume and in the ROC curve. In most AD classification studies to date, volume-based features, such as gray matter probability maps, are used with SVM to predict diagnosis [40, 42]. In the best ones, overall leave-one-out accuracy was 89–96% [43]. More relevant to the

current study, some recent works instead classified disease according to hippocampal shape-based features [44, 45] , with point-wise displacements forming the feature set. Li et al., used patch-averages of local displacement vectors projected onto the average normal. SVM was then applied to these local features to separate AD subjects from controls. Using hand-traced surfaces, the best leave-one-out accuracy reached 94.9%.

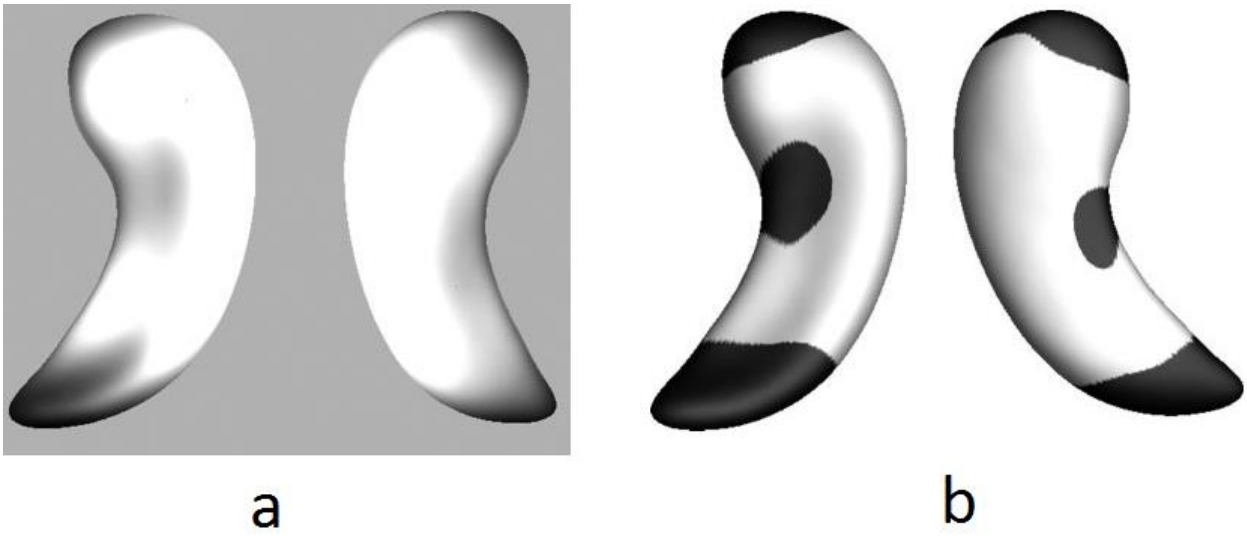

Fig 8. (a) Average displacement map. Displacements are computed from each mixed reconstruction and the average surface. Darker areas correspond to greater displacement; hence, these areas contribute more to the discriminant harmonics.

(b) Dark spots correspond to regions above a peak histogram value from (a). These regions contribute the most to the discriminant harmonics.

Davies et al. studied effects of Schizophrenia using the minimal distance length approach to statistically align hippocampal parameterizations in [46]. For classification, Linear Discriminant Analysis (LDA) is used to find the discriminate vector in the feature space for distinguishing diseased subjects from controls. The work is compared to the SPHARM technique [6] with both approaches yielding a Student's t-statistic for the group difference of less than 2.3 along the

discriminate vector. Classification rates are not reported, but the authors claim that an SVM classifier on this feature space yielded practically the same results.

An interesting study by Gorczowski [47] et al. recently appeared on classification using multi-object complexes. Their approach is advantageous to ours in that it takes into account the relative position of several subcortical structures with respect to each other, while we can only combine several shape invariants from every structure individually. This study, however, acknowledged that the classification results are improved when pose is eliminated from the feature space and only structure-intrinsic features (here, radii of m-reps developed by Gerig, Styner and coworkers) are used. Though the validation method is more robust than ours, the accuracy is inferior: 75%.

A unifying aspect of the studies above is their emphasis on locally-based features: in each case a feature corresponds to either a voxel or a point on the surface. While this facilitates visualization, it may not take full advantage of some pattern a shape exhibits globally. Shenton et al. [48] perhaps comes closest to our approach in that he uses two nonlocal shape features to classify Schizophrenia subjects and controls. This study does indeed use a spherical harmonic representation, specifically SPHARM, to align the left and flipped right amygdala-hippocampal surfaces for each subject. However, once the shapes are aligned, the study again returns to a simple spatial measure — not spherical harmonics or any features derived from them — to classify the shapes. Two asymmetry measures, volume difference and mean square distance with the volume normalized, are used in an SVM classification. A good accuracy of 87% is achieved.

2.4.2 Limitations

Our study was slightly biased since the same small subgroup we used in setting the selection threshold was used in classification testing. Although our shape measures separated patients from healthy subjects substantially better than hippocampal volume, other AD classification studies have had better results with volume alone, including one that used the ADNI dataset [49]. Since using volume for classification is straight-forward, it is hard to point to our classification technique as the cause of such discrepancy. More likely, this is because our hippocampal segmentation produces lower-quality models than the hand-traced models used in other studies. Better segmentation techniques will likely lead to better classification results by this method.

Since the number of features in these studies is much larger than in the present article, a very robust feature selection is required as a preprocessing step before SVM can give reasonable results. The usual and evidently quite reliable means of doing this is the recursive feature elimination (RFE), as in [40, 42, 45]. Here, an SVM model is iteratively trained and at each step the weakest (least-weighted) feature is removed. This is repeated until the classification rate or the particular cost function of the SVM model stop improving. Though this method is SVM-centric and well-suited for the problem, it is more expensive than our simple feature selection technique. Our naive selection may in part explain why our accuracy is inferior to some of the best results in the above studies.

2.4.3 Visualization

Perhaps the most interesting part of this work is the visualization of a meaningful variation in global description as local variation. The ability to represent an individual hippocampus as a

mixture of average and individual shape effectively allows for a representation that minimizes meaningless shape variation while maximizing variation that has some significance for disease detection. Using such a representation could potentially have clinical uses; for example, it could make visualization-based diagnosis by non-experts possible.

When comparing our visualization results with those based on local measures [8, 50], it is important to keep in mind that they are unlikely to be identical. For example, Wang et al. [50] use a pairwise displacement map as a measure of local atrophy after a large diffeomorphism-based registration and rigid alignment. In both local and global approaches we can expect some degree of noise to contribute to the true effect. Since our noise occurs in the spectral domain, unlike the studies above, the effect of our noise can be quite different. It is also important to keep in mind that we are measuring different things. First, in [50] change over time is compared between the two groups, whereas we compare shape difference at a single point in time. Second, we measure the effect of certain spectral components on the overall shape, making our visualization a much less direct way to see disease effects. Locally-based studies measure spatial differences more directly.

2.5 Conclusion and future work

It would be interesting to do a cross-validation study to see whether mixed reconstruction increases discriminative power of locally-based features. For example, one could break a dataset in two parts, use one to find discriminate aspects of a global description and create mixed reconstructions of subjects in the other. Ideally, an SVM-based feature selection would be used. For an objective comparison, the average shape used for reconstructing the second data set would be taken entirely from the first set.

Alternatively, we can expand our bag of features by incorporating scalar maps that are intrinsic to the surface. For example, incorporating mean curvature and conformal factor would create a brand new description and at the same time allow for mixed reconstructions, since it is possible to uniquely reconstruct a conformally mapped surface based only on those two features. Again, so far the only scalar map we have used is distance to surface average value.

We have presented an alternative means of disease classification based on a global description of hippocampal shapes. In experiments, our method's accuracy, though respectable, remained inferior to some of the best reported AD classification results. However, the novelty of the information contained in our measures means that our feature set may well be useful in complementing existing classification methods. In contrast, our visualization of a global description appears to be the first of its kind. Although admittedly simple to do, it gives a new interpretation of a global description. Our mixed reconstructions allow us to deliberately reduce meaningless shape variation while preserving variation that is of interest to the researcher.

CHAPTER 3

Shape Matching With Medial Curves and 1-D Group-Wise Registration

This chapter is adapted from:

B. A. Gutman, W. Yalin, P. Rajagopalan, A. W. Toga, and P. M. Thompson, "Shape matching with medial curves and 1-D group-wise registration," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, 2012, pp. 716-719.

Abstract

We present a method for shape matching that approximates group-wise shape registration by reducing the problem to a 1-D registration. First, a novel medial curve method is proposed for computing a 1-D description of the shape. Second, a group-wise registration of the geometric descriptor is performed by directly minimizing the group variance. The resulting registration is used to adjust a global shape feature used to compute the final correspondence. Thus, the problems of description, registration, and statistical analysis are solved in one framework, while reducing the computational problem of group-wise registration of shapes substantially. We validate our method on 620 lateral ventricles extracted from the ADNI MRI dataset and 19 lateral ventricles from patients with HIV/AIDS and matched controls, also scanned with MRI. We show that our group-wise approach leads to improved statistical results, and also compare it to the SPHARM method.

3.1 Introduction

In computational anatomy, it is often desirable to locally compare the geometry of anatomical shapes. Essentially, anatomical shape analysis poses three related problems: (1) defining a meaningful and intuitive geometric descriptor, (2) registering the shapes in an unbiased manner, and (3) statistically comparing the resulting description. These three problems are interconnected, but existing methods generally solve these sequentially, or combine at best two of the three steps.

Several approaches have been proposed towards (1), including medial core-type methods [31, 51] which define a local thickness of a shape, and an extension of Tensor Based Morphometry (TBM) to 2-manifolds [52]. The latter may require some imagination to be made intuitive, but it can be more statistically powerful for detecting correlations between subcortical shape and clinical, cognitive and CSF biomarkers.

In this work we choose a thickness-based approach based on a 1-dimensional single curve skeleton, as many anatomical shapes are oriented and therefore admit such a description. Thus, our geometric descriptor is the medial thickness, which is generally accepted as the most intuitive measure of shape morphometry. 1D medial representations in medical imaging were popularized by Stephen Pizer with the M-reps algorithm [51]. M-reps are comprised of a discrete web of "atoms," each of which describes position, width and local directions to the boundary, and an object angle between corresponding boundary points. However, the requirement that the direction between each atom's position and each corresponding surface point must be normal to the surface makes this method quite constrained even for anatomical shapes. To endow a shape with an M-reps description, the shape must either be quite simple, or it must be excessively

smoothed. For example, such a description is not feasible for branching shapes such as the lateral ventricles, with all three horns prominently represented.

CM-reps, popularized by Paul Yushkevich and colleagues, is an extension of M-reps to the 2-D continuous medial core. CM-reps offer a way to derive boundaries from skeletons, by solving a Poisson-type partial differential equation with a nonlinear boundary condition [31]. The medial axis-anchored 3D parameterization of the shape-enclosed volumetric region is continuous, and allows a body-centered coordinate system for analyzing shape and appearance. However, because CM-reps do not reduce dimensionality, they do not produce significant computational speed-ups.

A number of methods have also been developed towards the registration step (2), such as Minimum Description Length (MDL) [23], intrinsic parameterization methods such as SPHARM [6] and conformal maps [52], direct shape mapping based on Laplace-Beltrami eignefunctions (LBE) [53], Spherical Demons (SD) [13], and q-maps [54]. MDL is an information-theoretic approach for group-wise shape registration. It leads to an unbiased correspondence within a sample, but it can be quite slow as it relies on simulated annealing, which may take many hours for just a handful of shapes. This method is not feasible for a large cohort such as ADNI where N is several hundred. SPHARM, conformal maps, and more recently q-maps compute optimal shape parameterizations, either by minimizing angle or area distortion, or by optimizing over a metric space of reparameterizations. These methods are quite general, but they do not exploit the shape descriptor, and thus do not couple steps (1) and (2). LBE-based direct maps generate feature functions to be used for direct registration using level set embeddings. Similarly, SD performs shape registration on the sphere by registering spherical images in the Diffeomorphic

Demons framework. Thus, both SD and LBE are capable of combining description with registration. However, these methods are designed to register an image to either another image or, in the case of SD, to an atlas formed by a set of pre-registered images. This biases the resulting parameterization, as the statistical analysis step (3) remains uncoupled with the rest of the processing. Spherical Demons perhaps comes closest to coupling all 3 steps above. Further, LBE is only feasible for fairly simple shapes with a clear spatial orientation and no branching, a constraint similar to M-reps' limitations.

To mitigate this problem, we maintain the reduced 1-D representation with a single curve, but relax the M-reps' normal-to-surface constraint. Instead, our medial representation is based on a variational framework in which the curve is defined by minimizing a weighted total distance. Thus, the geometry of the curve and the shape are tied more loosely, and more classes of shapes admit such a representation. Although a single curve representation may appear inferior to 2-D representations due to the ill-posed nature of the 1-D problem, in fact the reduced dimensionality is of great benefit for subsequent processing.

Our registration step combines many of the registration methods above. We use the spherical domain and a combination of area and angle- preserving regularization, while minimizing the L^2 distance between pairs of feature functions. To enable fast spherical registration, we modify the unconstrained spherical parameterization (USP) algorithm [55] by simply adding the L^2 fidelity terms to the cost function. The resulting registration remains nearly as fast as the original USP tool.

The feature functions are induced by our medial curve, in the spirit of LBE. In our case, however, the curve is computed first, while in LBE the curve is induced by the function. Finally,

before registering the spherical images to a target, we perform a 1-D to 1-D group-wise registration in the spirit of MDL. Our 1-D registration non-linearly remaps the scalar feature to be used in spherical registration by minimizing the variance of the feature to be used as a geometric descriptor, i.e., the feature that will be compared statistically. Thus, this step combines all 3 shape comparison problems.

Our contribution is threefold: first, we develop a relaxed medial curve framework which allows noisy and branching shapes to be parameterized consistently by a single curve. This allows the computation of intuitive shape description by medial thickness and natural feature functions for registration. Second, as the curve has lower dimensionality, we enable a quasi-group-wise shape registration by group-wise registration of 1-D functions, thereby incorporating variance reduction and improved statistical sensitivity into the registration. And third, using a modification of the USP algorithm, we show that the spherical registration resulting from the 1-D registration step leads to improved statistical results based on ADNI and a dataset of lateral ventricles from HIV+ subjects and age-matched controls.

3.2 Medial Curve Framework

Finding the curve-skeleton of an orientable surface is not a well-defined problem, but some properties are generally accepted as desirable [9]:

(1) Centered: we would like our curve to be “locally” in the middle of the shape. In medical imaging, numerical accuracy is vital when estimating local thickness on boundaries of shapes.

(2) Onto, smooth mapping: there must be a surjective, smooth mapping from the surface to the curve. This enables us to use the medial curve for registration.

(3) Consistent geometry: this property requires that the geometry of the curve depend continuously on the shape

We assume that our anatomical shape can be represented by a single curve. We also assume that the ends of the curve lie on the surface. With the topology of the curve fixed, we focus on property (1) above. Intuitively, we can say that a curve is the *medial* curve if it is smooth and every point on it is "locally in the middle" of the shape. Given a surface *S*, the curve $\boldsymbol{c}(t),\ t \in [0,1]$, should be a global minimum of $R(\boldsymbol{c}, \boldsymbol{c}', S) =$

$$\int_0^1 \int_{\boldsymbol{p} \in S} w(\boldsymbol{c}(t), \boldsymbol{c}'(\boldsymbol{t}), \boldsymbol{p}, S) \|\boldsymbol{c}(t) - \boldsymbol{p}\|^2 \, dS dt \tag{3.1}$$

$$\boldsymbol{c}(0) \in S, \boldsymbol{c}(1) \in S$$

Here, $w(\boldsymbol{c}, \boldsymbol{c}', \boldsymbol{p}, S)$ is the weight defining "localness" of point $\boldsymbol{p}$ relative to $\boldsymbol{c(t)}$. A variety of weighting functions can be devised, one of the simplest ones being $w(\boldsymbol{c}, \boldsymbol{c}', \boldsymbol{p}, S) =$

$$\begin{cases} w_1(\boldsymbol{c}(t), \boldsymbol{c}'(\boldsymbol{t}), \boldsymbol{p}, S), & if\ min\{D[\boldsymbol{g}(u, \boldsymbol{p}, \boldsymbol{c}(t)), S] | u \in (0,1)\} > 0 \\ 0, & otherwise \end{cases}$$

$$w_1(\boldsymbol{c}, \boldsymbol{c}', \boldsymbol{p}, S) = \delta_\alpha \left(\frac{\langle \boldsymbol{c} - \boldsymbol{p}, \boldsymbol{c}' \rangle}{\|\boldsymbol{c} - \boldsymbol{p}\| \|\boldsymbol{c}'\|} \right) \tag{3.2}$$

where $D(\boldsymbol{x}, S)$ is the signed distance of S at $\boldsymbol{x}$, and $\boldsymbol{g}(u, \boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a} + u(\boldsymbol{b} - \boldsymbol{a})$. Here, δ_α is a continuous approximation of the Dirac delta. We used $\delta_\alpha = \alpha e^{-x^2\alpha^2}$ in our implementation. This

particular weighting function is only concerned with the differential properties of the curve, not those of the surface, making it quite stable when applied to noisy surfaces. The weighting function decreases continuously as the vector from the surface to the curve becomes less normal to the curve, and vanishes when there is no line of sight to the surface.

To ensure that the curve is smooth, we add a smoothing term, so that our final cost function takes the form:

$$L(\boldsymbol{c}(t), S) = R(\boldsymbol{c}, \boldsymbol{c}', S) + \beta \int_0^1 \|\boldsymbol{c}''(t)\|^2 \, dt \tag{3.3}$$

Solving for the Euler-Lagrange equation of (3.3), we derive the descent direction for the curve:

$$\frac{d\boldsymbol{c}}{d\tau} = \int_{\boldsymbol{p}\in S} \left[\left(\frac{d}{dt}\frac{dw}{d\boldsymbol{c}'} - \frac{dw}{d\boldsymbol{c}} \right) \|\boldsymbol{c} - \boldsymbol{p}\|^2 - w(\boldsymbol{c} - \boldsymbol{p}) \right] dS - \beta \boldsymbol{c}^{(4)}. \tag{3.4}$$

3.3 Group-wise 1-D registration

To enable curve-based registration for shapes described by a single curve, we derive two surface based functions. The first is the global orientation function (GOF), defined as

$$G(\boldsymbol{p}) = \arg\min_t \{\|\boldsymbol{c}(t) - \boldsymbol{p}\|, t\epsilon[0,1]\}. \tag{3.5}$$

This is similar to the first LB eigenfunction [6], except it can have more than two local extrema. Our geometric descriptor is then the familiar thickness measure:

$$D(\boldsymbol{p}) = \left\|\boldsymbol{c}\big(G(\boldsymbol{p})\big) - \boldsymbol{p}\right\|. \tag{3.6}$$

As we are interested in improving statistical sensitivity of medial thickness, we seek a set of mappings $r_n: [0,1] \rightarrow \mathbb{R}$, such that the total sample variance of a set of functions $f_n: [0,1] \rightarrow \mathbb{R}$ is reduced by minimizing the cost function $C(\{f_n, r_n\}_{n=1}^N) =$

$$\int_0^1 \frac{1}{N}\sum_{n=1}^N f_n^2[t - r_n(t)] - \frac{1}{N^2}\left(\sum_{n=1}^N f_n[t - r_n(t)]\right)^2 + \frac{\sigma^2}{N}\sum_{n=1}^N [r_n'(t)]^2 \, dt \qquad (3.7)$$

$$[t - r_n(t)] \in [0,1] \;\; \forall n$$

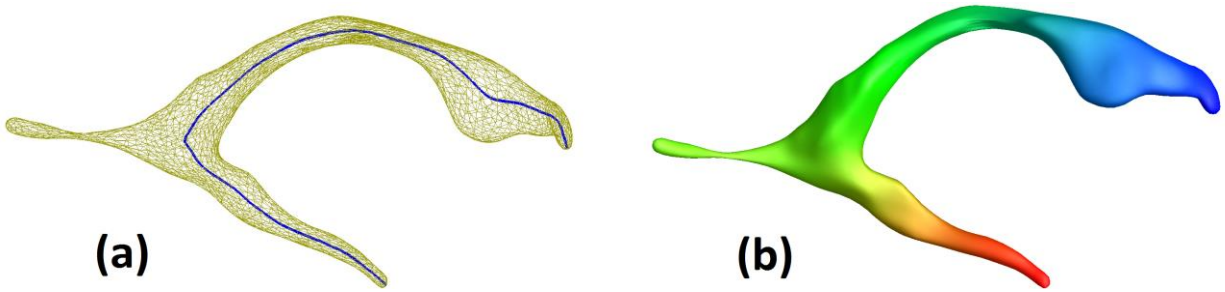

Fig 9. (a) Medial curve of a left lateral ventricular surface, based on the proposed framework; (b) the resulting global orientation function.

The last term simply represents elastic regularization. To enable minimization in thickness variance, we set $f_n(t) = \frac{\int_{\{\boldsymbol{p} \in S \,|\, G_n(\boldsymbol{p}) = t\}} D_n(\boldsymbol{p}) dS}{\int_{\{\boldsymbol{p} \in S \,|\, G_n(\boldsymbol{p}) = t\}} dS}$. This may be thought of as a curve-based average thickness. Having obtained $\{f_n\}$, we minimize (3.7) according to the descent direction $\frac{dr_n}{d\tau} = \frac{2}{N}\left[f_n'\left(f_n - \bar{f}\right) - \sigma^2 r_n''\right]$, where $\bar{f}(t) = \frac{1}{N}\sum_{n=1}^N f_n[t - r_n(t)]$. We then adjust the GOF based on the 1-D maps by $G_{adj_n} = h_n^{-1} \circ G_n$, $h_n(t) = t - r_n(t)$.

3.4 Non-linear Spherical Registration

We modify an existing, highly robust framework for unconstrained spherical mapping proposed by Freidel et al. [55], by adding L^2 terms to the cost. Using Freidel's notation, the functional

becomes $E = aE_{conformal} + bE_{area} + \sum_{i=1}^{K} \omega_i \int_{\mathbb{S}^2} \left(f_i(\boldsymbol{p}) - g_i(\boldsymbol{p})\right)^2 dS$, where f_i, g_i are scalar features of the moving and stationary spherical image. We use G_{adj} and D, the adjusted GOF and medial thickness features. Details on the optimization of this cost can be found in [8]. This is not a particularly innovative step, but a final surface-to-surface registration based on scalar features is needed to complete our shape matching. Our choice of registration framework is driven by practical concerns – alternative frameworks exist. In practice, this has proven to be a good choice. We use thickness as a feature in spherical registration, but thickness alone would not be sufficient to map the surfaces correctly, as it is a local measure. The global measure (GOF) primarily drives the registration.

3.5 Implementation and experiments

We have applied our method to two datasets of left lateral ventricles: the ADNI baseline MRI dataset consisting of 391 subjects with mild cognitive impairment (MCI), and 229 age-matched controls; and a dataset of 11 HIV subjects and 8 age-matched controls [56]. Parameters were set according to **Table 3** for both datasets. Our medial curve is evolved in a multi-resolution fashion with respect to both the curve and the surface: first, the surface is decimated to a small number of triangles (e.g., 300), and the curve is evolved at several resolution levels. We super-sample the curve by a pre-set ratio, as convergence is reached at the current level. The resulting curve is then used as initialization for the full mesh, at which point only a few descent steps are typically required. We show decrease in energy **(3)** with each step in **Figure 11** for an HIV+ subject. Typical execution time is 2 minutes for a 10K triangle mesh on an AMD Opteron 152 2.61 GHz

single core workstation with 4 Gb of RAM. An example of resulting curves and their GOFs can be seen in **Figure 9**.

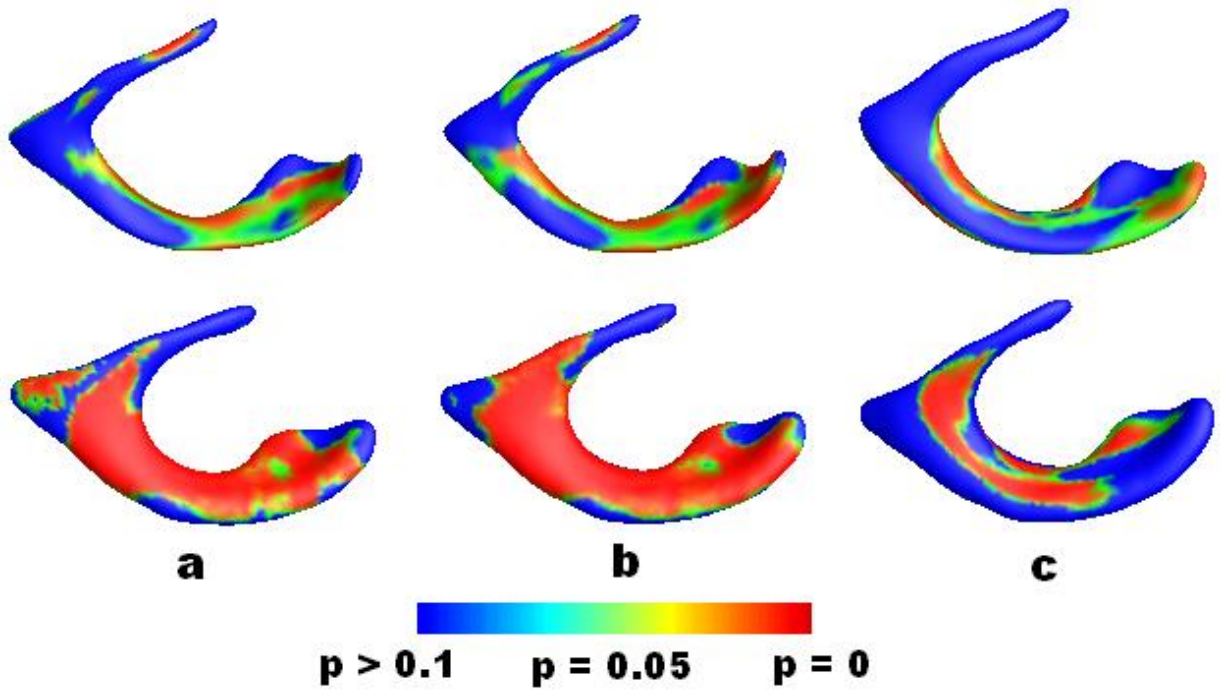

Figure 10. P-maps of HIV-NC (top row) and MCI-NC (bottom row) group difference after registering with (a) group-wise method, (b) unadjusted GOF + modified USP, (c) SPHARM

Group-wise registration reduced the variance of curve-based thickness by nearly 50% in the HIV cohort, and about 30% in ADNI. This step is computed in less than 1 minute. Our 1-D registration was done blindly with respect to both diagnosis and which subject was to be used as a target in spherical registration, to avoid "double-dipping". We show 3 average thickness functions from the ADNI dataset before and after the group-wise step in **Figure 12**.

Table 3. Parameters for group-wise shape registration

α	β	σ^2	ω_{thick}	ω_{GOF}	a	b
10	10^{-6}	0.1	10^2	10^6	5	2

Our final spherical registration step was initialized by an unconstrained spherical map [8], followed by a spherical cross-correlation to remove the rotational component [3]. We then registered each subject to an arbitrary target shape from the same cohort using our modified USP framework. The L^2 energy was reduced by 20-80% depending on the subject and the cohort. Typical execution time was 10-20 seconds.

The statistical analysis consisted of computing overall p-value for group differences based on 100000 permutations, as in [4]. For comparison, we also applied the SPHARM method to the same two datasets. We then computed Cumulative Distribution Functions (CDF) of significance maps for each method (**Fig. 13)**. CDF curves are a way of visualizing the multiple comparisons problem, so that one sees the tradeoff between the statistical threshold and the spatial extent of the effect. While the permutation test remains the gold standard, CDF is still a useful visualization. As expected, group-wise registration improved the overall p-value in both studies, as seen in **Table 2**. However, CDF plots show that using the unadjusted GOF, and keeping all other parameters the same produces the dominant curve. This is expected: as group-wise registration actually localizes the effect, the resulting p-maps have smaller significant regions, with greater effect sizes. By focusing the significant regions, the group-wise method makes the

p-maps more useful, as we see the true locations of significant change, rather than a washed out effect due to mis-registration. This effect is seen in **Figure 10.**

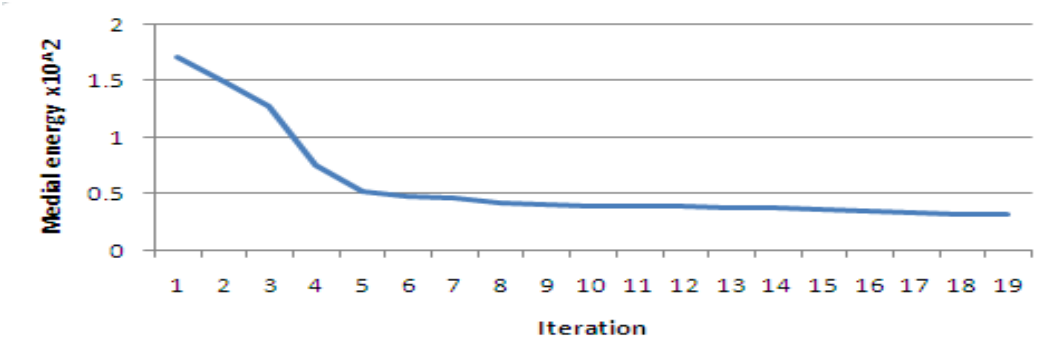

Fig. 11. Medial energy **(3.1)** vs. gradient descent iterations

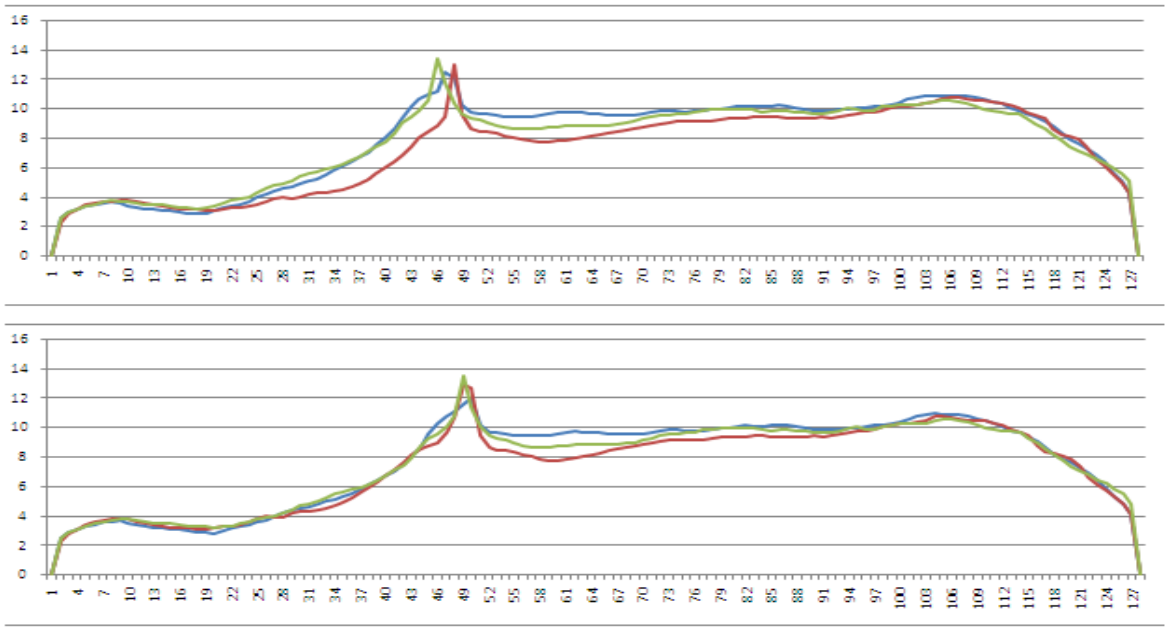

Fig. 12. 1D Thickness maps of 3 ADNI subjects before (top) and after (bottom) group-wise registration of all 620 subjects. Note that the peaks corresponding to the posterior horn are well-matched.

Table 4. P-values for group difference after 100K perms

	Group	No Group	SPHARM
HIV	**0.00988**	0.01039	0.0149
ADNI	**0.00029**	0.00046	0.0068

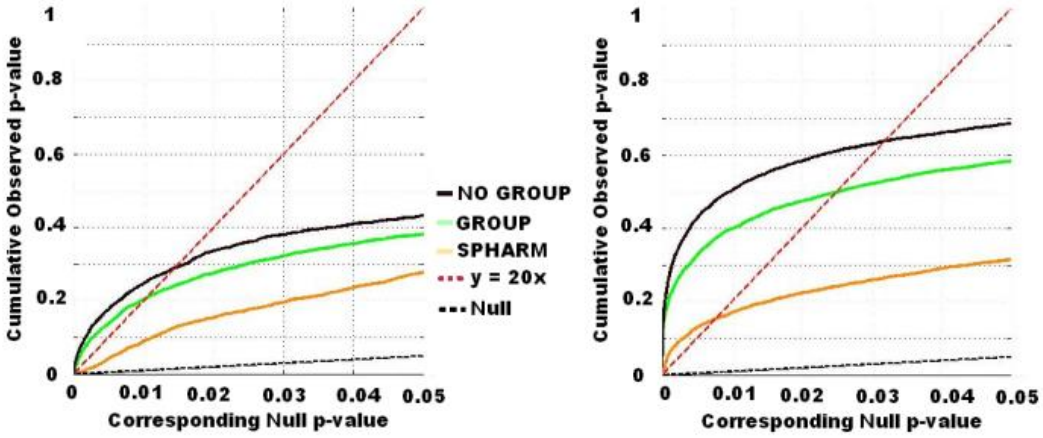

Fig. 13 CDF plots of HIV vs. NC (left) and MCI vs. NC (right)

3.6 Conclusion

We presented a framework to register anatomical shapes that combines the description, registration, and statistical analysis aspects of shape comparison in one step. Our approach does not lead to true group-wise shape registration, but it approximates the solution to the group-wise problem by exploiting the specifics of the geometry of anatomical shapes - they are often approximately tubular. We further developed a medial curve method for computing intuitive

descriptions of shapes so that the tubular assumption can be relaxed, and show that it works stably on the complicated lateral ventricles. As many anatomical shapes are inherently 1D, we reduce the computational problem significantly for group-wise registration. Thus, we are able to approximately group-wise register over 600 shapes in a few minutes.

We chose the most intuitive geometric measure, but other measures such as medial eccentricity, TV norm of the thickness, curvature-based features, functional and DTI-based measures, and others can all be incorporated easily into the framework. Future developments will include exploration of additional descriptors, as well as incorporating distance field information into the medial curve framework, and using a probabilistic atlas in the spherical domain.

CHAPTER 4

A Family of Fast Spherical Registration Algorithms for Cortical Shapes

This chapter is adapted from:

B. A. Gutman, S. K. Madsen, A. W. Toga, and P. M. Thompson, "A Family of Fast Spherical Registration Algorithms for Cortical Shapes," in *Multimodal Brain Image Analysis 2013*., accepted for **oral presentation**, Nagoya, Japan Sept. 22, 2013

Abstract

We introduce a family of fast spherical registration algorithms: a spherical fluid model and several modifications of the spherical demons algorithm introduced in [13]. Our algorithms are based on fast convolution of tangential spherical vector fields in the spectral domain. Using the vector harmonic representation of spherical fields, we derive a more principled approach for kernel smoothing via Mercer's theorem and the diffusion equation. This is a non-trivial extension of scalar spherical convolution, as the vector harmonics do not generalize directly from scalar harmonics on the sphere, as in the Euclidean case. The fluid algorithm is optimized in the Eulerian frame, leading to a very efficient optimization. Several new adaptations of the demons algorithm are presented, including compositive and diffeomorphic demons, as well as fluid-like and diffusion-like regularization. The resulting algorithms are all significantly faster than [13], while also retaining greater flexibility. Our algorithms are validated and compared using cortical surface models.

4.1 Introduction

Non-rigid shape registration represents an important area of research in medical imaging, in particular for cortical shape analysis. The highly variable, convoluted geometry of the cortical boundary together with an abundance of MR data present a significant challenge for fully automating reliable correspondence searches. Current methods for parametric shape registration range from conformal maps and intrinsic embeddings via the Laplace-Beltrami (LB) operator [53, 57, 58] to the more direct adaptations of image registration algorithms in the Euclidian domain. The latter approach is appealing, as non-linear medical image registration has by now become a mature field with several well-validated methods. For example, an approach taken by [59] maps subcortical shapes directly to the 2D plane, and performs the usual fluid registration of mean curvature and conformal factor features following [11] to achieve final correspondence. The method is fast and reliable, provided a consistent set of boundaries is introduced to enable the initial parameterization. The boundary constraint requires a strong prior on the final correspondence search, before any registration can be attempted at all.

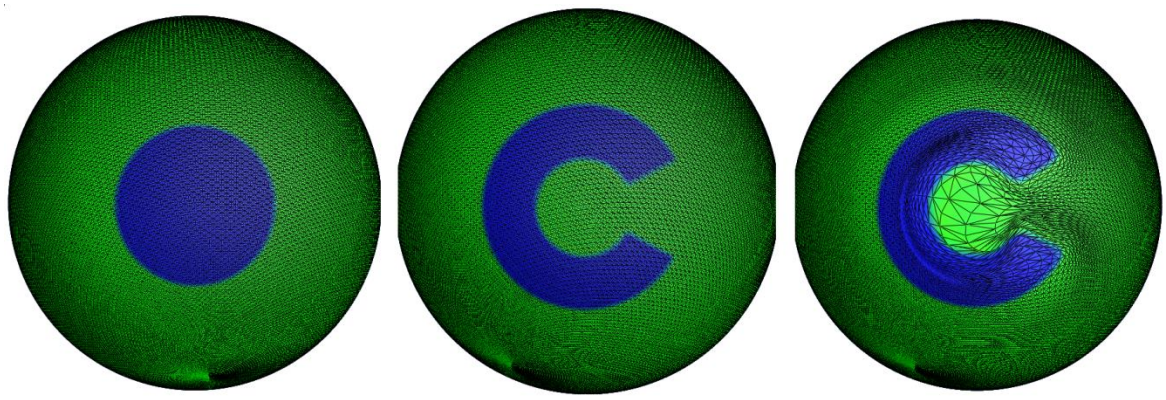

Fig. 64. Smooth spherical circle to C. No registration paper is complete without the "circle to C." Here, we used diffeomorphic demons with fluid- and diffusion-like regularization.

This illustrates the significant advantage that using a parametric domain of the same topology as the shape – for example the 2-sphere – offers compared to the Euclidian domain. In effect, using the spherical domain enables truly automated parametric registration of genus-zero shapes. As a result, several adaptations of Euclidian registration to the sphere have been proposed. Thompson et al. [1] proposed elastic matching based on sulcal landmarks, constraining the curve-induced flow by the Cauchy-Navier differential operator, and correcting explicitly for metric distortions. In [58, 60], the authors extend Miller's LDDMM framework [61] to point and curve-set registration on the sphere. Closer to this work, in [21, 22] the authors propose landmark-free methods on the sphere, minimizing the sum of squared distances (SSD) between corresponding curvature maps and curvature-derived feature functions. The optical flow algorithm is adapted to the sphere in [21], and solved using a narrow band approach, while in [22] a straight-forward optimization of coordinates is performed directly on the surface. In a more recent effort, Yeo extended the very efficient diffeomorphic demons algorithm [12, 62] to spherical images, and showed that registering curvature and thickness features of the cortex leads to robust shape registration [13]. The resulting algorithm can accurately register two cortical surfaces of high resolution (150K vertices) in under 5 minutes, while maintaining invertible warps. This is quite an impressive result, since FreeSurfer [22], perhaps the most popular tool for cortical surface alignment, takes on the order of 1 hour.

Inspired by the recent work on spherical shape registration, we revisit the spherical registration problem from the perspective of adapting well-known Euclidean registration approaches. At the heart of many image registration algorithms, one finds a Gaussian convolution of either the displacement field itself, or the update step/instantaneous velocity of the field. On the other hand, it is often convenient to decompose the velocity or the displacement

over an orthogonal basis, formed by the eigenfunctions of a suitable differential operator. As the most basic example, Gaussian convolution can be performed quickly in the Fourier domain; this is often used to approximate the solution to the fluid equation [63]. A more precise solution is offered in [64], where the decomposition is in eigenfunctions of the fluid operator itself. The many variants of the demons algorithm likewise require a convolution of the field ("diffusion-like" demons) or the update step ("fluid-like" demons) with a smoothing kernel for regularization, which is typically done in the Fourier domain [12]. However, the adaptation of this idea to the sphere is not trivial, since the well-known scalar spherical harmonics cannot be applied directly to canonical coordinates of tangential vector fields. To mitigate this problem, Yeo et al. [13] use a straightforward recursive smoothing scheme. This limits the possible kernel range, as the execution time depends directly on the size of the kernel. In this work, we derive a smoothing technique for tangential vector fields based on *vector* spherical harmonics (VSH). Vector spherical harmonics form a suitable basis in which to perform the required convolution, as they are eigenfields of the Casimir operator restricted to the sphere. This allows for a natural extension of Mercer's theorem to spherical vector fields for the purpose of fast spherical field regularization. Such an approach has all the usual benefits of performing convolution in the spectral domain.

We implement a fast VSH transform, and apply our smoothing in several natural adaptations of the demons algorithm and the spherical fluid algorithm. We compare the performance of the proposed algorithms based on 100 white matter (WM) surfaces, using both synthetic warps and true cross-subject registration. We conclude that the combined fluid- and diffusion-like diffeomorphic demons offer the best accuracy, followed closely by the fluid algorithm. Both of these approaches significantly outperform compositive demons, as well as

diffeomorphic demons with only fluid-like or diffusion-like regularization, such as implemented in [13].

The remainder of the chapter is organized as follows. The first section describes fast heat kernel smoothing for spherical fields. The second and third sections outline the demons and fluid adaptations to the sphere based on the proposed vector smoothing. The fourth section describes some implementation details. The fifth section compares the results across our methods, and the sixth concludes the chapter.

4.2 Heat Kernel for Spherical Vector Fields

In direct analogue to Fourier series on $\mathbb{R}^n$, and scalar spherical harmonics – the eigenfunctions of the Laplacian and the scalar LB operator on $\mathbb{S}^2$, VSH can be derived from the Casimir operator, or the Laplacian operator on spherical tangential fields. The curvature of the sphere implies a non-trivial parallel transport, which complicates the vector Laplacian form and distinguishes it from the scalar case. In canonical coordinates, the vector Laplacian of $\mathbf{v} = \mathrm{v}_\theta \boldsymbol{e}_\theta + \mathrm{v}_\varphi \boldsymbol{e}_\varphi$ is written as

$$\Delta \mathbf{v} = \left[-\left(\frac{\partial^2}{\partial \theta^2} + \cot\theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial \varphi^2} - \frac{1}{\sin^2\theta} \right) v_\theta + \frac{2\cot\theta}{\sin\theta} \frac{\partial v_\varphi}{\partial \varphi} \right] \boldsymbol{e}_\theta$$

$$+ \left[-\left(\frac{\partial^2}{\partial \theta^2} + \cot\theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial \varphi^2} - \frac{1}{\sin^2\theta} \right) v_\varphi - \frac{2\cot\theta}{\sin\theta} \frac{\partial v_\theta}{\partial \varphi} \right] \boldsymbol{e}_\varphi. \quad (4.1)$$

Vector spherical harmonics satisfy $\Delta \mathbf{A}_{\boldsymbol{lm}} = l(l+1)\mathbf{A}_{\boldsymbol{lm}}$, and can be defined as the gradient of the scalar harmonics Y_{lm}, and its orthogonal complement (see, e.g. [65]).

$$\mathbf{B}_{lm} = \frac{1}{\sqrt{l(l+1)}}\left[\frac{\partial Y_{lm}}{\partial \theta}\boldsymbol{e}_\theta + \frac{1}{\sin\theta}\frac{\partial Y_{lm}}{\partial \varphi}\boldsymbol{e}_\theta\right], \tag{4.2}$$

$$\mathbf{C}_{lm} = -\boldsymbol{e}_r \times \mathbf{B}_{lm}.$$

This leads to harmonic decomposition of a spherical vector field into

$$\mathbf{v} = \sum_{\substack{l=1 \\ |m|\le l}}^{\infty} [\mathbf{B}_{lm} f^B(l,m) + \ \mathbf{C}_{lm} f^C(l,m)], \quad f^A(l,m) = \langle \mathbf{v}, \mathbf{A}_{\boldsymbol{lm}} \rangle_{L^2(\mathrm{T}\mathbb{S}^2)} \tag{4.3}$$

Taking advantage of the eigenfields, we can extend Mercer's Theorem [66] to spherical fields, and define the heat kernel as

$$\boldsymbol{K}_\sigma(p,q) = \sum_{l=1}^{\infty} e^{-l(l+1)\sigma} \sum_{m=-l}^{l} \mathbf{B}_{lm}(p) \otimes \mathbf{B}_{lm}(q) + \mathbf{C}_{lm}(p) \otimes \mathbf{C}_{lm}(q), \tag{4.4}$$

where $\otimes$ is the tensor product. The kernel in (5) represents Green's function of the vector isotropic diffusion equation $\frac{\partial \mathbf{v}}{\partial t} = -\Delta \mathbf{v}$, $\sigma = \sqrt{2t}$. Applying the kernel to a field leads to an expression which is similar to the scalar harmonics case [67]:

$$\mathbf{K}_\sigma * \mathbf{v}(p) = \sum_{\substack{l=1 \\ |m|\le l}}^{\infty} e^{-l(l+1)\sigma} \left[\mathbf{B}_{lm}(p) \int_{\mathbb{S}^2} \mathbf{B}_{lm}(q)\mathbf{v}(q)d\mu(q) + \ \mathbf{C}_{lm}(p) \int_{\mathbb{S}^2} \mathbf{C}_{lm}(q)\mathbf{v}(q)d\mu(q)\right]. \tag{4.5}$$

It is easy to see that all that is required for an efficient heat kernel smoothing of a spherical field is a forward harmonic transform followed by an $O(n)$ operation and an inverse transform.

4.3 Spherical Demons

The general idea behind a demons approach [62] is a two-step optimization, in which the first step represents a search for the update direction of the current warp, and the second – the regularization of the new warp resulting from this update. Thus, for fields $\boldsymbol{u}, \boldsymbol{g}, \boldsymbol{G}: \mathbb{R}^n \to \mathbb{R}^n$, we have the following optimization problems

$$\boldsymbol{u} = \arg\ \min_{\boldsymbol{u}} \left(\|S - T * \{\boldsymbol{g} * \boldsymbol{u}\}\|_2^2 + \frac{1}{\sigma} \mathrm{dist}(\boldsymbol{g}, \{\boldsymbol{g} * \boldsymbol{u}\}) \right), \tag{4.6}$$

where $\boldsymbol{u}$ is a "hidden" transformation, and the regularization

$$\boldsymbol{G} = \arg\ \min_{\boldsymbol{G}} \left(\frac{1}{\sigma_{\boldsymbol{G}}} R(\boldsymbol{G}) + \frac{1}{\sigma} \mathrm{dist}(\boldsymbol{G}, \{\boldsymbol{g} * \boldsymbol{u}\}) \right), \tag{4.5}$$

with the update $\boldsymbol{g}_{t+1} = \boldsymbol{G}_t$. Here S, T are fixed and moving images with $T * \boldsymbol{g}: \mathbb{R}^n \to \mathbb{R}$ defined using the Lagrangian frame by $\{T * \boldsymbol{g}\}[x + \boldsymbol{g}(x)] = T[x]$. The optimized field $\boldsymbol{g}$ is the warp bringing the two images into correspondence. The regularization term $R(\boldsymbol{G})$ is generally taken as a norm of a differential operator, so that the minimization can be achieved with a convolution. A well-known example, minimizing the harmonic energy in $\mathbb{R}^n$ is equivalent to a Gaussian smoothing of the displacement field $\boldsymbol{G}$. Likewise, the second term in the first equation (6) can be interpreted as a penalty on the harmonic energy of $\boldsymbol{u}$, as well as its norm, and can be smoothed with a Gaussian kernel. Smoothing the displacement field is often termed "diffusion-like regularization," and smoothing the update, "fluid-like regularization" [12]. The unique advantage of the demons family of algorithms is precisely the separation of the two optimization problems: each cost can be optimized very efficiently with either a linear approximation or a fast

convolution. Lastly, a more recent modification of the demons framework [12] introduced the idea of maintaining diffeomorphic warps by passing each update step $\boldsymbol{u}$ through the exponential $\boldsymbol{u} \to exp(\boldsymbol{u})$, thus ensuring invertibiliy. Since diffeomorphisms form a Lie group under composition, this approach guarantees a smooth invertible final warp $\boldsymbol{g}$. Adapting the demons approach to spherical images $S, T: \mathbb{S}^2 \to \mathbb{R}$, we optimize over $\boldsymbol{u}, \boldsymbol{g}, \boldsymbol{G}: \mathbb{S}^2 \to T_x\mathbb{S}^2$, and following the convention in [13], define $T * \boldsymbol{g}$ by $\{T * \boldsymbol{g}\}[p(x, \boldsymbol{g}\{\boldsymbol{x}\})] = T[x]$, where

$$p(x, \boldsymbol{g}\{\boldsymbol{x}\}) = x\sqrt{1 - \|\boldsymbol{g}\{\boldsymbol{x}\}\|^2} + \boldsymbol{g}\{\boldsymbol{x}\}. \tag{4.8}$$

Although such a parameterization of the warp contains a nonlinearity, as the geodesic length of the displacement is the arcsine of $\|\boldsymbol{g}\{x\}\|$, it leads to significantly simpler computations than an arc length parameterization. Indeed, given $p(x, \boldsymbol{g}\{x\})$, it is easy to compute $\boldsymbol{g}\{x\}$ by $\boldsymbol{g}\{\boldsymbol{x}\} = -\mathcal{G}^2(x)p(x, \boldsymbol{g}\{x\})$, where $\mathcal{G}$ is the cross-product matrix, as suggested in [13]. In solving the first optimization problem (6), we deviate from [13], who optimize the problem directly in the original image space, and follow [12] more closely: we reformulate the problem as

$$\boldsymbol{u} = \arg\min_{\boldsymbol{u}} \left(\|S - [T * \boldsymbol{g}] * \boldsymbol{u}\|_2^2 + \frac{1}{\sigma}\|\boldsymbol{u}\|^2 \right). \tag{4.9}$$

This leads to a straightforward linear problem, following the linearization of $\|S - [T * \boldsymbol{g}] * \boldsymbol{u}\|_2^2$, which can be solved separately for every point on $\mathbb{S}^2$.

$$\boldsymbol{u}\{p\} = \frac{S\{p\} - [T * \boldsymbol{g}] * \boldsymbol{u}\{p\}}{\left\|\vec{\nabla}[T * \boldsymbol{g}]\{p\}\right\|^2 + \frac{1}{\sigma^2\{p\}}} \vec{\nabla}[T * \boldsymbol{g}]\{p\}, \tag{4.10}$$

where $\sigma^2\{p\}$ is a normalization term controlling for image noise. Note that we have omitted the matrix $\frac{\partial p(x, \boldsymbol{u})}{\partial \boldsymbol{u}}$, because at $\boldsymbol{u} = \boldsymbol{0}$ it is simply the identity.

The second optimization step consists entirely in applying a smoothing kernel to the composition $\boldsymbol{g_t} * \boldsymbol{u_{t+1}}$. In this sense, the energetic norm used in $R(\boldsymbol{G})$ is in practice defined by the kernel of choice, rather than the other way around [58]. However, as mentioned earlier, convolving tangential vector fields on the sphere does not generalize directly from scalar convolution as in $\mathbb{R}^n$ due to the non-trivial parallel transport operator $\boldsymbol{T}(p,q)\colon T_p\mathbb{S}^2 \to T_q\mathbb{S}^2$. A straightforward solution is to recursively approximate the kernel by repeated application of $\boldsymbol{T}(p,q)$ over some neighborhood for each vertex on a mesh. In this case, the effective size of the kernel depends directly on the number of smoothing iterations performed. This is an approach taken in [13], and allows one to create custom kernels depending on the weighting function. Thus, equation (7) is approximately solved by

$$\widetilde{\boldsymbol{G}}(p) = \sum_{q \in \mathcal{N}(p)} \lambda(p,q)\boldsymbol{T}(p,q) * \{\boldsymbol{g} * \boldsymbol{u}\}(q), \tag{4.11}$$

where $\mathcal{N}(p)$ is the 1-ring of the vertex p, and $\lambda(p,q)[\sigma]$ is normalized to add up to 1 over the 1-ring, and monotonically increasing with σ. While this appears to work well in practice, the approach can only be applied to kernels of a limited size. We replace this with vector heat kernel smoothing via the VSH, which eliminates the limitation on the kernel size and speeds up the process considerably. Further, the energy minimized can actually be defined in closed form, and computed essentially for free as part of the VSH transform. One additional advantage of having a low-cost smoothing technique for spherical fields is that it enables us to naturally extend the fluid-like regularization of Euclidean demons to the sphere. This modification turns out crucial for improving cortical correspondence accuracy, as we will see shortly. We thus have our final demons algorithm family:

Algorithm 1 (VSH-based spherical demons)

Given images $S, T: \mathbb{S}^2 \to \mathbb{R}$, max step R_{max}, tolerance

Initialize $\boldsymbol{u}, \boldsymbol{g_0}: \mathbb{S}^2 \to T_x\mathbb{S}^2$ to be uniformly 0.

While(t < max_iterations AND $\frac{\|S - T*\{\boldsymbol{g_{t-1}}\}\|_2^2 - \|S - T*\{\boldsymbol{g_t}\}\|_2^2}{\|S - T*\{\boldsymbol{g_0}\}\|_2^2}$ > tol.)

1. Compute update $\boldsymbol{u}$ using (10)
2. Find σ^2, so that $R_{max} \leq \|\boldsymbol{u}(p)\| \; \forall p \in \mathbb{S}^2$ [12]
3. For fluid-like registration, $\boldsymbol{u} \to \mathbf{K}_{\sigma \text{ fluid}} * \boldsymbol{u}$ using (5)
4. For diffeomorphic registration, $\boldsymbol{u} \to exp(\boldsymbol{u})$ [13]
5. Compute $\boldsymbol{g_{t+1}}\{x\} = -\mathcal{G}^2(x) p(x, \boldsymbol{g_t}\{x\} * \boldsymbol{u})$
6. For diffusion-like regularization, $\boldsymbol{g_{t+1}} \to \mathbf{K}_{\sigma \text{ diff}} * \boldsymbol{g_{t+1}}$

End While

Return $\boldsymbol{g}$

4.4 Spherical Fluid Registration

The fluid registration paradigm, first introduced in [11], differs markedly from the demons family in that the fidelity term $\|S - T * \boldsymbol{g}\|_2^2$ is only represented as defining the body force $\boldsymbol{F}(\boldsymbol{g})$ of the simplified Navier-Stokes equation

$$\mu \Delta \boldsymbol{u} + (\lambda + \mu) \vec{\nabla}\left(\vec{\nabla} \cdot \boldsymbol{u}\right) = -\boldsymbol{F}(\boldsymbol{g})$$

$$\boldsymbol{F}[x, \boldsymbol{g}(x,t)] = (S[x] - T[\, p(x, -\boldsymbol{g}\{x,t\})]) \vec{\nabla} T[\, p(x, -\boldsymbol{g}\{x,t\})]. \tag{4.12}$$

The equations model the behavior of a viscous fluid, driven by the instantaneous forces resulting from image mismatch. Unlike demons, the fluid regularization has no memory of the previous step. Instead, the instantaneous velocity is explicitly integrated over time, allowing for very large deformations. The power and popularity of the fluid framework are largely due to this flexibility.

Optimization is performed in the Eulerian reference frame. In this frame, the velocity $\boldsymbol{u}(x,t)$ describes the motion of the particle at position x and time t. The field $\boldsymbol{g}(x,t)$ implies that from time 0, this particle underwent a transformation parameterized by $\boldsymbol{T}(x, x_0) * \boldsymbol{g}(x,t)$, using our convention (8). It is easy to see that $x_0 = p(x, -\boldsymbol{g}\{x,t\})$.

The algorithm's only memory of previous iterations is expressed in the material derivative, which accounts for the field Jacobian $D\boldsymbol{g}$ to update the displacement.

$$\frac{\partial \boldsymbol{g}}{\partial t} = [Id - D\boldsymbol{g}]\boldsymbol{u}. \qquad (\mathbf{4}.6)$$

To update the field in the spherical domain, one must also account for the non-linearity in the displacement parameterization. Since the particle to reside at x after applying $\delta\boldsymbol{g} = \delta t \frac{\partial \boldsymbol{g}}{\partial t}$ is to be found at $p[r,-\boldsymbol{T}(x,r) * \delta\boldsymbol{g}]$ in the original image, where for clarity we mean $r = p(x,-\boldsymbol{g}\{x,t\})$, we must update the field by

$$\boldsymbol{g}(x,t+\delta t) = \mathcal{G}^2(x)p[r,-\boldsymbol{T}(x,r) * \delta\boldsymbol{g}]. \qquad (4.7)$$

Because the field is updated at fixed coordinate points on the sphere, the matrices $\mathcal{G}(x)$ can be pre-computed offline just as in the demons algorithm.

The most computationally intensive aspect of the fluid registration approach is solving for the velocity $\boldsymbol{u}$, given the body force. In the original formulation [11], the successive over-relaxation (SOR) approach was used, which, while accurate, proved prohibitively computationally expensive. A significant improvement on speed was achieved in [64], where the authors derived and applied eigenfunctions of the operator (12). However, the most common simplification of this problem in $\mathbb{R}^n$ is to apply a Gaussian filter to the field, which can be shown [63, 68] to approximate (12). Following similar arguments, we propose an analogue of fluid approximation on the sphere, using VSH-based vector heat kernel smoothing.

Algorithm 2 (spherical fluid registration)

Given images $S, T: \mathbb{S}^2 \to \mathbb{R}$, max step R_{max}, tolerance

Initialize $\boldsymbol{u}, \boldsymbol{g}: \mathbb{S}^2 \to T_x\mathbb{S}^2$ to be uniformly 0. n = 0.

While(n < max_iterations AND $\frac{\|S[x]-T[\,p(x,-\boldsymbol{g}\{x,t\})]\|_2^2 - \|S[x]-T[\,p(x,-\boldsymbol{g}\{x,t+\delta t\})]\|_2^2}{\|S[x]-T[\,p(x,\mathbf{0})]\|_2^2}$ > tolerance)

1. Compute body force $\boldsymbol{F}$ using (12)
2. Set $\boldsymbol{u} \to -\mathbf{K}_\sigma * \boldsymbol{F}$
3. Compute $\frac{\partial \boldsymbol{g}}{\partial t}$ using (13), set $\delta t = R_{max}/\max(\left\|\frac{\partial \boldsymbol{g}}{\partial t}\right\|)$
4. Compute $\boldsymbol{g}(x, t+\delta t)$ using (14), set n = n + 1

End While

Return $\boldsymbol{g}$

4.5 Implementation Issues

4.5.1 VSH Computation

We apply the exact quadrature method for spherical harmonics sampled at regular canonical coordinates [65]. We vectorize the SpharmonicKit implementation [29] following [65], except normalizing vector and scalar harmonics $\|\mathbf{B}_{lm}\|, \|\mathbf{C}_{lm}\|, \|\mathrm{Y}_{lm}\| = 1$. This results in slightly different weights for computing the vector coefficients from the scalar ones. Setting the auxiliary scalar coefficients as in [65], $g^{\theta,\varphi}(l,m) = \langle \frac{1}{\sin\theta} \mathrm{v}_{\theta,\varphi}, \mathrm{Y}_{lm} \rangle_{L^2(\mathbb{S}^2)}$, the vector coefficients can be obtained from

$$f^B(l,m) = \sum_{n=-1}^{1} W_1(l,m,n) g^\theta(l+n,m) + W_2(l,m) g^\varphi(l,m) \tag{4.8}$$

$$f^C(l,m) = \sum_{n=-1}^{1} -W_1(l,m,n) g^\varphi(l+n,m) + W_2(l,m) g^\theta(l,m),$$

with the weights W_1, W_2 pre-computed offline for repeated use. An analogous expression can be obtained for computing the auxiliary vector coefficients back from the VSH coefficients. The resulting convolution algorithm requires only two forward and two inverse scalar spherical harmonic transforms, with an $O(n)$ operation. For a bandwidth of 256, which corresponds to a

grid of 512x512 vertices, typical execution time for the full convolution is around a tenth of a second.

The requirement that the fields be defined at regular spherical coordinates suggests performing the entire registration on the regular grid, only interpolating the final warp back to the original mesh coordinates. This proves faster and more stable than performing forward and reverse sampling of a field at every iteration. Further, sampling on this regular grid is denser than the typical FreeSurfer resolution of 150K vertices.

4.5.2 Demons Optimization

Some notable differences exist between our implementation of spherical demons and [13]. We choose to interpolate the moving image at regular coordinates, computing the update step from the currently warped image. Thus, there is no need to interpolate the gradient or compute the field Jacobian. However, in our implementation we use only a first order estimate of the update, while Yeo et al. use the Gauss-Newton scheme. The latter relies on empirically estimating the Hessian. Thus, in [1] there is a heavier computational burden for each iteration, but fewer iterations are required. We note that a Gauss-Newton scheme could be applied directly to our framework as well. Also unlike [1], we follow [12] in setting the gradient term in (10) to be the symmetric gradient. This was found to improve convergence.

An additional aspect of the diffeomorphic demons approach is the "scaling and squaring" procedure. We follow [13], but perform the procedure based on regular spherical grids. The advantage of this approach is that sampling from a regular grid onto an irregular one is an $O(1)$ operation per vertex, whereas sampling the other way generally grows with the number of

triangles even for fast samplers. The repeated sampling during "scaling and squaring" means that using a warp defined on a regular grid leads to faster exponentiation. We note also that the "slow" kind of spherical sampling (irregular to regular) must still be done at least once per iteration in the demons pipeline. For this, our in-house fast sampling algorithm performs roughly 2.5 triangle intersection checks per vertex for the spherical grids we use, leading to compute times less than 1 second.

4.5.3 Fluid Optimization

Our fluid registration is adapted to the sphere to be very close to [63], with the exception of the fidelity term (here, SSD). The other main difference lies in not using regridding during our optimization. In [11, 63], the authors restart the optimization process setting $T[x] \rightarrow T[\,p(x, -\boldsymbol{g}\{x,t\})]$, $\boldsymbol{g} = \mathbf{0}$, if the Jacobian determinant $|D\boldsymbol{g}|$ falls below some threshold. This "regridding" step is used to alleviate problems caused by discretization, which can make the warp non-invertible. The final output warp $\boldsymbol{g}$ is then the composition of all the intermediate warps. Here, we chose not to do this for a more fair comparison to the demons algorithm, where regridding is not customary.

A significant advantage of the fluid algorithm over the demons specifically on $\mathbb{S}^2$ is that there is no need for "slow" sampling, as we only interpolate the original image T. This is due to the choice of the Eulerian frame, and leads to significantly faster compute times per iteration.

4.6 Experiments

For our experiments, we used 100 left white matter surfaces extracted using FreeSurfer from the ADNI dataset. The initial spherical map was computed as in [55]. Throughout the experiments, we use only mean curvature as the feature function. We apply a 3-level multi-resolution scheme, smoothing the curvature maps [67] and regularizing the registration more at the first level, and relaxing both at subsequent levels. A rotational pre-registration step was applied based on fast spherical cross-correlation of the smoothed curvature [69].

Three sets of experiments were done on the cortical shapes: (1) Recovering a synthetic warp, (2) Pairwise registration, and (3) All-to-one registration. In the first experiment each subject's spherical map underwent a unique synthetic spherical warp. To synthesize a relatively large warp smooth, we first randomly seeded a warp field by its VSH coefficients and computed a smoothed inverse. This was then composed with a spherical MRF that was passed through an exponential, as in [12]. The result was a diffeomorphic deformation that was both large and highly non-linear. Further, some noise was added to the original shapes. The SSD and Laplacian norm of the warps are plotted in **Figure 15. Table 5** shows the relative error and normalized cross-correlation with the original warps for each method.

In the second experiment, shapes were randomly paired and 50 registrations were performed for each pair. The SSD and Laplacian norm of the warp are plotted in **Figure 16**. For ease of reading, here we only compare the diffeomorphic variants of the demons algorithm and the fluid algorithm.

In the third experiment, we mapped all subjects to a random target brain, using the combined diffeomorphic demons approach, and the fluid approach. We computed the resulting

brain averages and compared the result to using only rotational registration. The resulting averages are shown in **Figure 17**. Step size, σ_{fluid} and σ_{diff} were tuned experimentally to maximize agreement between recovered warp and synthetic warps. This was done separately for each method with one exception. For "fluid-like" demons, parameters were set as in the fluid registration for fair comparison. In general we found that the diffeomorphic demons favors larger step sizes than fluid registration. For the combined demons algorithm, the optimal σ_{fluid} and σ_{diff} were lower than the corresponding σ_{fluid} and σ_{diff} in "fluid-like" and "diffusion-like" versions. Overall, we found that the fluid algorithm recovered synthetic warps most accurately. However, combined demons approach resulted in visually better results for pair-wise and all-to-one registration, as well as lower final SSD.

All diffeomorphic approaches and the fluid approach resulted in invertible warps with no triangle flips. Execution time per iteration was roughly 0.5 seconds for the fluid approach and 1.5 seconds per iteration for diffeomorphic demons. Thus, registration can be achieved in as little as 10 seconds per resolution level, which is an order of magnitude improvement over [13].

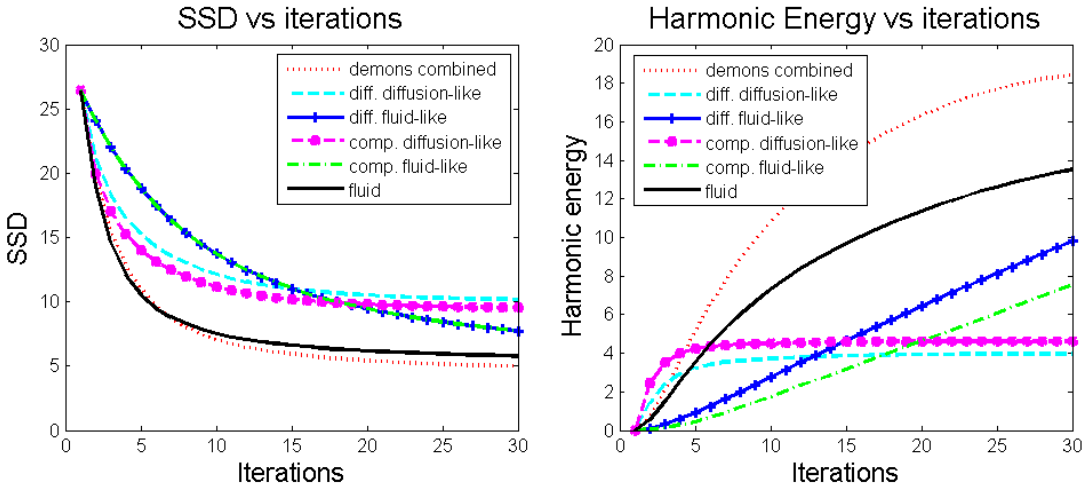

Fig. 15. Synthetic warp results. "Harmonic energy" is the vector Laplacian norm.

Table 5. Accuracy of synthetic warp recovery. Relative mean squared error and cross-correlation, averaged over 100 trials, with standard deviation. DF1 and FL1 stand for diffusion- and fluid-like diffeomorphic methods. DF2 and FL2 are the compositive versions.

	Combo	DF1	FL1	DF2	FL2	Fluid
$\frac{MSE}{\|g_{true}\|}$	0.192 +/-0.11	0.216 +/-0.11	0.181 +/-0.086	0.236 +/-0.11	0.181 +/-0.086	**0.162 +/-0.063**
CC	0.875 +/-0.086	0.833 +/-0.099	0.884 +/-0.062	0.807 +/-0.11	0.884 +/-0.062	**0.890 +/-0.046**

4.7 Conclusion

We presented a family of fast spherical registration tools for registering cortical surfaces, adapting several well-known image registration algorithms to the sphere. The full gamut of the demons approaches as well as the large deformation fluid approach are generalized to the 2-sphere, the latter being the first such adaptation. Our methods are based on fast convolution of spherical vector fields in the spectral domain, leading to perhaps some of the most efficient landmark-free cortical surface registration algorithms. The algorithms are validated on synthetic spherical warps, achieving an average normalized cross-correlation of nearly 0.9, where 1 would be a perfect recovery. Registration between pairs of real brains also shows promising results, leading to robust diffeomorphic registration in as little as 30 seconds.

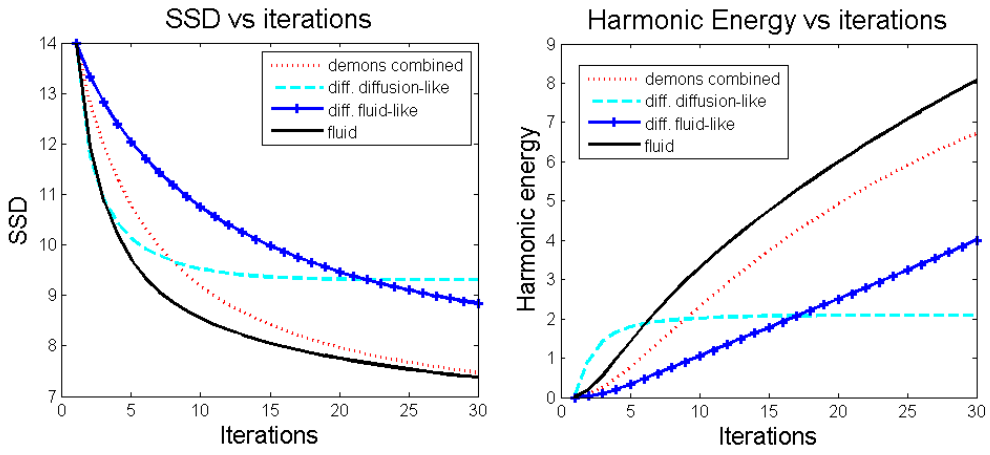

Fig. 16. Pair-wise warp results. "Harmonic energy" is the vector Laplacian norm.

A comparison of our methods reveals that the combined diffeomorphic demons and the fluid registration outperform the others in terms of minimizing geometric and image mismatch. A limitation of the tools tested here is their reliance on auxiliary measures of shapes, such as mean curvature. We recognize that an explicit correction for geometric distortion may improve our results. In particular, the fluid algorithm is flexible enough to handle any geometry-driven mismatch function without any additional modification.

A more complete version of this work will compare the implementation in [1] and [11] to our methods based on agreement with manually delineated cortical regions. A future direction of a more theoretical flavor would be to explore the relationship between the particular energetic norm used here – the vector Laplacian – and the harmonic energy of the automorphism encoded in the vector field. Such an exploration would complete the adaptation of the tools used here from a theoretical perspective in much the same way we have adapted these tools computationally.

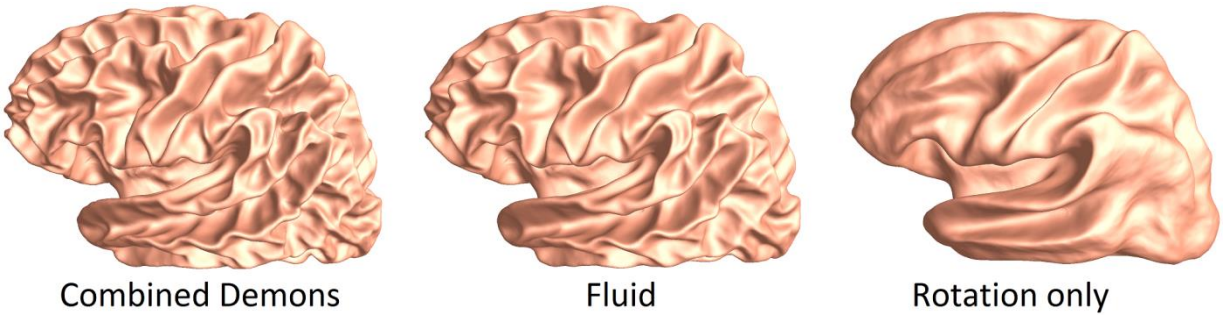

Fig. 17. 100-Brain Averages. More detailed geometry suggests better overall registration.

CHAPTER 5

Maximizing Power to Track Alzheimer's Disease and MCI Progression by LDA-Based Weighting of Longitudinal Ventricular Surface Features

This chapter is adapted from:

B. A. Gutman, X. Hua, P. Rajagopalan, Y.-Y. Chou, Y. Wang, I. Yanovsky, A. W. Toga, C. R. Jack Jr, M. W. Weiner, and P. M. Thompson, "Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features," *Neuroimage*, vol. 70, pp. 386-401, 2013.

Abstract

We propose a new method to maximize biomarker efficiency for detecting anatomical change over time in serial MRI. Drug trials using neuroimaging become prohibitively costly if vast numbers of subjects must be assessed, so it is vital to develop efficient measures of brain change. A popular measure of efficiency is the minimal sample size (n80) needed to detect 25% change in a biomarker, with 95% confidence and 80% power. For multivariate measures of brain change, we can directly optimize n80 based on a Linear Discriminant Analysis (LDA). Here we use a supervised learning framework to optimize n80, offering two alternative solutions. With a new medial surface modeling method, we track 3D dynamic changes in the lateral ventricles in 2,065 ADNI scans. We apply our LDA-based weighting to the results. Our best average n80 - in two-fold nested cross-validation - is 104 MCI subjects (95% CI: [94,139]) for a 1-year drug trial,

and 75 AD subjects [64,102]. This compares favorably with other MRI analysis methods. The standard "statistical ROI" approach applied to the same ventricular surfaces requires 165 MCI or 94 AD subjects. At 2 years, the best LDA measure needs only 67 MCI and 52 AD subjects, versus 119 MCI and 80 AD subjects for the stat-ROI method. Our surface-based measures are unbiased: they give no artifactual additive atrophy over three time points. Our results suggest that statistical weighting may boost efficiency of drug trials that use brain maps.

5.1 Introduction

Biomarkers of Alzheimer's disease based on brain imaging must offer relatively high power to detect longitudinal changes in subjects scanned repeatedly over time [15-17]. Even so, recruitment and scanning are costly, and a drug trial may not be attempted at all, unless disease-slowing effects can be detected in an achievable sample size, and in a reasonable amount of time. Imaging measures from standard structural MRI show considerable promise. Their use stems from the premise that longitudinal changes may be more precisely and reproducibly measured with MRI than comparable changes in clinical, CSF, or proteomic assessments; clearly, whether that is true depends on the measures used. Brain measures that are helpful for diagnosis, such as PET scanning to assess brain amyloid or CSF measures of amyloid and tau proteins, may not be stable for longitudinal trials that aim to slow disease progression. As a result, there is interest in testing the reproducibility of biomarkers, as well as methods to weight or combine them to make the most of all the available measures [70].

Recent studies have tested the reproducibility and accuracy of a variety of MRI-derived measures of brain change. Several are highly correlated with clinical measures, and can predict future decline on their own, or in combination with other relevant measures. Although not the only important consideration, some analyses have assessed which MRI-based measures show greatest effect sizes for measuring brain change over time, while avoiding issues of bias and asymmetry that can complicate longitudinal image analysis [71-73], and while avoiding removing scans from the analysis that may lead to unfairly optimistic sample size estimates [17, 71]. Promising MRI-based measures include the brain boundary shift integral [74, 75], the ventricular boundary shift integral [74] and measures derived from anatomical segmentation software such as Quarc or FreeSurfer, some of which have been recently modified to handle longitudinal data more accurately [76-79].

Although several approaches are possible, one type of power analysis, advocated by the ADNI Biostatistics Core [80], is to estimate the minimal sample size required to detect, with 80% power, a 25% reduction in the mean annual change, using a two-sided test and standard significance level $\alpha = 0.05$ for a hypothetical two-arm study (treatment versus placebo). The estimate for the minimum sample size is computed from the formula below. $\widehat{\beta}$ denotes the annual change (average across the group) and $\hat{\sigma}^2{}_D$ refers to the variance of the annual rate of change.

$$n = \frac{2\hat{\sigma}^2{}_D\left(z_{1-\alpha/2} + z_{power}\right)^2}{\left(0.25\hat{\beta}\right)^2} \quad (1)$$

Here z_α is the value of the standard normal distribution for which $P[Z < z_\alpha] = \alpha$. The sample size required to achieve 80% power is commonly denoted by n80. Typical n80s for competitive methods are under 150 AD subjects and under 300 MCI subjects; the larger numbers for MCI reflect the fact that brain changes tend to be slower in MCI than AD and MCI is an etiologically

more heterogeneous clinical category. For this reason, it is harder to detect a modification of changes that are inherently smaller, so greater sample sizes are needed to guarantee sufficient power to detect the slowing of disease. In addition, there is some interest in prevention trials targeting cognitively normal subjects who are at risk for AD by virtue of a family history or specific genetic profile (e.g., ApoE genotype); for these and other cohorts, efficiency must be a high priority, and measures that can distinguish AD from normal aging may be unable to track subtle changes efficiently in controls.

Many algorithms can detect localized or diffuse changes in the brain, creating detailed 3D maps of changes [53, 81, 82], but the detail in the maps they produce is often disregarded when making sample size estimates according to (1), as the formula expects a single, univariate measure of change. In other words, it requires a single number, or 'numeric summary' to represent all the relevant changes occurring within the brain. To mitigate this problem, Hua et al. [83] defined a "statistical ROI" based on a small sample of AD subjects by thresholding the *t*-statistic of each feature (voxel) and summing the relevant features over the ROI; this approach was initially advocated in the FDG-PET literature to home in on regions that show greatest effects [84]. In spirit, the statistical ROI is a rudimentary supervised learning approach, as it finds regions that show detectable effects in a training sample, and uses them to empower the analysis of future samples; the samples used are non-overlapping and independent, to avoid circularity. However, a simple threshold-based masking is known to potentially eliminate useful features, as binarisation loses a lot of the information present in continuous weights [85]. While many studies have used machine learning to predict the progression of neurodegenerative diseases and differentiate diagnostic groups such as AD, MCI, and controls [40, 86, 87], we found no attempts in the literature that used learning to directly optimize power to detect brain

change. The closest work is perhaps that of Hobbs et al. [88]. In this paper, SVM was used to separate subjects with Huntington's disease from controls, and the resulting score used to calculate sample size estimates. Our goal here was to generalize the very simple binary feature weighting in the stat-ROI approach by *directly maximizing* the power estimate in a training sample. A linear weighting that optimizes (1) directly, while using multiple features at once, is exactly analogous to a one-class Linear Discriminant Analysis (LDA), discriminating the disease class from an imaginary sample of zero mean whose covariance is identical to the disease group. We propose two approaches to perform this task: one optimizes (1) directly by Tikhonov regularization; the other is based on principal components analysis (PCA).

A common criticism of the power analysis provided by (1) is that it does not take into consideration normal ageing in non-high risk healthy subjects [73]. To mitigate this, several researchers have proposed simply subtracting the mean value of the change computed from controls, while using only the diseased subjects for a variance estimate. This same issue can be directly addressed in the LDA framework. In this case, the problem reduces to the usual 2-class LDA classification, except that the covariance structure is based on the diseased group only, and no assumption of homoscedasticity (equality of variance) is made. This modification is particularly useful for revealing subtle disease-specific atrophy in regions that also change, to some extent, with normal aging.

We apply our LDA approach to maps of surface-based "thickness" changes in the lateral ventricles over intervals of 1 and 2 years after a baseline scan. The analyses are performed on MRI scans from the ADNI-1 dataset. Using two follow-up time points, where available, in addition to the baseline scan allows us to estimate the presence of any longitudinal bias, or

intransitivity, which has been a subject of controversy in recent ADNI studies [71, 89, 90]. To register the ventricles and compute radial thickness measures, we modify the medial curve algorithm developed in Chapter 3 [91] for longitudinal registration. Our general approach is to compute a single continuous curve skeleton and use the curve to induce feature functions on the surface. Shape registration is then performed parametrically by minimizing the L^2 difference (summed squared difference) between corresponding feature functions of a pair of shapes.

We note that ventricular expansion is not specific to AD and the ventricles are often also abnormally enlarged in vascular dementia, frontotemporal lobar degeneration, traumatic brain injury, Huntington's disease, and schizophrenia, among other conditions. Even so, using detailed surface-based maps of the location of expansion – in conjunction with a modified 2-class LDA – helps to reveal aspects of ventricular expansion associated with the progression of Alzheimer's disease.

5.2 Materials and Methods

5.2.1. Alzheimer's Disease Neuroimaging Initiative

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI),

positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information, see www.adni-info.org.

Longitudinal brain MRI scans (1.5 Tesla) and associated study data (age, sex, diagnosis, genotype, and family history of Alzheimer's disease) were downloaded from the ADNI public database (http://www.loni.ucla.edu/ADNI/Data/) on July 1st 2012. The first phase of ADNI, i.e., ADNI-1, was a five-year study launched in 2004 to develop longitudinal outcome measures of Alzheimer's progression using serial MRI, PET, biochemical changes in CSF, blood and urine, and cognitive and neuropsychological assessments acquired at multiple sites similar to typical clinical trials.

All subjects underwent thorough clinical and cognitive assessment at the time of scan acquisition. All AD patients met NINCDS/ADRDA criteria for probable AD [92]. The ADNI protocol lists more detailed inclusion and exclusion criteria [93, 94], available online http://www.alzheimers.org/clinicaltrials/fullrec.asp?PrimaryKey=208). The study was conducted according to the Good Clinical Practice guidelines, the Declaration of Helsinki and U.S. 21 CFR Part 50-Protection of Human Subjects, and Part 56-Institutional Review Boards. Written informed consent was obtained from all participants before performing experimental procedures, including cognitive testing.

5.2.2. MRI acquisition and image correction

All subjects were scanned with a standardized MRI protocol developed for ADNI [95]. Briefly, high-resolution structural brain MRI scans were acquired at 59 ADNI sites using 1.5 Tesla MRI scanners (GE Healthcare, Philips Medical Systems, or Siemens). Additional data was collected at 3-T, but is not used here as it was only collected on a subsample that is too small for making comparative assessments of power. Using a sagittal 3D MP-RAGE scanning protocol, the typical acquisition parameters were repetition time (TR) of 2400 ms, minimum full echo time (TE), inversion time (TI) of 1000 ms, flip angle of 8°, 24 cm field of view, 192x192x166 acquisition matrix in the x-, y-, and z- dimensions, yielding a voxel size of 1.25x1.25x1.2 mm^3, later reconstructed to 1 mm isotropic voxels. For every ADNI exam, the sagittal MP-RAGE sequence was acquired a second time, immediately after the first using an identical protocol. The MP-RAGE was run twice to improve the chance that at least one scan would be usable for analysis and for signal averaging if desired.

The scan quality was evaluated by the ADNI MRI quality control (QC) center at the Mayo Clinic to exclude failed scans due to motion, technical problems, significant clinical abnormalities (e.g., hemispheric infarction), or changes in scanner vendor during the time-series (e.g., from GE to Philips). Image corrections were applied using a standard processing pipeline consisting of four steps: (1) correction of geometric distortion due to gradient non-linearity [96], i.e. "gradwarp" (2) "B1-correction" for adjustment of image intensity inhomogeneity due to B1 non-uniformity [95], (3) "N3" bias field correction for reducing residual intensity inhomogeneity [97], and (4) phantom-based geometrical scaling to remove scanner and session specific calibration errors [98].

5.2.3. The ADNI-1 dataset

For our experiments, we analyzed data from 683 ADNI subjects with baseline and 1 year scans, and 542 subjects with baseline, 1 year and 2 year scans. The former group consisted of 144 AD subjects (age at screening: 75.5 +/-7.4, 67 female (F)/77 male (M)), 337 subjects with Mild Cognitive Impairment (MCI) (74.9 +/-7.2, 122 F/215 M), and 202 age-matched healthy controls (NC) (76.0 +/-5.1, 95 F/107 M). The 2-year group (i.e., people with scans at baseline, and after a 1-year and 2-year interval) had 111 AD (75.7 +/-7.3, 52 F/59 M), 253 MCI (74.9 +/-7.1, 87 F/166 M), and 178 NC (76.2 +/- 5.2, 85 F/93 M) subjects. All raw scans, images with different steps of corrections, and the standard ADNI-1 collections are available to the general scientific community at http://www.loni.ucla.edu/ADNI/Data/. We used exactly all ADNI subjects available to us (on Feb. 1, 2012) who had both baseline and 12 month scans, and all subjects with 24 month scans (available July 1, 2012). The use of all subjects without data exclusion has been advocated by [17] and [71], because any scan exclusion can lead to power

estimates that are unfairly optimistic, and many drug trials prohibit the exclusion of any scans at all.

5.2.4 Surface Extraction

Our surfaces were extracted from 9-parameter affine-registered, fully processed T1-weighted anatomical scans. We used a modified version of Chou's registration-based segmentation [99], using inverse-consistent fluid registration with a mutual information fidelity term [81]. To avoid issues of bias and non-transitivity, we segmented each of our subjects' two or three scans separately. In this approach, a set of hand-labeled "templates" are aligned to each scan, with multiple atlases being used to greatly reduce error. There were an equal number of templates from each of the three diagnostic groups, with an equal number of males and females in each. However, using only AD or MCI templates instead is unlikely to have any measurable effect on the segmentation, due to the fact that many templates are used.

Table 1: Available scans for ADNI-1 on February 1, 2012, for 12 months and July 1, 2012, for 24 months. Total number of scans used: N = 2065.

	Screening	**12Mo**	**24Mo**
AD	200	144	111
MCI	408	337	253
Normal	232	202	178
Total	**840**	**683**	**542**

5.2.5 Medial Curve-Based Surface Registration

In this study, we focus on mapping changes in the lateral ventricles, a fluid-filled space that expands as brain atrophy progresses (**Figure 1**). Clearly, other features could be used with our multivariate approach, and it would be equally possible to apply the learning of

discriminative features from voxel-based maps of changes throughout the brain, as measured by tensor-based morphometry, for example. The method is completely general, and it could even be simultaneously applied to multiple types of features; for example, thickness measures from anatomical surfaces (cortical and subcortical), maps of volumetric changes throughout the brain, and any other biomarkers such as maps of brain amyloid or CSF analytes. In that case the meaning of a 25% slowing the pattern of change would be less intuitive, but it might identify biomarkers whose progression is slowed by a treatment. For simplicity we present our analysis on measures of ventricular expansion, computed from surface models of the ventricles in serial MRI. For completeness, we first explain some mathematical concepts from differential geometry – such as medial curves and mappings – that are useful when analyzing patterns of changes on these surface meshes.

Mathematical Preliminaries. Anatomical surfaces in the brain, such as the ventricles, hippocampus, or caudate, have often been analyzed using surface meshes and features derived from them, such as a medial curve, or "skeleton", that threads down the center of a 3D structure [8]. These reference curves are often used to compute the "thickness" of the structure, by assessing the distance from each boundary point to a central line or curve that runs through a structure.

The problem of finding the "medial curve" or "skeleton" of an orientable surface is not well-defined, but a few properties are generally accepted as desirable [100]. Here we focus on those properties that are particularly pertinent for registering and comparing surfaces across multiple subjects:

(1) *Centered*: we would like our curve to be "locally" in the middle of the shape. This is important for accurately estimating local thickness on boundaries of shapes.

(2) *Onto, smooth mapping*: There must exist a surjective, smooth mapping from the surface to the curve. This is essential in order to use the medial curve for registration.

(3) *Consistent geometry*: The geometry of the curve should vary smoothly with smooth variations of the shape.

Exploiting the approximately tubular structure of many subcortical regions of interest (ROI), we make the simplifying assumption that our skeleton is a single open curve with no branches or loops. While this is a strong assumption, it greatly simplifies computation, and allows us to focus on (P1) and (P3). In practice, single curve skeletons are robust, even for representing branching shapes like the ventricles [91]. Focusing on (P1), we say that a curve is the *medial* curve if it is smooth and every point on it is "locally in the middle" of the shape. Formalizing this intuition for approximately tubular shapes, we have the following expression for a medial cost function. Given a surface $\mathcal{M}$, the medial curve $\boldsymbol{c}: [0,1] \to \mathbb{R}^3$, should be a global minimum of

$$R(\boldsymbol{c}, \boldsymbol{c}', \mathcal{M}) = \int_0^1 \int_{\boldsymbol{p} \in \mathcal{M}} w(\boldsymbol{c}(t), \boldsymbol{c}'(t), \boldsymbol{p}, \mathcal{M}) |\boldsymbol{c}(t) - \boldsymbol{p}|^2 \, d\mathcal{M} dt, \tag{2}$$

where $\boldsymbol{c}(0), \boldsymbol{c}(1) \in \mathcal{M}$. Here, $w(\boldsymbol{c}, \boldsymbol{c}', \boldsymbol{p}, \mathcal{M})$ is the weight defining the "localness" of point $\boldsymbol{p}$ relative to $\boldsymbol{c}$. Our weight function is defined as in Chapter 3[91]. Adding a smoothness term penalizing curvature $\kappa_{\boldsymbol{c}}$, we have our final cost function:

$$L(\boldsymbol{c}, \boldsymbol{c}', \mathcal{M}) = R(\boldsymbol{c}, \boldsymbol{c}', \mathcal{M}) + \beta \int_0^1 |\kappa_{\boldsymbol{c}}(t)|^2 \, dt. \quad (3)$$

While (P3) is not formally satisfied, it generally holds in practice due to the regularization and the fact that $w(\boldsymbol{c}, \boldsymbol{c}', \boldsymbol{p}, \mathcal{M})$ is piecewise smooth. We equip the shapes with two scalar functions for registration based on the medial curve, the global orientation function (GOF) G and medial thickness D:

$$G(\boldsymbol{p}) = \arg\min_t \{|\boldsymbol{c}(t) - \boldsymbol{p}|, t\epsilon[0,1]\} \quad (4)$$

$$D(\boldsymbol{p}) = |\boldsymbol{c}[G(\boldsymbol{p})] - \boldsymbol{p}| \quad (5)$$

An example of a medial curve and the corresponding GOF is shown in **Figure 2(a)** and the weighting function is illustrated in **Figure 2(b).** To ensure (P2), we apply constrained Laplacian smoothing to the GOF if there are any local extrema not at curve endpoints. This step generally requires just a few iterations and is needed in only a small proportion of cases. We modify the registration presented in Chapter 3 by adding the longitudinal change term:

$$\Delta D(\boldsymbol{p}) = D_2(\boldsymbol{p}) - D_1(\boldsymbol{p}), \quad (6)$$

We first perform longitudinal registration as in Chapter 3 between each follow-up ventricle model and the corresponding baseline model. Thus, the group-wise registration step of [91] is done on only two shapes at a time. We then register each pair of shapes to a corresponding target pair. We do not use group-wise registration of Chapter 3 during the cross-sectional step to avoid "peaking," or unfairly biasing our n80 estimate by using information from the testing sample during the learning stage. Instead, we modify the GOF to minimize the L^2 difference between the 1D thickness and thickness change maps of the target surface and each new surface, expressed as

$$f_*(t) = \frac{\int_{\{\boldsymbol{p}\in\mathcal{M}\ |G(\boldsymbol{p})=t\}} * (\boldsymbol{p})\, d\mathcal{M}}{\int_{\{\boldsymbol{p}\in\mathcal{M}\ |G(\boldsymbol{p})=t\}} d\mathcal{M}}, \tag{7}$$

where * can correspond to D or $\triangle D$. The 1D registration minimizes $C(r, r') =$

$$\int_0^1 w_D(f_D[t - r(t)] - g_D[t])^2 + w_{\triangle D}(f_{\triangle D}[t - r(t)] - g_{\triangle D}[t])^2 + \sigma^2 r'(t)^2\, dt, \tag{8}$$

$$[t - r(t)] \in [0,1].$$

Here the functions f_* are the feature functions of each subject's surface, and g_* are the corresponding features of the target shape. The 1D displacement field r is restricted by $r: [0, 1] \to (-1,1)$. The GOF is adjusted by $G_{adj} = h^{-1} \circ G, h(t) = t - r(t)$. Surfaces are then registered parametrically on the sphere by simultaneously minimizing the L^2 difference between G_{adj}, D, and $\triangle D$ of the new shape and the target shape. The target shape is excluded from LDA training or testing. In this way, each time point and each subject are treated entirely independently; adding new subjects or time points to the dataset does not affect previous results.

5.2.6 LDA-based Feature Weighting

In designing an imaging biomarker, one generally seeks a balance between the intuitiveness of the biomarker and its power to detect disease or disease progression. A natural choice for ventricular shape-based features is radial expansion. It directly measures anatomical change that correlates with the severity of AD and MCI [8, 74, 101-105]. We use the thickness change defined in (7) as our local measure. Having made this choice, we would now like to find an optimal linear weighting for each vertex on the surface to maximize the effect size of our

combined global measure of change. A linear model may not have the intuitive clarity of a binary weighting (i.e., specifying or masking a restricted region to measure), but its meaning is still sufficiently clear and can be easily visualized. Thus we would like to minimize our sample size estimate (1) as a function of the weights, **w**:

$$n(\mathbf{w}) = C\frac{\frac{1}{N-1}\sum\left(\mathbf{x}_i^{\mathrm{T}}\mathbf{w} - \mathbf{m}^{\mathrm{T}}\mathbf{w}\right)^2}{(\mathbf{m}^{\mathrm{T}}\mathbf{w})^2} = \frac{1}{N-1}C\frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_W\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_B\mathbf{w}}. \quad (9)$$

Here $C = 32\left(z_{1-\alpha/2} + z_{power}\right)^2$, $\mathbf{x}_i$ is the thickness change for the i^{th} subject, **m** is the mean vector, the covariance matrix $\mathbf{S}_W = \sum_{i=1}^{N}(\mathbf{x}_i - \mathbf{m})\,(\mathbf{x}_i - \mathbf{m})^T$, and $\mathbf{S}_B = \mathbf{m}\mathbf{m}^{\mathrm{T}}$. Minimizing (9) is equivalent to maximizing

$$J(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_B\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_W\mathbf{w}}, \quad (10)$$

which is a special case of the LDA cost function, with a maximum given by

$$\mathbf{w} = \mathbf{S}_W^{-1}\mathbf{m}. \quad (11)$$

For our purposes, **m** represents the mean of the diseased group. We denote this by $\mathbf{m} = \mathbf{m}_{AD,MCI}$, where $\mathbf{m}_{AD,MCI}$ stands for the mean expansion vector in the combined MCI and AD group. We make no distinction between these two groups during LDA training. Maximizing (10) directly is generally not stable when $\mathbf{S}_W$ has a high condition number, as is typically the case when the number of features greatly exceeds the number of examples. For the same reason, even if a stable solution is found, it is unlikely to generalize to a new sample. This is indeed observed for our ventricle data: direct unregularized solutions yield 1-year training n80's between 10 and 30 for MCI subjects, but applying the weighting to a new, non-overlapping sample of MCI data

can lead to n80's over 200, comparable to the stat-ROI results. To mitigate this, we use two regularization approaches, one aimed at speed and scalability, and the other at precision and generalizability.

To avoid dealing with dense covariance matrices directly, we apply Principal Components Analysis to our training sample, storing the first *k* principal components (PCs) in the rows of a matrix, **P**, and computing the corresponding *k* eigenvalues λ_j. This is a standard approach when applying LDA to actual two-class problems, as it makes the mixed covariance matrix nearly diagonal. In our case, the covariance in PCA space is exactly diagonal, which reduces (11) to a direct computation:

$$\mathbf{w} = \mathbf{P}^{\mathrm{T}}\boldsymbol{\omega}, \text{ where } \omega_j = [\mathbf{Pm}]_j/\lambda_j. \tag{12}$$

This approach is very fast: one can compute the first *k* eigenvectors and eigenvalues of $\mathbf{S}_W$ without explicitly computing $\mathbf{S}_W$ itself. An alternative to PCA is to incorporate spatial smoothing into (11) as Tikhonov regularization. This approach is not as efficient as PCA, but allows us to incorporate prior knowledge about the spatial distribution of vertex weights into the solution. Thus it has better potential to generalize across samples. The regularized solution then becomes

$$\mathbf{w} = (\mathbf{S}_W^2 + a\mathbf{L}^{\mathrm{T}}\mathbf{L})^{-1}\,\mathbf{S}_W\mathbf{m}. \tag{13}$$

Here *a* is the smoothing weight, and **L** is the Tikhonov matrix. We use the matrix of surface Laplacian weights between vertices of the average shape computed from healthy controls. To avoid "peaking" with respect to the test n80's for control subjects, a different average shape and Laplacian matrix are computed for each fold during cross-validation. To address the potential lack of disease specificity of ventricular expansion and the power analysis of (1), we also

optimize (10) for NC-modified sample size estimates. In this case, the mean estimate is modified to

$$\mathbf{m} = \mathbf{m}_{AD,MCI} - \mathbf{m}_{NC}, \quad (14)$$

where $\mathbf{m}_{NC}$ is the mean expansion among controls.

The order of subjects in each diagnostic group is randomly changed to avoid any confounds such as scanner type, as well as the age or sex of the subject. This step is needed mainly because the standard ADNI subject order corresponds to scanning cites. Where the subjects are scanned is known to correlate with reliability in many morphometric measures. This is only done once before LDA training, with the same order and same subdivision of diagnostic groups used for each method. To validate our data-driven weighting approaches, we create two groups of equal size, with an equal number of MCI, AD and NC subjects in each. Each of these folds is then used to optimize the relevant parameter, i.e., the number of principal components *k*, the smoothing weight *a*, or the parametric *p*-value threshold for stat-ROI. The training fold is again divided into two groups with equal number of AD and MCI subjects in each, to tune the parameters. The best parameter is then used to train a model on the whole fold, and the model is tested on the other fold. We note that for the PCA approach, a different set of principal components is computed for each fold so that the covariance information from the test set is not used. We further stress that group-wise registration, even if it is blind to diagnosis and time, would constitute a circular analysis here, as the covariance structure of the test set would again be used to inform the training even if indirectly. In fact, the method proposed in [91] exploits covariance quite directly. One alternative would be to group-wise register each fold separately. The analysis would even remain objective if we then registered the test fold to a probabilistic atlas created with the

training fold, in which case each fold would have two independent homologies, one for testing, and the other for training. However, we did not pursue such complicated schemes, and simply registered all subjects to a single subject template.

To distinguish between the two novel aspects of the LDA model compared to the stat-ROI – the continuous weighting, and the multivariate analysis compared to the mass-univariate approach – we tested the effect of each separately. In practice, this suggests two additional weightings: a continuous t-statistic weighting, which can be computed directly with no parameter tuning; and a masked version of the LDA. In the latter case, two parameters need to be tuned. In addition to the single parameter already embedded in the model, we also need to find an optimal threshold. For computational speed, we choose to threshold the PCA model. To devise a reasonable set of mask thresholds to test, we compute the cumulative distribution functions (CDFs) of the vertex weights, and space our cutoff values at regular intervals along the y-axis. In other words, each subsequent threshold adds a surface region of a roughly constant area. Further, because the LDA maps are signed, we consider both signed and unsigned masks. For the unsigned case, we use the prior knowledge that ventricles are generally expected to expand, thus considering only positively weighted areas, and weighing the vertices by 1 when the threshold is exceeded. For the signed case, we also assign a value of -1 to vertices whose contraction rate exceeds a threshold. In the latter case, different CDFs and threshold magnitudes are used for the positive and negative regions.

To compute meaningful anatomical summary from the vertex weights of each weighting scheme, we normalized the weights by their 1-norm, which corresponds to averaging over the ROI for the discrete methods, assuming equal area elements for all vertices. This assumption,

however, is only approximately true. These results do not exactly correspond to mean sample sizes reported, since the mean n80 is defined as the average of the two folds' sample size estimates.

5.3 Results

To verify that our measure of annual thickness change has good potential as a biomarker of AD and MCI, we initially performed a group mean comparison of radial thickness change over 1 year using a permutation test as in [8]. This test relies on the standard *t* statistic at each vertex, and computes a non-parametric null distribution for the surface area that exceeds the given *t*-threshold. A threshold corresponding to $\alpha < 0.05$ was used as in [8]. We compared AD vs. NC groups, and MCI vs. NC. After 100000 random re-assignments of the group data, permutation-based *p*-values for the overall pattern of group difference for AD vs. NC and MCI vs. NC were below the threshold for each hemisphere, i.e., $p < 10^{-5}$. Localized *p*-maps of the results are shown in **Figure 3**, and are consistent with prior papers by Chou and others [52, 104].

Below we compare the performance of our PCA-based vertex weighting, the Tikhonov-regularized weighting, and the standard stat-ROI approach, as well as t-statistic weighting and signed and unsigned LDA-ROI weightings. In testing each of these weighting methods, we used nested 2-fold cross-validation. Only AD and MCI subjects were used in the training stage. Further, we restricted our training sample to include only 1-year changes. Twenty-four month data was only used for testing, applying 1-year models to the non-overlapping subgroups of the 24-month data.

Our 1-year sample size estimates based on cross-validation were nearly identical for the PCA and the Tikhonov approaches. During training of the PCA model, the optimal number of principal components was chosen to be 28, and 47 for folds 1 and 2, respectively. Maps of the weights averaged over the two folds are shown in **Figure 4a**. The Tikhonov approach resulted in predictably smoother weight maps, the mean of which is shown in **Figure 4b**. Twenty exponentially increasing values of smoothing weight a were tested, between $a = 10^{-2}$ and $a = 10^{7}$. The two folds returned $10^{3.5}$ and 10^{4} as the optimal values.

The "stat-ROI" approach led to inferior results with n80's notably higher for all three diagnostic groups, especially MCI. The optimal t-threshold was chosen to be $p = 10^{-6}$ in both folds. The range of tested p-values contained every power of 10 between $p = 10^{-4}$ and $p = 10^{-20}$, which corresponded to ROIs of every size, from patches covering nearly the entire surface (>95%) to just a handful of mesh vertices. The average stat-ROI mask is shown in **Figure 5c.** As an additional control, we computed sample size estimates based on ventricular volume change in both hemispheres. Volume-based estimates for the n80 were significantly higher than the three surface-based measures. **Table 2** shows a summary of all 1 year sample size estimates. **Figure 6** illustrates the mean sample size estimates for all four measures at 1 year. Bootstrapped 95% confidence intervals (CIs) for each fold were computed as in [74]. To estimate confidence intervals for the whole cohort in an unbiased way, we normalized the linear weights of each fold by their standard deviation. Although the scale of the weighting vector has no bearing on (1) within each fold, the relative scales of the two vectors can skew the CIs significantly when considered together. Thus, a similar scaling is necessary when computing overall CI's. After this step, the overall CIs were computed the same way as for each fold. We also computed bootstrapped mean n80 comparisons between the different methods for each fold, and for the

overall sample. While the two LDA approaches were not significantly different, both led to significantly lower n80's for an MCI clinical trial. The PCA method showed significantly lower estimates for NC subjects, if not corrected for multiple comparisons. P-values for n80 comparisons across the first three methods are summarized in **Table 3a**.

Table 2. Sample size estimates for clinical trials, using ventricular change over 12 months as an outcome measure. Depending on how we weight the features on the ventricular surfaces, the sample size estimates can be reduced, and the power of the study increased. The two LDA-based methods (*top two rows*) show lower sample size estimates (i.e., greater effect sizes) than the standard "statistical ROI" approach, which uses a binary mask to select a region of interest. The "t-stat" row shows results when weighting the vertex expansion rates with the t-statistic. "PCA sign." and "PCA uns." show results when thresholding PCA-LDA based weight maps, with "sign." meaning that negatively weighted areas were considered and assigned a weight of -1 when below the threshold. "Uns." means "unsigned," i.e. only positively weighed vertices were considered, and weighed with 1 if exceeding the threshold. All surface-based approaches (*top six rows*) outperform measures of change based on ventricular volume. The "mean" columns display n80's and CI's of the two folds' estimates averaged.

	MCI	AD	NC	Mean MCI	Mean AD	Mean NC
PCA	111/96 (85,150)/(75,127)	65/86 (46,92)/(64,128)	134/192 (106,177)/(150,260)	**104** **(94,139)**	**75** **(64,102)**	**163** **(114,190)**
Tik.	116/105 (92,154)/(81,146)	71/95 (49,100)/(65,155)	155/186 (121,205)/(156,247)	110 (92,135)	83 (63,110)	170 (119,196)
Stat-ROI	184/145 (143,256)/(108,215)	95/94 (64,143)/(67,143)	207/201 (159,279)/(155,271)	165 (134,209)	94 (72,125)	204 (156,273)
t-stat	205/151 (154,289)/(112,247)	91/99 (63,143)/(67,147)	218/212 (166,298)/(162,288)	178 (143,232)	95 (72,128)	215 (175,264)
PCA sign.	134/111 (104,184)/(83,157)	85/81 (61,124)/(55,128)	170/187 (128,242)/(144,251)	123 (100,154)	83 (63,111)	178 (146,222)
PCA uns.	226/161 (166,332)/(120,263)	95/100 (65,146)/(68,150)	242/233 (182,339)/(176,323)	193 (155,261)	98 (74,130)	237 (192,296)
Vol.	-	-	-	266 (216,355)	145 (108,199)	352 (262,533)

The sample sizes based on t-statistic weighting were very similar to stat-ROI results, with no significant difference, though stat-ROI n80's were generally slightly lower. These weights are visualized in **Figure 4c**. The LDA-ROI sample sizes were greater than the continuous weighting, though the difference only reached significance for the unsigned case. For the signed case, 81 and 71 PCs were selected during parameter tuning for the two folds, and the threshold

corresponded to 80% of the vertices. For the unsigned case, only 15 and 7 PCs were used, with the threshold set at 90% of the vertices. Comparisons of continuous vs. discrete weightings are presented in **Table 3b**. The masks resulting from the three discrete methods are visualized in **Figure 5**. P-values for n80 comparisons between the discrete and the corresponding continuous weightings are summarized in **Table 3b**.

Finally, the control-augmented sample sizes resulting from the 2-class LDA model resulted in noticeably different maps compared to the 1-class models. We again used the PCA approach, with 32 and 54 principal components used in the two folds. The pattern of the 2-class LDA model was characteristic of AD: significantly more weight was given to the inferior horns bordering the hippocampus, and more weighting was also given to the middle of the occipital horns, characteristic of white matter degeneration. These results are displayed in **Figure 4d**. For comparison purposes, we show the control-adjusted sample size estimates for our weightings in **Table 4b** and **Figure 7b**, and also those reported in [73], in **Figure 9b**.

All methods had good agreement between the two folds' models. The sample sizes in each fold were similar, and the weight patterns were also in good agreement. We used ordinary linear regression for each weighting scheme's pair of linear models. The null hypothesis that the regression model does not fit the data (F-test) returned $p < 10^{-20}$ for all weighting schemes. **Tables 4** and **5** show corresponding results for 24-month sample sizes. Here, the general trend is similar to 1-year, though the sample size estimates in controls are closer to estimates for MCI and AD. **Figure 7a** illustrates this effect graphically.

Table 3a. *P*-values estimating the evidence that the true 12-month n80 (sample size requirement) of the first method is equal to or greater than that of the second method. Null distributions were created by bootstrapping 100000 samples with replacement. Note that depending on how rigorous one is about hypothesis testing, the true *p*-values may need a Bonferroni correction by a factor of 3, if one accepts a separate correction for each subset of the data, or 9 (in which case the two *p*-values lower than 0.05/9, here, should be considered significant). In either case, n80's for NC subjects are not significantly different among the surface-based methods. At 12 months, the two LDA-based methods give statistically indistinguishable results.

PCA vs. Tikhonov			PCA vs. stat-ROI			Tikhonov vs. stat-ROI		
MCI	AD	NC	MCI	AD	NC	MCI	AD	NC
0.545	0.437	0.419	**0.00475**	0.200	**0.0366**	**0.0035**	0.256	0.0529

Table 3b. *P*-values estimating the evidence that the true 12-month n80 (sample size requirement) of the first method is equal to or greater than that of the second method. For LDA, continuous weighting gives better results, but the difference is only significant when using unsigned masking. For stat-ROI, masking is better, but the improvement is not significant.

PCA vs. signed LDA-ROI			PCA vs. unsigned LDA-ROI			Stat-ROI vs. t-stat weighting		
MCI	AD	NC	MCI	AD	NC	MCI	AD	NC
0.27697	0.43547	0.27105	**0.00037**	0.15134	**0.00421**	0.30627	0.48339	0.35982

Table 4a. Sample size estimates for clinical trials, using ventricular change over 24 months as an outcome measure. Depending on how the features on the ventricular surfaces are weighted, the sample size estimates can be reduced, and the power of the study increased. The two LDA-based methods (*top two rows*) show lower sample size estimates than the stat-ROI approach. Control subjects' average atrophy now approaches that of MCI subjects in magnitude.

	MCI	AD	NC	Mean MCI	Mean AD	Mean NC
PCA	80/62 (65,108)/(44,86)	67/47 (47,122)/(31,67)	69/74 (54,91)/(57,102)	71 (65,98)	57 (45,89)	72 (52,89)
Tik.	73/60 (57,99)/(42,87)	63/41 (44,117)/(26,63)	60/80 (47,82)/(62,105)	**67** **(54,84)**	**52** **(38,76)**	**70** **(48,82)**
Stat-ROI	141/96 (113,193)/(69,130)	93/66 (69,150)/(43,98)	119/122 (89,177)/(92,182)	119 (98,149)	80 (61,108)	121 (90,179)
Vol.	-	-	-	191 (157,258)	119 (88,169)	196 (155,253)

Table 4b. Sample size estimates for clinical trials, using ventricular change over 24 months as an outcome measure, modified by change in controls. The NC-modified analogues to table 4a show a marked increase in required sample size. The 2-class model greatly outperforms all other ventricular measures, with ventricular volume performing on par with the best 1-class surface measures: unsigned LDA-ROI and Stat-ROI.

	MCI	AD	Mean MCI	Mean AD
PCA	1170/1373 (659,2758)/(689,4240)	426/276 (240,1157)/(165,576)	1272 (760,2239)	351 (222,610)
Tik.	1661/1595 (859,4892)/(737,6306)	534/302 (286,1743)/(167,747)	1628 (902,3440)	418 (261,875)
Stat-ROI	852/1171 (529,1610)/(623,3171)	266/241 (169,509)/(150,449)	1011 (660,1704)	254 (176,379)
t-stat	840/1298 (517,1628)/(670,3708)	238/249 (152,433)/(157,459)	1069 (697,1880)	244 (171,355)
PCA sign.	1480/1380 (794,3803)/(681,4488)	482/244 (257,1309)/(140,527)	1430 (835,2674)	363 (229,676)
PCA uns.	841/1133 (514,1650)/(611,2958)	238/233 (150,444)/(147,423)	987 (641,1696)	235 (167,357)
2-class	**461/628** **(290,841)/(380,1244)**	**115/174** **(76,191)/(112,306)**	**544** **(371,802)**	**145** **(101,192)**
Vol.	--	--	916 (658,1385)	253 (182,372)

Table 5. *P*-values estimating the chance that the true 24 month n80 of the first method is equal to or greater than that of the second method. Null distributions were created by bootstrapping 100000 samples with replacement. Note that depending on how rigorous one is about hypothesis testing, the true *p*-values may need a Bonferroni correction by a factor of 3, if one accepts a separate correction for each subset of the data, or 9. The improvement of the Tikhonov-LDA method over the stat-ROI approach reaches significance, when uncorrected, for AD subjects. At 24 months, the improvement in power when using Tikhonov-regularized LDA model over the PCA model approaches trend levels for MCI subjects.

Tikhonov vs. PCA			PCA vs. stat-ROI			Tikhonov vs. stat-ROI		
MCI	AD	NC	MCI	AD	NC	MCI	AD	NC
0.129	0.257	0.334	**0.00249**	0.11	**0.00278**	**0.0001**	**0.0276**	**0.00076**

To assess whether there is any evidence of longitudinal bias of our weighted measures, we applied our 1 year models to healthy controls at 12 and 24 months. Using a method similar to [89], we used the *y*-intercept of the linear regression as a measure of bias (bearing in mind the caveats noted that there may be some biological acceleration or deceleration that could appear to be a bias). We again used bootstrapping to estimate the intercept and linear fit confidence intervals [106]. **Figure 8** shows the regression plots for all surface models over the two follow-up time points. Confidence intervals for the linear fits are shown in dotted green lines. The units of change along the *y*-axis represent the weighted ventricular expansion, normalized by the 1-norm of the weight vectors. The 95% confidence interval for the PCA method was (-0.0218, 0.036) mm, with a mean expansion of 0.111 mm at 1 year and 0.214 mm at 2 years. For the Tikhonov model, the 95% CI was (-0.0126, 0.0242), with mean change at 1 and 2 years of 0.0704 mm and 0.1351 mm, respectively. The stat-ROI summary resulted in a 95% CI of (-0.0411, 0.0645) mm, mean expansion 0.158 mm at 1 year, and 0.306 mm at two years. The bias test results are summarized in **Table 6**. Group averages for atrophy rates, with each model, are reported in **Tables 7a** and **b**.

Table 6. Longitudinal bias analysis of ventricular surface-based measures**.** Change in healthy controls is linearly regressed over two time points. The intercept is very close to zero, with the confidence interval clearly containing zero for each method. The surface-based measures do not show any algorithmic bias according to the CI test.

PCA	Tikhonov	T-stat	2-class LDA	Signed LDA-ROI	Unsigned LDA-ROI	Stat-ROI
0.0064 (-0.0218, 0.06)	0.0048 (-0.0126, 0.0242)	-0.0102 (-0.0617, 0.0416)	0.0172 (-0.0031, 0.036)	-0.0031 (-0.0208, 0.0143)	-0.0189 (-0.0665, 0.0304)	0.0115 (-0.0411, 0.0645)

Table 7a. Ventricular surface summary atrophy measures for continuous weightings**.** Averages and standard deviations for atrophy rates are in millimeters of radial expansion. The vertex weights are normalized by their 1-norm, which corresponds to averaging over the ROI for the stat-ROI method, assuming equal area elements for all vertices.

	Tikhonov			**PCA**			**t-stat**			**2-class LDA**		
	MCI	**AD**	**NC**	**MCI**	**AD**	**NC**	**MCI**	**AD**	**NC**	**MCI**	**AD**	**NC**
12 mo	0.0872 ±*0.060*	0.105 ±*0.062*	0.0704 ±*0.054*	0.151 ±*0.10*	0.191 ±*0.11*	0.111 ±*0.084*	0.247 ±*0.21*	0.349 ±*0.21*	0.146 ±*0.14*	0.058 ±*0.06*	0.0835 ±*0.073*	0.0299 ±*0.05*
24 mo	0.146 ±*0.079*	0.181 ±*0.086*	0.135 ±*0.067*	0.257 ±*0.14*	0.33 ±*0.17*	0.214 ±*0.11*	0.46 ±0.32	0.68 ±0.37	0.304 ±*0.21*	0.101 ±*0.1*	0.175 ±*0.13*	0.0412 ±*0.08*

Table 7b. Ventricular surface summary atrophy measures for discrete weightings. Discrete analogues of results in table 7a, because negative weights are allowed in signed LDA-ROI, the normalized average expansion is closer to the 2-class LDA result, while the unsigned version is closer to stat-ROI and t-statistic weighting.

	stat-ROI			**Signed LDA-ROI**			**Unsigned LDA-ROI**		
	MCI	**AD**	**NC**	**MCI**	**AD**	**NC**	**MCI**	**AD**	**NC**
12 months	0.262 ±*0.21*	0.367 ±*0.22*	0.158 ±*0.14*	0.0602 ±*0.042*	0.0796 ±*0.046*	0.0411 ±*0.035*	0.226 ±*0.2*	0.328 ±*0.2*	0.132 ±*0.13*
24 months	0.468 ±*0.32*	0.690 ±*0.39*	0.306 ±*0.21*	0.108 ±*0.06*	0.14 ±*0.068*	0.083 ±*0.046*	0.438 ±*0.3*	0.652 ±*0.36*	0.284 ±*0.19*

5.4 Discussion

Here we introduced and tested an approach to increase the efficiency of clinical trials in Alzheimer's Disease and MCI, based on multiple neuroimaging features, with a straightforward application of Linear Discriminant Analysis (LDA). We applied our measure of brain change to a surface-based measure of atrophy in the lateral ventricles. Despite the simplicity of our approach, the resulting sample size estimates are significantly better than the stat-ROI approach, which has been the standard feature weighting method to date. The linear feature weighting also produces an intuitive, univariate measure of change – a single number summary that can be

correlated to other relevant variables and outcome measures. The linear weights can be easily visualized, adding insight into the pattern and 3D profile of disease progression. Our longitudinal ventricular morphometry showed high sensitivity to local differences in shape change due to AD. Local maps of shape change were consistent with previous studies.

We applied our LDA approach to local ventricular shape change features, with promising results. We used two alternative methods for solving the LDA optimization problem. The first approach, based on principal components analysis, is very fast and scalable to larger feature sets, such as dense Jacobian determinant maps in volumetric Tensor Based Morphometry (TBM). The other optimization method exploits the relatively sparse nature of our surface data by adding Tikhonov regularization in the form of surface-based scalar Laplacian smoothing.

To distinguish between the two novel aspects of the LDA approach with respect to the stat-ROI – multivariate analysis and continuous weighting – we compared three additional weighting schemes. The first is simply the continuous version of the mass-univariate approach used in the stat-ROI, the t-statistic weighting. In cross-validation, this weighting performed slightly worse than the stat-ROI, but the difference did not reach significance. The second and third weighting schemes were designed to be discrete analogues of the PCA-LDA model. These performed worse than the continuous PCA model, with the difference between the unsigned LDA-ROI and the PCA models reaching significance for MCI and NC subjects. However, the signed LDA-ROI model performed notably better than the stat-ROI. These results together suggest that both the continuous and the multivariate aspects of the LDA models contribute to sample size reduction, but the multivariate aspect may play a larger role.

5.4.1 Shape Analysis

We have modified a shape registration approach for longitudinal shape analysis in the lateral ventricles. A variety of "medial-curve type" analysis methods for subcortical shapes have been developed over the years. A discrete approach called *M*-reps was popularized by Pizer et al. [51] and extended to the continuous setting by Yushkevich et al. [31]. *M*-reps consist of a discrete web of "atoms," each of which describes the position, width and local directions to the boundary, and an object angle between corresponding boundary points. The approach leads to an extremely compact representation of the shape model. However, the method requires a specific *m*-rep model for each type of shape. For a given brain region, this model may need to be modified before it can be applied to a different dataset, if the geometry of the new set of shapes is slightly different, e.g. after being segmented using a different protocol. This drawback is partially overcome when the medial core is continuous.

"*CM*-reps" are an elegant extension of M-reps to the 2-D continuous medial core. CM-reps offer a way to derive boundaries from skeletons, by solving a Poisson-type partial differential equation with a nonlinear boundary condition [31]. The resulting 2D medial "sheet" continuously parameterizes the shape-enclosed volumetric region, as well as the surface. Thus in spirit it is very similar to our approach: a particular topology of the continuous medial model is assumed, and the model is deformed to fit each shape. However in certain practical applications, the 2D aspect of the *cm*-reps model can become a liability, leading to inconsistent parameterization for a family of similar shapes. One such case is the lateral ventricle, where the 2D medial sheet can twist unpredictably around the junction of the superior and occipital horns. Instead, our general approach is to compute a single 1D continuous curve skeleton and use the curve to induce feature functions on the surface. Shape registration is then performed

parametrically by minimizing the L^2 difference between corresponding feature functions of a pair of shapes. Wang et al. [52] used radial distance in conjunction with a conformal parameterization and surface-based TBM as an improved measure of ventricular expansion. In this case, a single-curve skeleton was computed based on the conformal parameterization for each ventricular horn. Curve-skeletons of fixed topology have also been used before in medical imaging with distance fields (DFs) [107]. Our approach avoids the use of DFs, and defines a cost function relating the skeleton directly to the surface, eliminating the imprecision associated with the additional discretization due to DFs. Further, relying only on the discretization of the surface allows us to greatly speed up computation, making analysis of many hundreds of 3D shapes with the continuous medial axis achievable in little time without using large computing clusters.

5.4.2 Machine Learning in Shape Analysis and Alzheimer's Disease

PCA-type approaches have been used in prior shape analyses. A regularized components analysis approach called LoCA [108] is similar to sparse PCA [109]. The idea, similar to PCA, is to generate an orthogonal basis for shape space, while adding a penalty term. However, instead of penalizing the number of non-zero weights in each basis vector, as is done in sparse PCA, LoCA instead forces all the non-zero components to be spatially clustered on a surface, which gives each component a clearer anatomical meaning. However, both of these methods come at a much higher computational cost than ordinary PCA: they are iterative, while PCA only relies on eigen-decomposition, making it feasible for much larger feature sets. LoCA has been applied to shape analysis in AD, including ADNI hippocampal data [110], finding specific component associations with AD and other biomarkers. Here, the measure used was very similar to ours, the

radial distance. Only baseline data were used in assessing morphometric differences, while our current study is longitudinal.

Sparse basis decomposition has also been used as a preprocessing step for training an AD classifier, for example by applying Independent Component Analysis (ICA) to gray matter density maps before using machine learning methods, such as Support Vector Machines (SVM), for classification [111]. Many machine learning approaches have been applied to image-based diagnosis or classification of AD and MCI. Davatzikos and colleagues applied SVM to RAVENS maps [112], an approach similar to modified VBM [113] which assigns relative tissue composition to every voxel after a high-dimensional warp. A similar approach was used by Vemuri [40], using tissue probability maps (TPMs) - essentially the same VBM measures that are constructed using the SPM package. Kloppel et al. [42] further showed that such a model can be stable across different datasets. In general, classification algorithms can achieve AD-NC cross-validation accuracy in the mid-nineties (~95%) within the same dataset, although performance inevitably degrades when applied to new datasets, especially if the cohort demographics or scanning protocols are different.

Cuingnet et al. [114] developed a Laplacian-regularized SVM approach for classifying AD and NC subjects, which is very similar in spirit to our Tikhonov-regularized LDA. They show that using the Laplacian regularizer improves classification rates for AD vs. NC subjects. SVM has also been used, in our prior work, to separate AD and NC subjects based on hippocampal shape invariants and spherical harmonics [115]. Another recent surface-based classification effort by Cho et al. [116] uses an approach very similar to our PCA method, where surface atlas-registered cortical thickness data is smoothed with a low-pass filter of the Laplace-

Beltrami operator, computed on the atlas shape. Following this procedure, PCA is performed on the smoothed surface thickness data and LDA is performed on a subset of the PCA coefficients to train a linear classifier. The resulting classification accuracy is very competitive. Another surface-based classifier [117] uses the SPHARM-PDM approach to classify AD and NC subjects based on hippocampal shape. SPHARM-PDM [118] computes a small number of spherical harmonic coefficients based on an area-preserving surface map, and normalizes the spherical correspondence by aligning the first-order ellipsoid with the poles. The result is a rudimentary surface registration and a spectral decomposition of the shape. Gerardin et al. reported competitive classification rates compared to whole-brain approaches. Shen et al. [119] recently used a Bayesian feature selection approach and classification on cortical thickness data and showed that AD-NC and MCI-NC classification accuracy remains competitive with SVM. Finally, to combine multiple modalities for classification, Zhang et al. [120] developed a multiple kernel SVM classifier to further improve diagnostic AD and MCI classification.

It is important to stress that while many studies have used machine learning to derive a single measure of "AD-like" morphometry for discriminating AD and MCI subjects from the healthy group, no study we are aware of has used machine learning to maximize the power of *absolute atrophy rates* in AD. We have attempted this by using a straightforward application of LDA, using either PCA or Tikhonov regularization. The Tikhonov approach was intended to improve generalization relative to PCA, but surprisingly, there were essentially no major differences in test sample size estimates for AD and MCI subjects between the two methods. Only one subgroup at 24 months approached a trend level for a difference in efficiency (sample size difference) between the Tikhonov and PCA model. The Tikhonov method was slightly better for reducing sample size estimates in MCI, generalizing better - as expected. A potential

cause of this may be an insufficiently thorough optimal parameter search for the Tikhonov approach, as the search must remain fairly coarse due to computational constraints. On the other hand, it is possible that the covariance structure of the training samples captures the spatial priors sufficiently well, and an explicit prior does not significantly improve generalizability.

Outside of Alzheimer's literature we found one approach for explicitly minimizing sample size estimates [121], and another that uses SVM for classification of Huntington's disease patients versus controls, with reduced sample sizes as a by-product [88]. The first paper is methodologically closest in spirit to this work: a fidelity term is explicitly defined to be the control-adjusted sample size estimate. A number of non-linear constraints are then added: the total variation norm (TV1-norm), sparsity and non-negativity. While the first two have analogues that can be linearly optimized as we do here (TV2 and L^2 norm), the third constraint forces the authors to use non-linear conjugate gradient (CG), which leads to far slower convergence than the linear CG we use. More importantly, due to the differences in the nature of their data – knee cartilage CT images – and ours, the sparsity and non-negativity constraints are perhaps not appropriate for brain imaging. We expect the effect over soft tissue to be diffuse without many discontinuities, and non-negativity is generally not appropriate in brain MR either. Admittedly, though, as we have focused only on the ventricles, non-negativity would probably be appropriate here, though it would lead to slower convergence. The second paper [88], which we mention in the introduction, simply uses leave-one-out linear SVM weighting of fluid registration-based TBM maps to derive an atrophy measure. No spatial regularization, or sample size-specific modification to the learning approach is used. In both of these cases the measure used is based on the difference between the mean of controls and the diseased group, which is not the main goal of the present work. Though we have used the NC-adjusted measure here as well to show the

potential to reveal AD-specific ventricular change patterns, the main goal was to optimize detection of absolute change.

5.4.3 Other Ventricular Measures

Several studies have used a ventricular measure alone, as a predictor of cognitive decline. Ott et al. [101] showed that ventricular volume is associated with CSF Aβ. Carmichael [104] compared ventricular volume and ventricle-brain ratio (VBR) across MCI converters and non-converters with significantly increased volume, and VBR at baseline among converters. Nestor et al. [102] used a semi-automated highly precise ventricular segmentation to estimate differences in rates of volumetric change. Rates of volumetric increase in the ventricles were significantly greater in MCI and AD subjects compared to NC, which is in line with expectations about the rates of atrophy in each group. Chou et al. [103] performed a cross-sectional ventricular study on the baseline ADNI dataset, using a surface-based model. Surfaces were registered – in a similar way to [8] - by separating each horn and computing three separate medial axes. The radial distance measure was shown to be significantly different between NC and AD, and NC and MCI subjects. Several other cognitive measures and CSF biomarkers were shown to correlate significantly with the local ventricular surface expansion, in a direction expected from the advancing pathology and the intensification of the disease. Ferrarini et al. [122] showed differences in local ventricular surface morphometry between AD and NC subjects using permutations testing and a novel algorithm – known as "GAMEs" - for surface meshing and matching.

5.4.4 Power Estimates of Other Measures in AD

Our ventricular change measures outperformed other common ventricular measures as an AD biomarker with respect to the sample size requirements, assuming of course that the reference data are comparable. Schott [74] reported 1-year n80's for the "Ventricular Boundary Shift Integral" (VBSI) of 118 (92, 157) for AD and 234 (191,295) for MCI at 1 year. Holland et al. [73] reported a Quarc ventricle measure of 92 (69, 135) for AD and 183 (146, 241) for MCI for a 2-year trial. FreeSurfer ventricular measures give similar 2-year estimates of 90 (68,128) for AD and 164 (133, 211) for MCI. Our approach performs comparably well or better than many other imaging measures, in particular those using the entire cortex. An FSL tool, known as SIENA [78, 123], achieved a 1-year point estimate for sample size of 132 for AD and 278 for MCI. Quarc achieved 2-year whole brain estimates of 84 (63, 123) for AD and 149 (121, 193) for MCI. FreeSurfer is reported [73] to achieve 2-year whole brain estimates of 252 (175, 408) for AD and 384 (294, 531) for MCI. Schott reported that BBSI, a whole brain gray matter atrophy measure [74], required 1-year samples of 81 (64, 109) for AD and 149 (122, 188) for MCI. Hua et al. [71] used improved Tensor Based Morphometry (TBM) with the stat-ROI voxel weighting to achieve 1-year sample sizes of 58 (45,81) for AD and 124 (98,160) for MCI. These comparisons are summarized in **Figures 9 and 10.** Though all sample size estimates mentioned here are based on the same ADNI-1 dataset, different studies were done on ADNI subsamples of different size. To shed some light on this, we give the number of subjects used for each study in **Table 8.** Comparison between our study and others is more meaningful where there are fewer exclusions. We note that while the Quarc and Freesurfer results from [73] are for a 2-year study, and using a slightly different power calculation, in fact even subjects who only had scans up to 1 year were considered. The variance and disease effect in these calculations were based on a

mixed effect model using all available time points for each subject. However, there are still roughly 10% fewer subjects used for Quarc and 7% fewer for Freesurfer compared to the full ADNI dataset.

Table 8. Number of ADNI subjects used for computing sample size estimates in Figures 9 and 10.

	AD	MCI
Quarc	131	311
FreeSurfer	135	320
BSI	144	334
SIENA	85	195
TBM Stat-ROI	138	326
Medial Vent-LDA	144	337

5.4.5 Algorithmic Bias

Importantly, we showed that our surface-based measures are longitudinally unbiased according to the intercept CI test [124], alleviating common concerns about overly optimistic sample size estimates due to, for example, additive algorithmic bias. The fact that the baseline and follow-up scans were processed identically, and independently, avoids several sources of subtle bias in longitudinal image processing that can arise from not handling the images in a uniform way [90]. Some issues have been raised regarding the validity of the intercept CI test as a test for bias in estimating rates of change. The CI test assumes that the true morphometric change from baseline increases in magnitude linearly over time in healthy controls. Relying on this assumption, the test examines whether the intercept of the linear model, fitted through measures of change at successive time intervals in controls, is zero. If this is not the case, the measure of change is said to have additive bias. There are two common criticisms of this test. First, the linearity assumption may not always be valid, i.e., true biological changes may be nonlinear. In this case, a truly unbiased algorithm could fail the test, while giving accurate

results. For example, the loss of tissue may be proportional to the amount of tissue left, so the change in volume as measured by TBM, or medial distance to a ventricular boundary relative to baseline might decay or expand exponentially. Alternatively, if loss of tissue volume is linear in time, radial distance measures might be expected to change in proportion to the cube-root of time, or to vary according to some empirical power law relating the distance to volumetric measures, depending on which directional changes contribute most to the overall change [125, 126]. This may partially explain the slight additive "bias" that is detectable in AD and MCI subjects (though not in controls). In disease, the power law describing changes as a function of time may be different compared to controls due to disease effects. As a result, only control subjects should be used when using the linear fit CI test, but even in that case, it is not a perfect test, in that an accurate algorithm could fail the test.

The second common criticism of the CI bias test is that non-additive, atrophy-dependent type of bias may not be detected with the test. In other words, if there exists a complex transfer function between the true change and measured change, a simple linear regression through several time points may not reveal this as a non-zero y-intercept. However, care must be taken when such a transfer function is discovered in an algorithm. It is not simply enough to show the existence of such a function to call an algorithm "biased"; one must also show that the function systematically scales the square of the mean and the variance of the whole sample differently over any number of time points. To address this, we will break the possible transfer functions into four somewhat-overlapping categories, ordered by plausibility, and discuss each. We assume that in all cases an observed transfer function for the mean change is similar in form to the true transfer function for individual subjects.

1. *No additive bias and a transfer function whose second derivative has the same sign everywhere.* This is perhaps the most likely scenario. For example, measures derived from non-linear deformation fields such as TBM are known to undergo this kind of bias when inverse-consistency is not enforced; page 5 of [127] gives a clear explanation for this. In this case, assuming a linear true change, the measured change will follow a non-decreasing (or non-increasing) curve with the same convexity everywhere. Since we assume a zero y-intercept for the curve, a linear fit through two or more time points will inevitably show a non-zero intercept, given enough true change. Thus, the CI intercept test can detect this kind of bias even with just two time points in addition to baseline. Most "known" cases of algorithmic bias that the CI test has so far revealed in literature [73, 124] may well fall in this category.
2. *An additive bias and a transfer function whose second derivative has the same sign everywhere.* In this case, the additive bias may indeed cancel out the effect of the atrophy-dependent bias, if the second derivative and the additive bias have the same sign. However, this would require quite a special set of circumstances: the sampled time-points and the additive portion of the bias would have to be just right for the linear fit to have a zero y-intercept. With sufficiently many time points, this situation becomes virtually impossible. However, as we have only used three time points, we must admit that this is a possibility, however unlikely, in our study.
3. *A strictly linear transfer function with or without additive bias.* This scenario is the clearest example of when the CI test would fail. In this case, the measured

mean change would be a constant multiple of true mean change. However, it is unclear whether this kind of transfer function can be properly called a "bias." If such a transfer function had similar form for individual subjects as their sample mean, the sample size estimates would remain unchanged, as the variance would scale in the same way as the square of the mean. On the other hand, it is somewhat difficult to imagine a linear bias that scales the variance and the square of the mean differently for many time points. This would violate our assumption.

4. *"Other:" A transfer function with one or more inflection points, with or without additive bias.* Though experience with image processing algorithms shows that this scenario is highly improbable, in the interest of completeness we note that this kind of bias can indeed be missed by the CI test, while unfairly inflating power estimates. This is also the only scenario in which our assumption regarding individual bias and bias of the mean will be invalid.

Having gone through this exercise, we hope to have shown that while certain kinds of non-additive bias can potentially be left undetected by the intercept CI test, such a bias is either improbable or does not lead to unfair sample size estimation. Far more likely is the breakdown in the linearity assumption of the CI test, which renders the test inappropriate. Using a subset of the data in which the change is small enough to be at least approximately linear in time, i.e. the control subjects, alleviates the latter issue.

We deliberately chose to use only 12-month data in training our optimized atrophy models. This choice was motivated by practical concerns: we wanted to show that using LDA for optimizing power generalizes sufficiently well to later time points without requiring real drug

trials to recompute the atrophy measure for all scans with each new batch of data. This latter situation could be considered a "moving target." While it may be argued that a better, more parsimonious model could be trained on a sample that includes all available time points, perhaps even incorporating the time variable explicitly, such an approach potentially hides any methodological bias that is inherent in a particular algorithm. Further, a real clinical trial would likely require an ability to apply the same measure to new follow-up scans without requiring one to recompute the measure for all time points every time a new batch of follow-up scans becomes available. In this way, we avoid the moving target problem that bedevils methods requiring scans from all time points. Further, because such methods rely on data-driven techniques to reduce longitudinal bias, for example by computing a per-subject brain atlas based on all available time points, the perceived absence of bias does not indicate that the imaging algorithms themselves are unbiased. This may be called a kind of "peaking" or circularity, as the data used to assess the bias is also used to compute the atrophy measure. However, we do not rely on any data-driven techniques which explicitly force the results to be transitive over time for all available scans. In this sense, the absence of bias in our measures is more indicative of algorithmic objectivity with respect to the amount and direction of morphometric change.

5.4.6 Total and Relative Atrophy

There has been some recent debate regarding the need to subtract the mean of the healthy controls when estimating sample sizes for a drug trial. Some ADNI collaborators seem to have rejected this idea [17, 71], in part because real drug trials do not tend to enroll controls, and even if they did, many controls already harbor incipient Alzheimer pathology or some degree of

vascular pathology that may also be resisted by treatment. However, the idea is not completely without merit, since all meaningful trials must compare a treatment against another (placebo or established) treatment group. Further, any additive algorithmic bias could be excluded by subtracting the mean rate of controls. In this paper, we addressed this issue by computing an additional linear ventricular expansion model, specific to AD and MCI progression. We did this by directly applying a 2-class, as opposed to 1-class, LDA with the covariance defined strictly by the diseased group, as required by the current practice of NC-adjusted sample size estimates. The resulting power estimates for NC-adjusted atrophy outperform all previous ventricular measures.

5.4.7 Future Work

Future work will include combination of multiple biomarkers, including other imaging biomarkers, such as TBM or even non-imaging biomarkers (such as CSF or proteomic measures), into the framework. For example, there is no mathematical reason why only the ventricular measures should be included, as any pertinent structural or functional measure could be incorporated. In the former case, only the PCA-based approach would be feasible due to the high computational cost. It is also important to validate the LDA-weighted measures as well as stat-ROI measures across different data sets. Further, we would like to extend the use of supervised learning to further reduce our sample size estimates. For example, in the PCA experiment, we simply used all principal components - up to a cutoff value. Although the resulting patterns lead to low sample sizes, their spatial distribution contained high-frequency components that may not have anatomical meaning. A more advanced approach would be to perform a greedy boosting-type search over the principal components, with the goal of making

the pattern more generalizable and more congruent across the folds. Such a blended LDA-boost approach with built-in cross validation for better fold agreement was proposed in [128]. As our linear weighting is likely to contain a combination of disease effect and systematic registration artifact, a boosting approach over the principle components could potentially isolate and discount any PCs containing the artifactual portion of the variance. Further improvements in sample size estimates could potentially be achieved by modeling out confounding factors such as age and sex, as is done in [74], and by enrichment techniques that take into account ApoE genotype or family history of AD.

There are two other caveats about comparing n80s that have been raised before. First, basing a measure of brain change on a certain region or parameter of the brain may overlook valuable disease-modifying effects that affect other regions or measures. This is an inevitable consequence when choosing any surrogate outcome measure. Second, the slowing of a change measure by 25% may have different value to the patient, depending on whether the measure is volumetric loss, amyloid clearance, or decline in cognition. As a result, the n80 is a guide to biomarker utility that must be weighed against other relevant criteria, in much the same way as we advocated the weighting of multiple features within an image here, rather than relying on any one marker of disease progression.

5.5 Figures

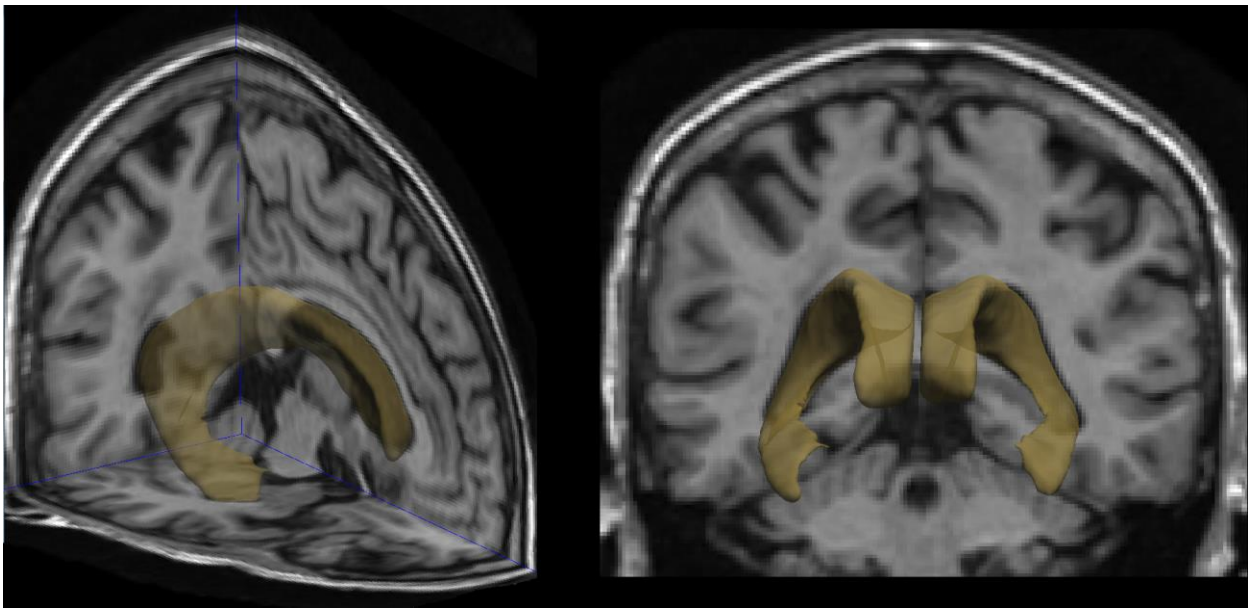

Fig. 1. Lateral ventricles in the human brain.

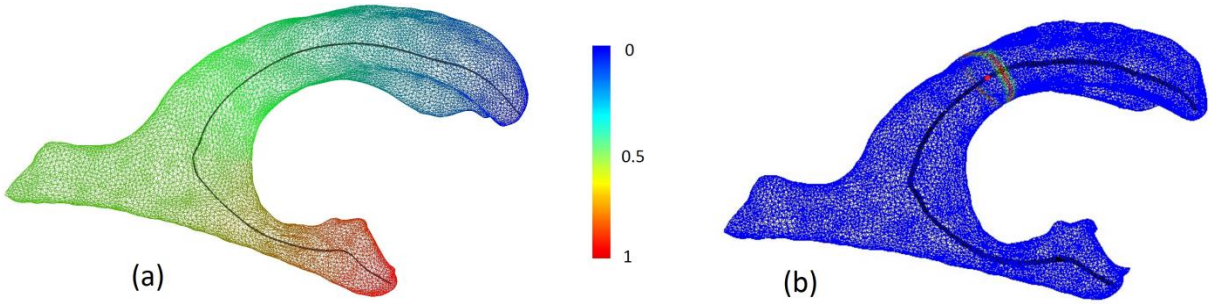

Fig. 2. The medial curve of a lateral ventricle surface in one subject from the ADNI cohort. (a) Mesh vertices are colorized by the corresponding Global Orientation Function. (b) Surface weight map from equation 2 corresponding to the curve point marked in red. The weight is maximal at the cross-section of the surface with the normal plane of the curve, and decays quickly away from the normal plane.

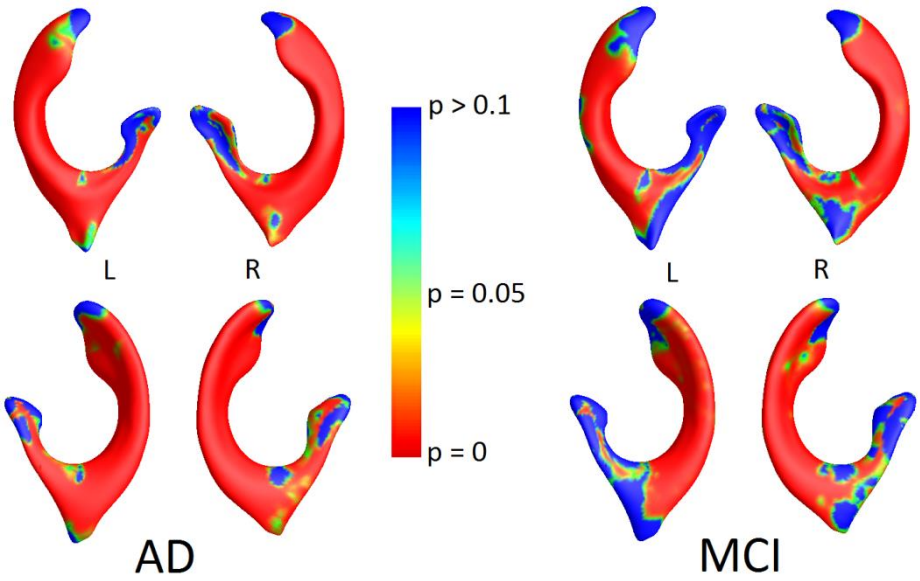

Fig. 3. P-maps show the group differences in annual atrophy rates between healthy controls and (*left*) AD, and (*right*) MCI subjects. The progressive expansion from normal aging to MCI/AD is in line with prior reports. Loss of significance near the ends of the medial curve are likely due to the nature of the measurement rather than true anatomical change.

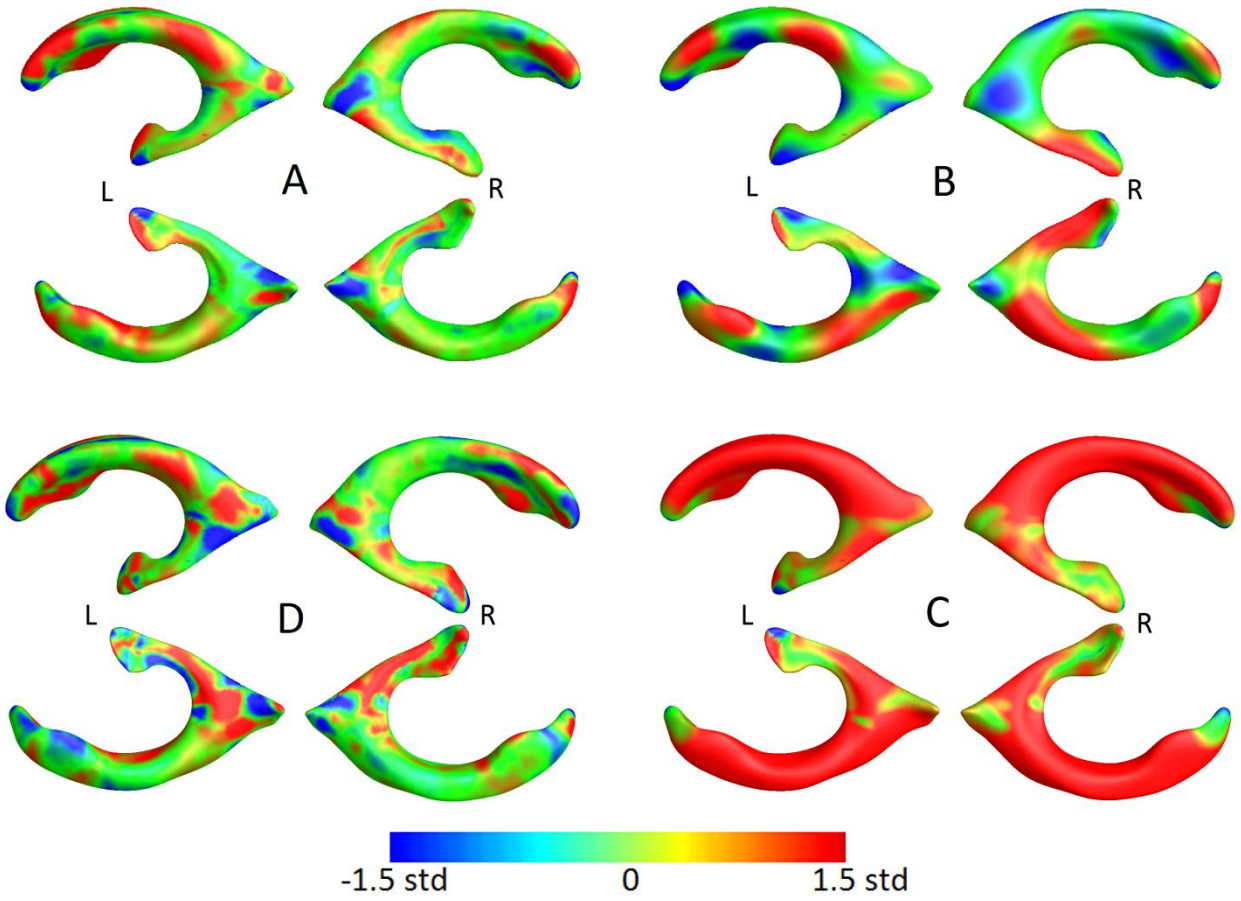

Fig. 4. Continuous weight maps, scaled by standard deviation of the weights**.** (a) PCA-LDA; (b) Tikhonov-LDA; (c) t-statistic weighting; (d) 2-class PCA-LDA. The weights in (a) and (b) are quite different from the stat-ROI, which indicates that areas of importance in detecting atrophy do not always correspond to the area with highest *t*-statistic. Compared to (a), (b) shows a similar, but smoother pattern. Most of the area is positively weighted - as expected - though some ventricular contraction is used for a scalar measure as well. This may be partially explained by registration artifact and imprecision in the medial axis, as there is no obvious biological explanation. (d) shows a more disease-specific atrophy pattern. Significantly more weight is given to the inferior horns bordering the hippocampus, and more weighting is also given to the middle of the occipital horns, characteristic of white matter degeneration. Unlike (a)-(c), this map is directly comparable with figure 3. The pattern is again different compared to a mass-univariate weighting.

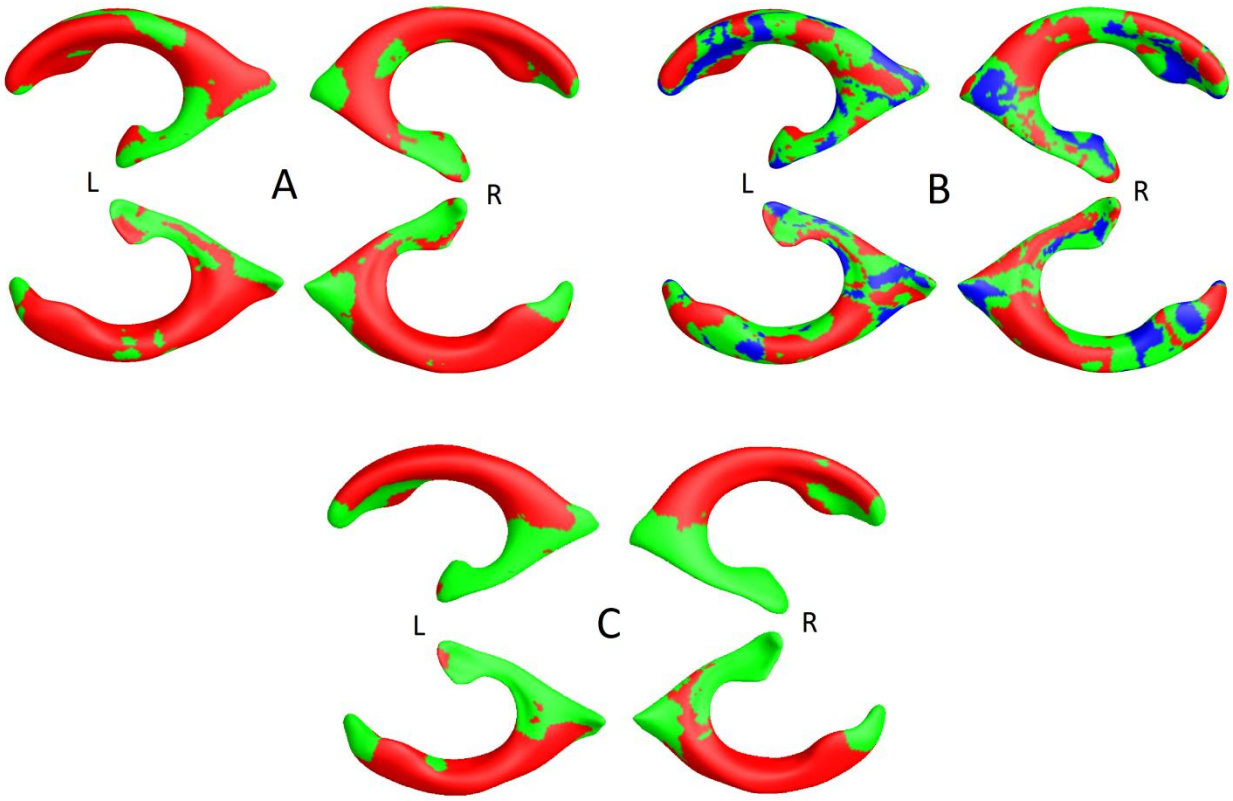

Fig.5. Discrete weight maps. (a) Unsigned LDA-ROI; (b) Signed LDA-ROI; (c) Stat-ROI. Positive ROI is colored in red, negative in blue.

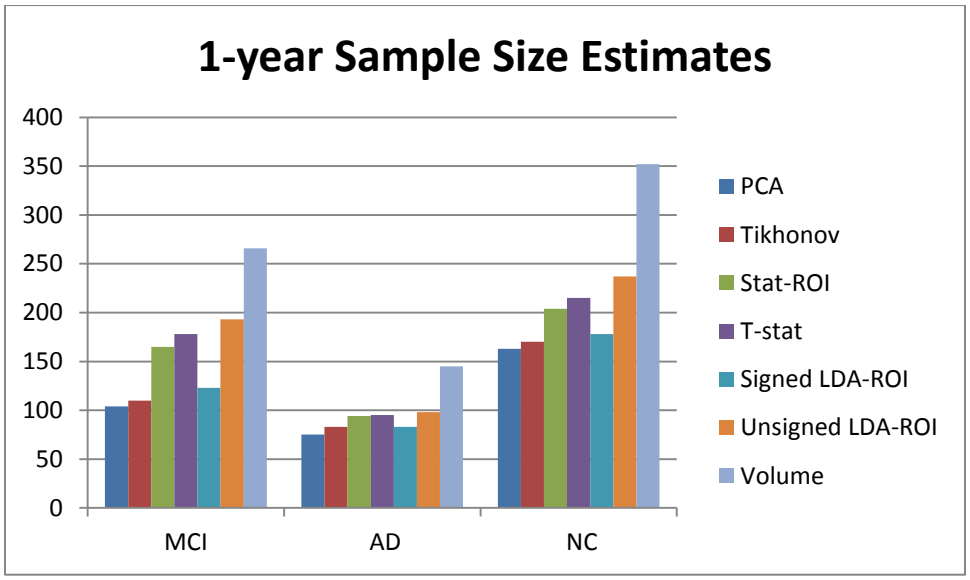

Fig. 6. Sample Size Estimates at 1 year. The surface-based estimates are based on nested cross-validation, and averaging the estimates in each non-overlapping fold. PCA and Tikhonov methods require nearly identical samples, while Stat-ROI requires a significantly larger sample of MCI subjects. All surface-based methods need fewer subjects than ventricular volume.

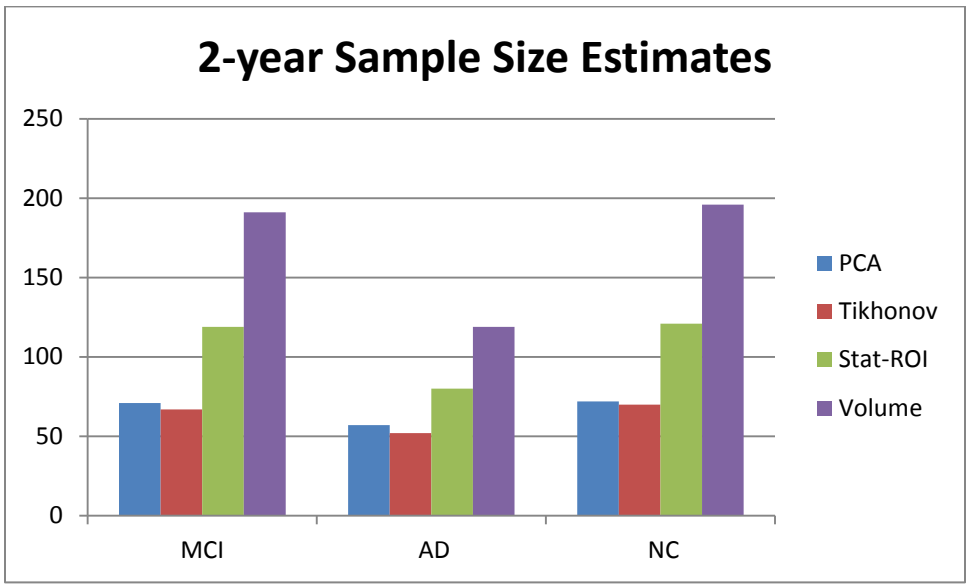

Fig. 7a. Sample Size Estimates at 2 years. The pattern is similar to estimates at 1 year. The relative difference between stat-ROI and LDA-based methods is greater.

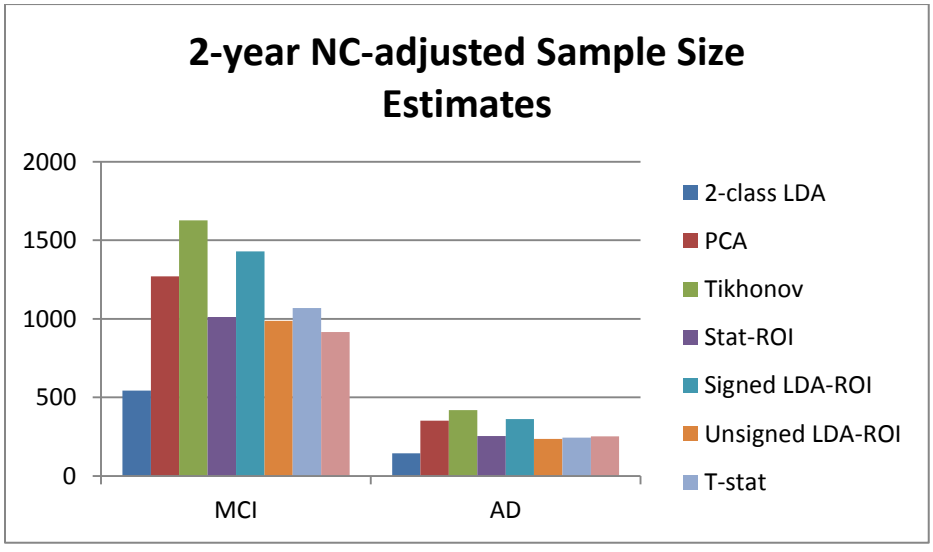

Fig. 7b. NC-adjusted Sample Size Estimates at 2 years. The 2-class model outperforms all other measures. Volume is second best for MCI and on par with the best 1-class measures for AD.

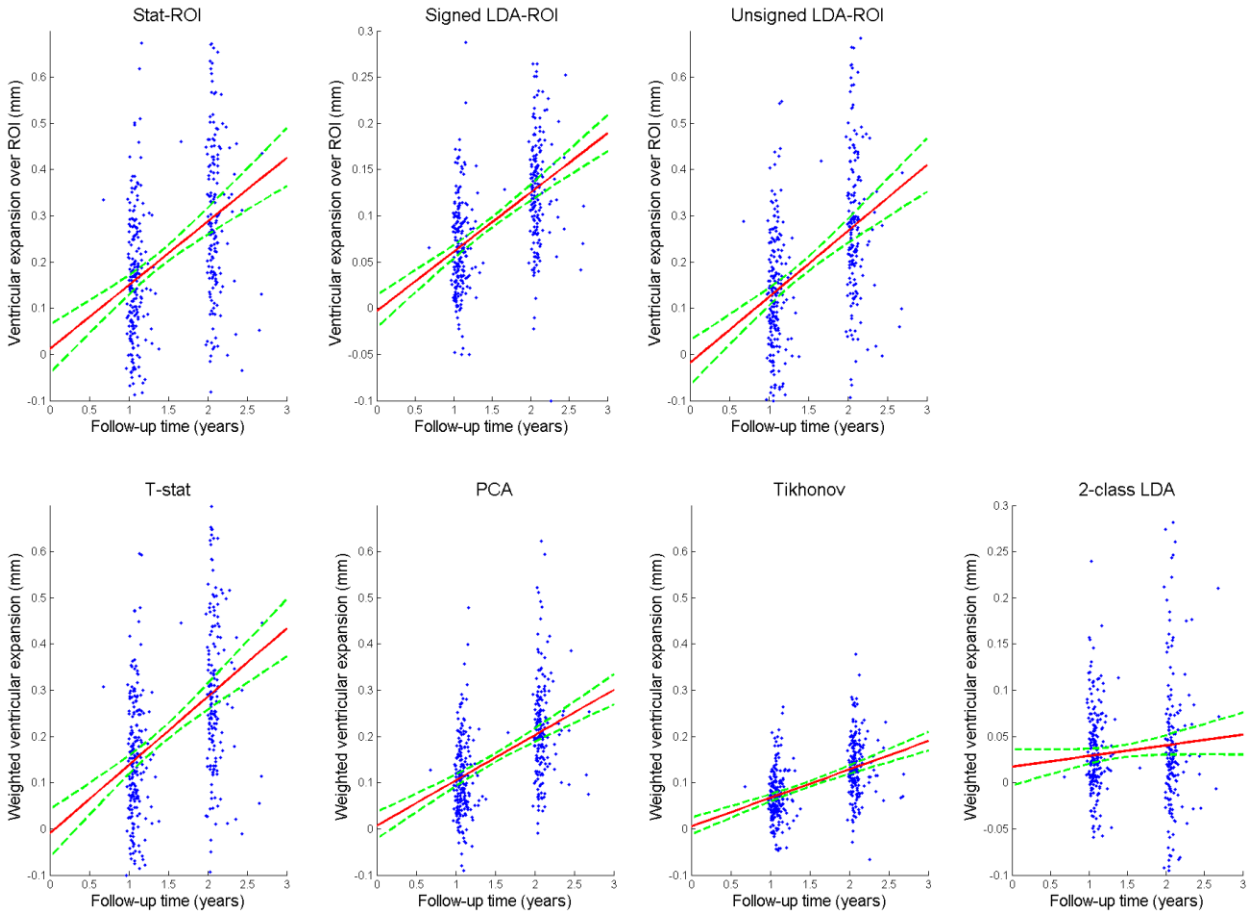

Fig. 8. Regression plots for surface-based ventricular expansion measures in controls**.** 95% confidence belts for the regression models are shown with dotted green lines. All surface models are longitudinally unbiased, since the zero intercept is contained in the 95% confidence interval on the intercept, for each of the methods. The 2-class model is trending on additive bias; however, in this model the mean of controls is subtracted for power estimates.

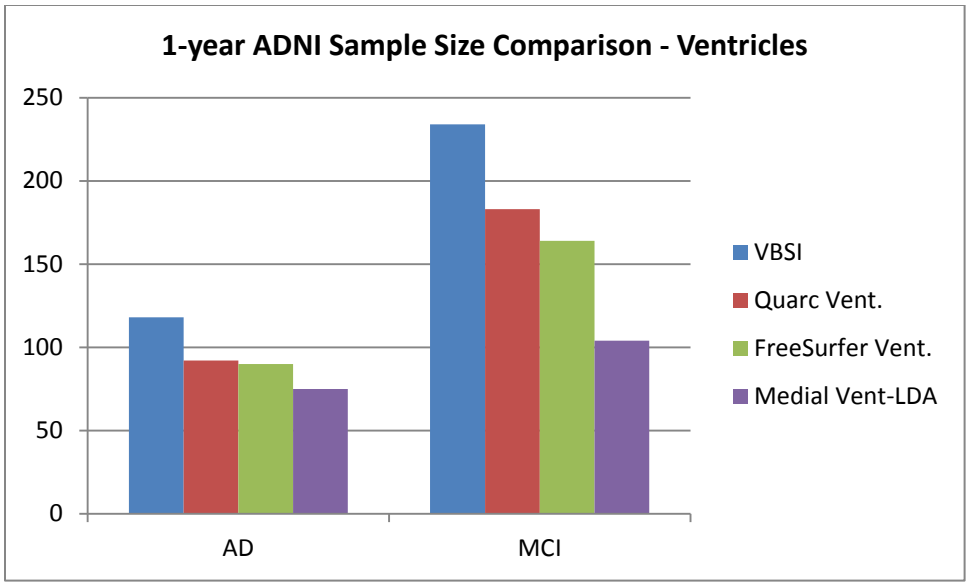

Fig. 9a. Sample size estimates for different ventricular biomarkers. The PCA method is designated here as "Medial Vent-LDA." The Quarc and Freesurfer measures are for a two-year trial.

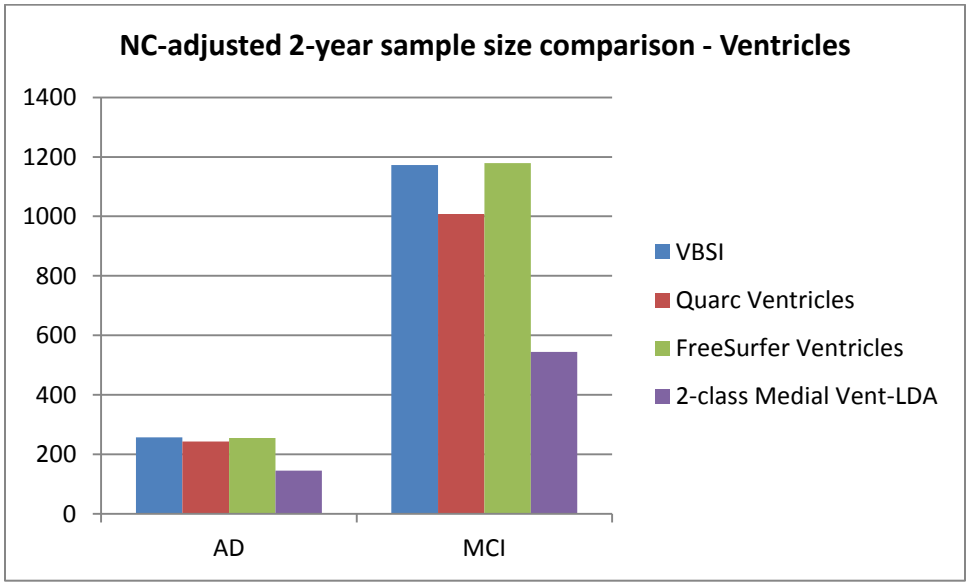

Fig. 9b. NC- adjusted sample size estimates for different ventricular biomarkers. The 2-class PCA method is designated here as "Medial Vent-LDA." All other results are taken from [73].

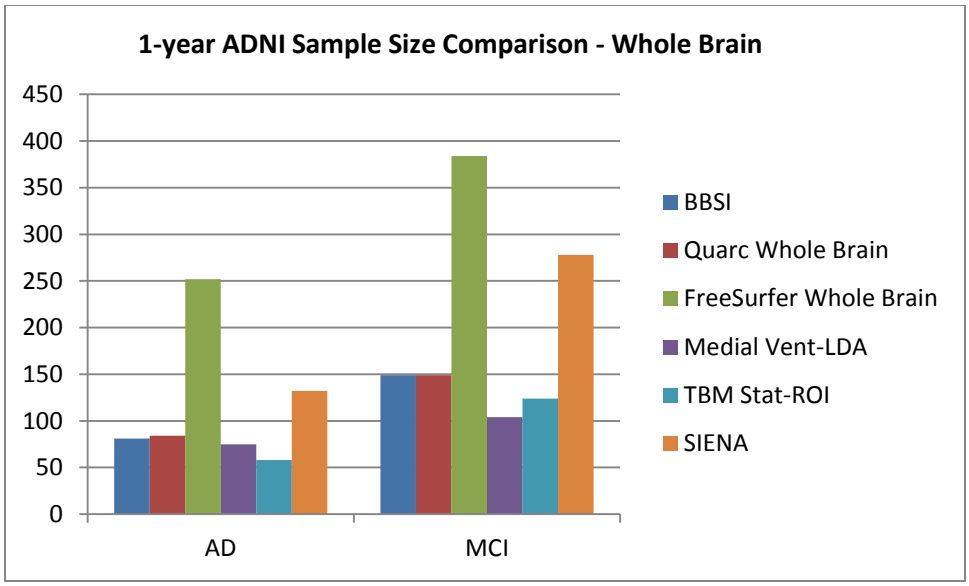

Fig. 10. Sample size estimates for different whole biomarkers compared to our ventricular measure. The PCA method applied to ventricular surfaces is designated here as "Medial Vent-LDA." The Quarc and Freesurfer measures are for a 2-year trial. SIENA estimates are computed from the means and variances reported in [123]

CHAPTER 6

Robust Shape Correspondence via Spherical Patch Matching for Atlases of Partial Skull Models

This chapter is adapted from:

B. Gutman, R. McComb, J. Sung, W. Moon, and P. Thompson, "Robust Shape Correspondence via Spherical Patch Matching for Atlases of Partial Skull Models," in *Mesh Processing in Medical Image Analysis 2012*. vol. 7599, J. Levine, R. Paulsen, and Y. Zhang, Eds., ed: Springer Berlin Heidelberg, 2012, pp. 89-100.

Abstract

Problems of dense partial correspondence for meshes of variable topology are ubiquitous in medical imaging. In particular, this problem arises when constructing average shapes and probabilistic atlases of partial skull models. We exploit the roughly spherical extrinsic geometry of the skull to first approximate skull models with shapes of spherical topology. The skulls are then matched parametrically via a non-local non-linear landmark search using normalized spherical cross-correlation of curvature features. A dense spherical registration algorithm is then applied for a final correspondence. We show that the non-local step is crucial for accurate mappings. We apply the entire pipeline to low SNR skull meshes extracted from conical CT images. Our results show that the approach is robust for creating averages for families of shapes that deviate significantly from local isometry.

6.1 Introduction

Registration of topologically variably shapes is an often-encountered problem in medical imaging. Unlike registration of topologically equivalent models, the problem requires additional flexibility and the ability to perform partial matching. A basic solution is the well-known iterative closest point algorithm (ICP) [7]. Applying the EM algorithm to the problem of point correspondence, ICP alternates between rigid alignment and local correspondence search. The major drawback of ICP is the very local nature of its search, and the use of Euclidian distance in the ambient space, which can be a crude approximation to true geodesic distance. A more general solution proposed by Bronstein et al. [129] is to generalize the multidimensional scaling algorithm (MDS) from Euclidian space to a general 2-manifold, replacing the Euclidian metric with geodesic distance. Such a framework naturally leads to full and partial matching between surfaces potentially differing in connectivity. A further improvement upon this framework was proposed in [130], where the geodesic distance is replaced with diffusion distances, while keeping polynomial complexity. Such a metric relies on the average path length between two points, rather than just the shortest, making it much more robust for cases of variable connectivity. The drawback here lies in the assumption that locally the surfaces are sufficiently isometric, an assumption that cannot always be satisfied by anatomical shape models.

Another traditional alternative is to embed surfaces in a canonical space [13, 28], and perform registration in this space. While this eliminates the requirements that the shapes be near-isometries, or that the search be purely local, it requires that we approximate our shapes with shapes of appropriate topology to enable their embedding. Since we are interested in creating a statistical average from a population, we must already assume a particular topology for the atlas, without necessarily knowing a "correct" topology a priori. Thus, a canonical space approach may be most appropriate for our needs.

There are also practical considerations. Our goal is to create a fast pipeline that is immediately useable in the clinic without the need to tune any parameters, and robust enough to handle meshes of any quality, like the four real skull meshes in **Figure 1**. We would like to create a pipeline that is robust with respect to the quality of triangulation and segmentation, and low in complexity. For example, computing the Laplace-Beltrami eigenfunctions required for diffusion distances is highly sensitive to triangulation quality [5], and so may not be suitable for an initial correspondence search. On the other hand, a purely statistical model, the Active Appearance Model (AAM) [131] can be used to establish correspondence without relying on high mesh quality. AAM couples texture and spatial information to create a compact model of appearance. A correspondence search is then driven by maximizing the agreement between the new shape and the learned model. AAM works well when the new example to be registered is sufficiently similar to the training examples, but generally does not handle outliers very well.

For these reasons we choose to go back to a common canonical space – the 2-sphere. Having chosen this space, we must now ensure that our initial correspondence search is not restricted to be local, and is robust with respect to outliers and incomplete matches, as required

by our data. A correlation-type search is ideally suited for this purpose: while the localization may be coarse, correlation is guaranteed to find a global optimum in the space of allowable transformations. This is the key idea of our approach.

Briefly, our method consists of the following steps. We first use a straight-forward spherical mapping, similar to the star-map to initialize a partial spherical parameterization of non-genus zero skull models. We then approximate our skull models with shapes of spherical topology with an approach similar to minimal surfaces, based on the spherical heat kernel. Next, the genus-zero shapes are registered parametrically with a novel weighted normalized cross-correlation (WNCC) algorithm. The idea is to use spherical maps of curvature features and a global search to optimize positions of multiple surface patches simultaneously. A dense initial correspondence is found with spherical splines with anchor points defined by patch centers [132]. An existing spherical registration algorithm is then applied for a final dense correspondence.

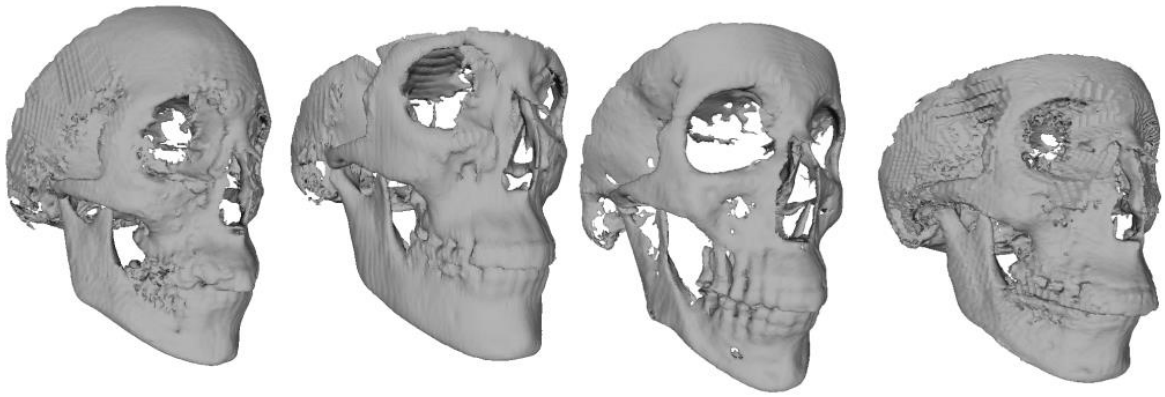

Fig. 1. Four partial skull models.

The remainder of the paper is organized as follows. In the second section we describe briefly our initial spherical mapping and spherical shape approximation. In the third, we go through the mathematical preliminaries of weighted normalized cross-correlation on the 2-sphere. The fourth section is devoted to our non-local landmark correspondence search for multiple landmarks. We briefly touch on the approach for dense spherical correspondence. Section 5 presents results on 67 real skull models, and section 6 concludes the paper.

6.2 Spherical Parameterization

Our spherical parameterization is a slightly generalized star map. It is based on the principal axis of the shape. By principal axis, we mean the line through the center of mass, in the first principal direction of the shape's inertia matrix, parameterized by $\theta \in [0, \pi]$. Given a surface $\mathcal{M}$, we define the axial star map and a corresponding "filler" mask as

$$S_{\boldsymbol{a}}(\theta,\varphi) = \begin{cases} Star_{\boldsymbol{a}}(\theta,\varphi), & if\ \exists\ R\ \text{s.t.}\ \boldsymbol{a}(\theta) + R\boldsymbol{u}_G(\theta,\varphi) \in \mathcal{M} \\ \boldsymbol{a}(\theta) + \overline{r(\theta)}\boldsymbol{u}_G(\theta,\varphi), & otherwise \end{cases} \tag{9}$$

$$M(\theta,\varphi) = \begin{cases} 1, & if\ \exists\ R\ \text{s.t.}\ \boldsymbol{a}(\theta) + R\boldsymbol{u}_G(\theta,\varphi) \in \mathcal{M} \\ 0, & otherwise \end{cases} \tag{10}$$

Here, $\boldsymbol{a}(\theta)$ is the principal axis, $\boldsymbol{u}_G$ is a unit vector normal to the tangent direction of $\boldsymbol{a}(\theta)$, rotated around the tangent direction by φ. $Star_{\boldsymbol{a}}(\theta,\varphi) = \boldsymbol{a}(\theta) + r\boldsymbol{u}_G(\theta,\varphi)$, $r = max\{R \in \mathbb{R} \mid \boldsymbol{a}(\theta) + R\boldsymbol{u}_G(\theta,\varphi) \in \mathcal{M}\}$, and $\overline{r(\theta)}$ is the mean radial extension at θ. Though simplistic, this mapping is a robust and fast initialization for our spherical approximations. The principal axis-based star map is well suited for skull models and leads to much smaller metric distortions than the usual centroid-based star map in practice.

To smooth the regions of the spherical shape which do not correspond to real data, we approximate heat kernel smoothing with boundary conditions. This is accomplished by applying successive heat kernels on the sphere. The following pseud-code describes the approximation. We assume that we have a set of $N_\sigma$monotonically decreasing radii of the heat kernel $\sigma(i)$, and a given tolerance ε for convergence at all levels of smoothing.

```
Inputs: N_σ, ε, σ(i), spherical map S: 𝕊² ⟶ ℝ³,
        binary map M: 𝕊² ⟶ {0,1};
i = 0;
WHILE (i < N_σ)
  Set σ = σ(i);
  j = 0;

  WHILE (‖S_j − S_{j−1}‖ > ε)
    Compute S_σ = K_σ ∗ S_{j−1}
    Set S_j by S_j(p) = { S_σ(p)      if M(p) = 0
                        { S_{j−1}(p)  otherwise
    j = j + 1;
  ENDWHILE

  i = i + 1;
ENDWHILE
```

The result is similar to directly approximating minimal surfaces of the filler mask boundary curves. Because S_σ is computed very quickly on the sphere [133], the entire approximation is very fast and robust.

6.3 Weighted Normalized Cross Correlation on $\mathbb{S}^2$

To initialize a robust correspondence between highly variable noisy shapes, we develop a weighted normalized cross correlation on the sphere for use with multiple informative surface patches. Some of the details in this section are given for completeness – the main goal of this

tool is to enable an initialization that is (a) robust, (b) fast, and (c) based only very loosely on statistical learning, to enable reasonable initialization for outlier cases.

We follow the normalized cross-correlation (NCC) algorithm on the 2-sphere presented in [134]. In chapter 1 we introduce the ordinary correlation function between a spherical function *I* and a rotated version of another function, or template *T*.

$$C(R) = \int_{\mathbb{S}^2} I(\omega)\Lambda(R)T(\omega)d\omega. \tag{11}$$

Alternatively, this can be expressed in terms of spherical harmonic coefficients as

$$C(R) = \sum_{l} \sum_{|m|\leq l} \hat{I}_{l,m}\overline{\widehat{\Lambda(R)T}_{l,m}}. \tag{12}$$

Here as before, $\Lambda(R)$ is the rotation operator on a spherical function. In chapter 1 we saw that this fact can be exploited to derive a fast *O(B^4)* spherical cross-correlation algorithm for computing $C(R)$ non-iteratively over the entire space of rotations.

Suppose now the support of the template function *T* is not the entire sphere, but a proper subset, or patch $\mathcal{P} \subset \mathbb{S}^2$. In this case (7) would not be sufficient as a true measure of match quality, as it is biased by variability in local intensity. We would instead like to compute the normalized cross correlation

$$NC(R) = \frac{\int_{\mathbb{S}^2}(I(\omega) - \bar{I}_{\mathcal{P}})(\Lambda(R)T(\omega) - \bar{T})d\omega}{\sqrt{\int_{\mathcal{P}}|I(\omega) - \bar{I}_{\mathcal{P}}|^2 d\omega \ \int_{\mathcal{P}}|T(\omega) - \bar{T}|^2 \, d\omega}}. \tag{13}$$

The difficulty, compared to Euclidian NCC, lies in computing the term $\int_{\mathcal{P}}|I(\omega) - \bar{I}_{\mathcal{P}}|^2 d\omega$, as the shape of the window $\mathcal{P}$ changes for every rotation. As suggested in [135] and done in [134], this

can be done by computing two additional correlation functions between a binary mask of the support patch $M_{\mathcal{P}}$, and both the image I and its square:

$$\int_{\mathcal{P}} |I(\omega) - \overline{I_{\mathcal{P}}}|^2 d\omega = \int_{\mathbb{S}^2} I^2(\omega)\Lambda(R)M_{\mathcal{P}}(\omega)d\omega - \frac{\left[\int_{\mathbb{S}^2} I(\omega)\Lambda(R)M_{\mathcal{P}}(\omega)d\omega\right]^2}{\int_{\mathbb{S}^2} M_{\mathcal{P}}(\omega)\, d\omega}. \quad (14)$$

If in addition we wish to weigh our NCC with some weighting function W: $\mathcal{P} \subset \mathbb{S}^2 \to \mathbb{R}$, equation (11) becomes

$$NC(R,W) = \frac{\int_{\mathbb{S}^2} W(\omega)\big(I(\omega) - \overline{I_{\mathcal{P},\mathrm{W}}}\big)(\Lambda(R)T(\omega) - \overline{T_W})d\omega}{\sqrt{\int_{\mathcal{P}} W(\omega)\big|I(\omega) - \overline{I_{\mathcal{P},\mathrm{W}}}\big|^2 d\omega \;\int_{\mathcal{P}} W(\omega)|T(\omega) - \overline{T_W}|^2\, d\omega}}, \quad (15)$$

where $\overline{I_{\mathcal{P},\mathrm{W}}} = \int_{\mathcal{P}} I(\omega)\Lambda(R)W(\omega)d\omega / \int_{\mathbb{S}^2} W(\omega)d\omega$, and $\overline{T_W}$ is defined analogously. Expanding the term $\int_{\mathcal{P}} W(\omega)\big|I(\omega) - \overline{I_{\mathcal{P},\mathrm{W}}}\big|^2 d\omega$, it is clear that to compute it one simply has to replace $M_{\mathcal{P}}$ in (12) with W:

$$\int_{\mathcal{P}} W(\omega)\big|I(\omega) - \overline{I_{\mathcal{P},\mathrm{W}}}\big|^2 d\omega = \int_{\mathbb{S}^2} I^2(\omega)\Lambda(R)W(\omega) - \frac{\left[\int_{\mathbb{S}^2} I(\omega)\Lambda(R)W(\omega)\right]^2}{\int_{\mathbb{S}^2} W(\omega)}. \quad (16)$$

With $\overline{T_W}$ and the other term in the denominator of (13) being constant, this makes the number of operations required for a spherical WNCC identical to the usual NCC. While this is a trivial extension of [10], we have not seen it in literature.

6.4 Non-Local Correspondence Search with Multiple Patch Matching

Suppose we have selected multiple corresponding landmarks on several example skull models (**Figure 2**), and wish to find their counterparts on a new skull. A direct approach would be to select a patch of a spherical curvature map around each landmark, and approximate its location

with NCC over the new skull map. While this is straightforward, this approach fails in practice for several reasons. First, a straightforward NCC does not take into account the plausibility of a given patch transformation, given that the two shapes are already rigidly aligned on the sphere. Second, NCC does not consider the relative stability of curvature features at any given point in a patch. And third, the approach does not take into account the relative positions of all other patches that are being matched, which could lead to highly improbable landmark placement. We develop an approach for multiple patch matching that addresses each of these issues in turn.

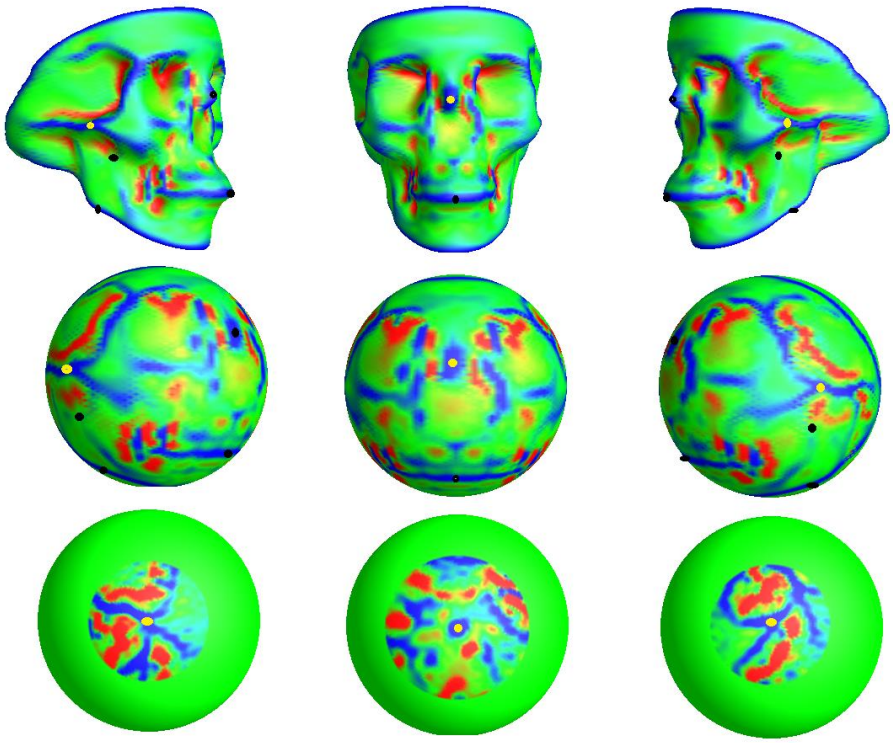

Fig. 2. Spherical Patches. First row: spherical skull model with landmarks marked by black and yellow dots. Second row: Spherical curvature map with corresponding landmarks. Third row: patches around landmarks marked with yellow dots above.

6.4.1 Single Patch Priors

To deal with the first issue, we modify each individual NCC with two prior terms for spatial congruence:

$$NC_{mod}(R) = NC(R)\left(\frac{a_\gamma - 1}{a_\gamma - \cos\gamma}\right)\left(\frac{a_p - 1}{a_p - \langle p_o, p_R\rangle}\right). \quad (17)$$

Here $a_\gamma, a_p \in (1, \infty)$ are parameter weights, p_o is the original spherical position of the patch, γ has the same meaning as in equation 5, and $p_R = R \circ \eta$ is the North Pole rotated by *R*. As all patches are moved to the North Pole for correlation, this represents the new position of p_o. The first term penalizes implausible patch orientations, while the second does the same for patch placement relative to the original position of the template. In practice, even this modification alone turns out to be sufficient for correct placement of skull patches in many cases.

To address the issue of spatially variable feature stability in a patch, we apply the weighted version of NCC. The weights in this case should reflect the variability at each patch point, which could be learned from examples. This is the only aspect of our approach that is at all akin to statistical learning: given multiple examples of the same patch, we weigh NC_{mod} with the inverse of the standard deviation of the template function *T*, i.e. $W(\omega) = \frac{1}{\sigma_T(\omega)}$.

6.4.2 Combining Multiple Patches

Finally, to address the problem of congruence *across* patches, we apply two more prior terms for a final combined score of match quality. Having computed WNCC for several patches

T^i, $NC^i_{mod}(R^i, W^i)$, we would like to find the set of optimal patch rotations $\{R^i_{opt}\}_{i=1}^N$ for each of N landmarks. We define this optimal set as the maximal argument of the combined match score:

$$NC_{combined}\left(\{R^i\}_{i=1}^N\right) = \prod_i NC^i_{mod}(R^i, W^i) \prod_{j \neq i} C_{orient}\left(R^i, R^j\right) C_{dist}(R^i, R^j). \quad (18)$$

The two terms C_{orient}, and C_{dist} serve an analogous purpose to the two prior terms for single patch correlation in equation (15):

$$C_{orient}(R^i, R^j) = \exp\left[-a_{orient} \cos^{-1}\langle v_o^{i,j}, v_{R^i,R^j}^{i,j}\rangle\right], \quad (19)$$

$$C_{dist}(R^i, R^j) = \exp\left[-a_{dist}\left(\cos^{-1}\langle p_o^i, p_o^j\rangle - \cos^{-1}\langle p_{R^i}^i, p_{R^j}^j\rangle\right)^2\right]. \quad (20)$$

The arguments of the inner product in equation (17) represent tangent vectors of the geodesics between the original and new positions of landmarks i and j. The original tangent vector is transported to lie in the tangent space of $p_{R^i}^i$:

$$v_o^{i,j} = \mathcal{T}(p_o^i, p_{R^i}^i) \circ \frac{Proj_{T_{p_o^i}}(p_o^j - p_o^i)}{\left|Proj_{T_{p_o^i}}(p_o^j - p_o^i)\right|}, \quad (21)$$

$$v_{R^i,R^j}^{i,j} = \frac{Proj_{T_{p_{R^i}^i}}(p_{R^j}^j - p_{R^i}^i)}{\left|Proj_{T_{p_{R^i}^i}}(p_{R^j}^j - p_{R^i}^i)\right|}, \quad (22)$$

where by $Proj_{T_p}$ we mean the projection onto the tangent plane of p, and $\mathcal{T}(v, w)$ is the parallel transport operator along the great circle from v to w.

6.4.3 Search Strategies for Multiple Patch Placement

In order to optimize (16), we must select an appropriate search strategy. It is clear that an exhaustive search over all possible locations for every patch is implausible even for the coarsest tessellation of *SO(3)* and $\mathbb{S}^2$. This is because an exhaustive search has exponential complexity. Given L possible locations on the sphere, and N patches, such a search requires $\mathcal{O}(N^2L^N)$ operations. A straightforward simplification is to select only the top few peaks from each individual correlation function. While this makes the search possible when there are only a few patches, this strategy also becomes impractical when N exceeds 10-15, depending on L.

A further reduction in complexity can be achieved by adopting a hierarchical search approach similar to fast sorting strategies. Instead of considering all N patches at once, we spatially cluster the patches into groups of K elements, and find the top L sets of locations for each cluster. We repeat this process so that the previous clusters form elements in the new clusters, again selecting only the top L sets of locations. This process is repeated until all patches are incorporated in a final search. The complexity of this strategy is $\mathcal{O}(N^2L^K log_K N)$. For a reasonable L and K, e.g. $L, K = 6$, this makes the search quick even for relatively large N.

6.4.4 Dense Surface Correspondence

To initialize a dense correspondence, we extrapolate a spherical warp from landmark displacements using the spherical splines [132]. To achieve a final correspondence, we modify an existing, highly robust framework for unconstrained spherical mapping proposed by Freidel et al. [55], by adding L^2 terms to the cost. Using Freidel's notation, the cost function becomes

$$E = aE_{conformal} + bE_{area} + \sum_{i=1}^{n} w_i \int_{\mathbb{S}^2} \big(f_i(\omega) - g_i(\omega)\big)^2 \, d\omega, \qquad (23)$$

where f_i, g_i are scalar features of the moving and stationary spherical image. We use mean curvature κ and a smoothed filler mask $K_\sigma * M$. More details on this step may be found in [55] and [91].

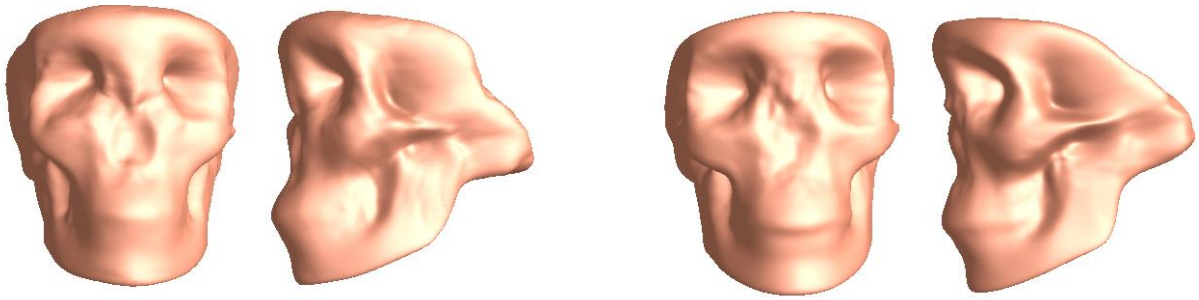

Fig. 3. Two skull averages. Left: local search only. Right: non-local initialization

6.5 Experiments

We applied our registration framework to 67 skull models of healthy subjects, acquired with a NewTom 3G conical CT scanner. The surface models were extracted from the CT images using Dolphin software. In the non-local step of our pipeline, we experimented with different numbers of landmarks and examples for each patch. A reasonable mapping for our data could be achieved with as few as 8 landmarks and only one example subject. However, for added robustness we use 25 landmarks and 10 subjects. We used circular patches of pre-set radius around each landmark, with the radius varying between $\pi/7$ and $\pi/5$. Landmark origins were set to be the geodesic average of the training landmarks. The average patch for each landmark was computed by warping each example curvature map to the average position. This was done by computing landmark displacement to the average and using spherical splines. The entire training process took less than 1 minute. We used the hierarchical search with $K=5$ elements per cluster and $L=10$ top sets of landmark placements. Using a bandwidth of B=64, our total time for the

multiple patch matching stayed between 2 and 3 minutes, running on a Lenovo W520 ThinkPad with an Intel i7 2820QM processor. RAM requirements for 25 landmarks are approximately 500 MB, although using the hierarchical search allows this to be greatly reduced if required.

To assess the need for our non-local initialization step, we compared the proposed method to using only local correspondence of equation (21). We first computed shape averages using both methods, and computed average vertex-wise distance between the average and each shape after Procrustes alignment. A paired t-test revealed that the distance decreases significantly when the non-local step is applied, with $p = 0.026$. The vertex-wise distance decreased by 0.63 mm on average with the use of multiple patch matching. Training subjects were excluded from this comparison. The resulting atlases are shown in **Figure 3.**

In the second part of the comparison, we assessed the compactness of our two atlases with Principal Components Analysis. A model is said to be more compact if a greater proportion of the total variance in the data is captured by a given number of principal components (PC) [23]. We plot the total percent of variance explained versus the number of PC's in **Figure 4**, disregarding the first component as it primarily captures affine misalignment. Note that the curve resulting from the non-local step dominates, which shows that this model is more compact. **Figure 5** shows this effect visually. By varying each shape atlas along its first 3 PC's, we see that the non-local atlas captures anatomically plausible variability, while the local-only atlas primarily captures error in registration.

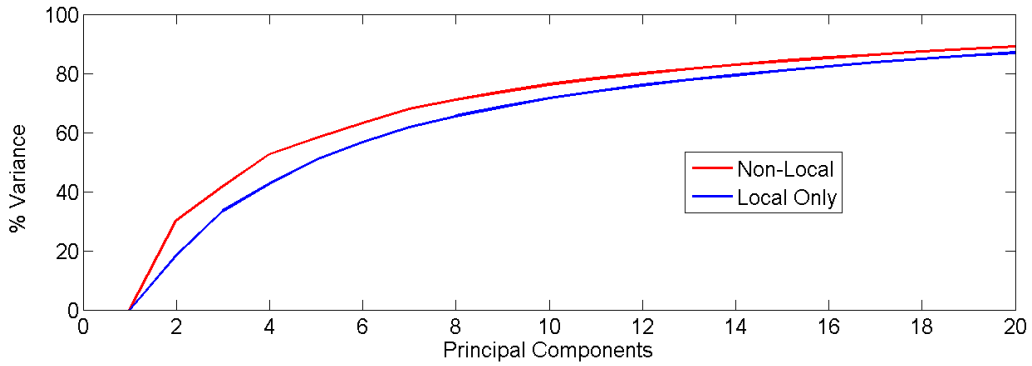

Fig. 4. Model compactness (% of explained variance vs. number of principal components). The non-local model is more compact.

6.6 Conclusion

We have presented a robust pipeline for quickly computing shape atlases from incomplete, highly variable, and noisy skull models. The key feature of our pipeline is a non-linear non-local correspondence search, made fast by exploiting analytic properties of the sphere and the rotation group. Our approach is based on very few training data (as low as 1) and few manually labeled landmarks, which makes it more practical than models more rooted in statistical learning, such as the Active Appearance Model. The basic idea is to compute global correlation functions for patches around example landmarks for locating corresponding landmarks on new shapes automatically. We modify an existing fast normalized cross correlation algorithm on the sphere to enable its weighted version. The weighted NCC helps to further improve the robustness of our pipeline. We have applied our pipeline to 67 human skull models, and showed that an anatomically plausible shape atlas is only achieved when the non-local initialization is used in shape registration. Our non-local search leads to statistically significant improvements in distance to the average shape.

A major drawback of this framework is the admittedly ad-hoc nature of our prior terms for individual and combined match scores (eqns. 15 and 16), and the weighting function. A more principled derivation of prior terms may be possible if one assumes a binary spherical feature function. In such a case, the normalized cross-correlation is equivalent to a difference of conditional probabilities, which may enable us to cast this problem in a Bayesian framework, while keeping NCC as the main workhorse of the algorithm. In this case, probabilistic spatial priors could be learned from the training data to use in place of the average case. Another direction for future work would be to replace manual labeling with automatic selection of informative patches.

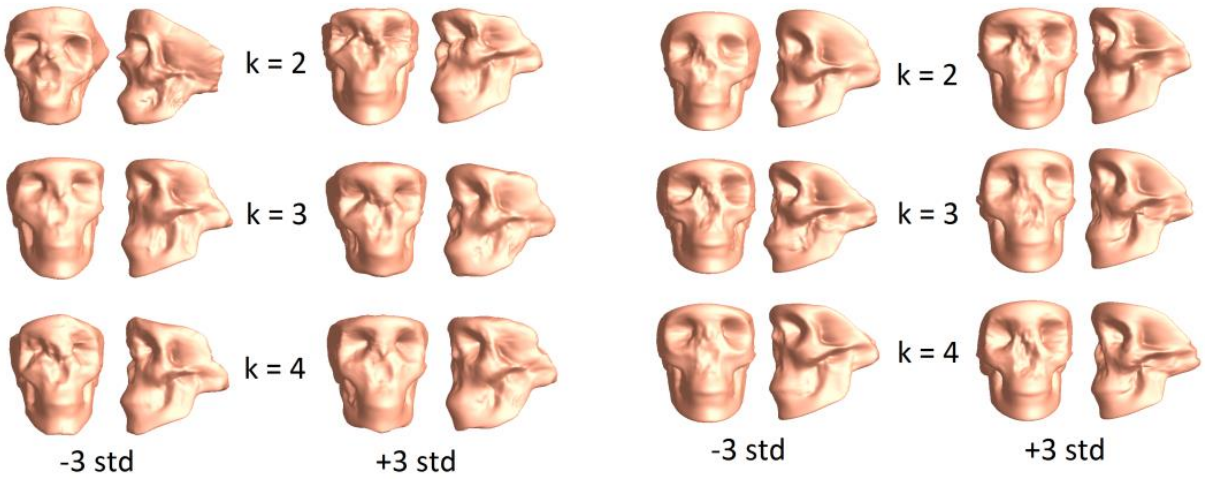

Fig. 5. Shape variation along principal components. Left: local only. Right: Non-local.

CONCLUSION OF THE DISSERTATION

With the goal of developing a variety of tools for analyzing and comparing anatomical shapes, I have presented several complimentary approaches, and tested them on cortical surfaces, boundaries of subcortical brain regions, and 3D human skull models. The resulting tools showed great promise, in some cases outperforming existing methods.

Focusing on the shape of human hippocampus – a deep structure in the brain implicated in neurodegenerative disease – I showed that global shape description based on properties of spherical harmonics can augment volume in discriminating hippocampal shapes of Alzheimer's patients from those of controls.

I developed a robust non-iterative registration technique for shapes based on spherical cross-correlation. The technique is general, and is shown to improve significantly the results of spherical rigid registration compared to existing methods in this category, when using spherical conformal parameterization.

To develop an improved medial shape representation, a variational framework was proposed. In this framework, a medial core of fixed topology is evolved to fit the geometry of the shape. Using this medial representation, surface-based feature functions are induced and used for dense non-linear surface registration on the sphere. An additional 1-D group-wise registration step was added, exploiting the generally elongated shape of subcortical structures. I showed that the new medial-spherical registration leads to greater statistical sensitivity in an Alzheimer's study, compared to an existing method. Further, I showed that the 1-D step adds additional sensitivity to the local measures, which are easily visualized as "p-maps," or heat maps of significance on the surface.

Cortical surfaces were registered using a novel adaptation of well-known Euclidean registration algorithms to the sphere. Curvature features were used to drive the registration. It was shown that the registration recovers complicated synthetic warps, and performs very well across different subjects. In addition, the registration is the fastest such tool for cortical shapes to date, able to register two cortical meshes of 200K+ vertices in under 1 minute.

In a deviation from brain-related work, I also developed a robust non-linear non-local registration approach for human skull meshes. I developed a new weighted normalized cross correlation method for the 2-sphere which is sped up considerably with appropriate use of spherical harmonic-based correlation. I showed that the new registration is crucial for skull shapes derived from conical CT scans in creating population-based averages.

BIBLIOGRAPHY

[1] P. M. Thompson, R. P. Woods, M. S. Mega, and A. W. Toga, "Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain," *Human Brain Mapping,* vol. 9, pp. 81-92, 2000.

[2] U. Grenander and M. I. Miller, *Computational Anatomy: An Emerging Discipline*, 1998.

[3] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," presented at the Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing, Aachen, Germany, 2003.

[4] M. Novotni and R. Klein, "3D zernike descriptors for content based shape retrieval," presented at the Proceedings of the eighth ACM symposium on Solid modeling and applications, Seattle, Washington, USA, 2003.

[5] M. Reuter, F.-E. Wolter, and N. Peinecke, "Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids," *Comput. Aided Des.,* vol. 38, pp. 342-366, 2006.

[6] C. Brechbuhler, G. Gerig, and O. Kubler, "Parametrization of closed surfaces for 3-D shape description," *Comput. Vis. Image Underst.,* vol. 61, pp. 154-170, 1995.

[7] S. Granger and X. Pennec, "Multi-scale EM-ICP: A Fast and Robust Approach for Surface Registration," in *Computer Vision — ECCV 2002*. vol. 2353, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds., ed: Springer Berlin Heidelberg, 2002, pp. 418-432.

[8] P. M. Thompson, K. M. Hayashi, G. I. De Zubicaray, A. L. Janke, S. E. Rose, J. Semple, M. S. Hong, D. H. Herman, D. Gravano, D. M. Doddrell, and A. W. Toga, "Mapping

hippocampal and ventricular change in Alzheimer disease," *Neuroimage,* vol. 22, pp. 1754-66, Aug 2004.

[9] M. Styner and G. Gerig, "Three-dimensional medial shape representation incorporating object variability," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. II-651-II-656 vol.2.

[10] G. Gerig, M. Styner, D. Jones, D. Weinberger, and J. Lieberman, "Shape Analysis of Brain Ventricles Using SPHARM," presented at the Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'01), 2001.

[11] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinematics," *Image Processing, IEEE Transactions on,* vol. 5, pp. 1435-1447, 1996.

[12] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *Neuroimage,* vol. 45, pp. S61-S72, 2009.

[13] B. T. T. Yeo, M. R. Sabuncu, T. Vercauteren, N. Ayache, B. Fischl, and P. Golland, "Spherical Demons: Fast Diffeomorphic Landmark-Free Surface Registration," *Ieee Transactions on Medical Imaging,* vol. 29, pp. 650-668, Mar 2010.

[14] C. W. Zhu and M. Sano, "Economic considerations in the management of Alzheimer's disease," *Clinical Interventions in Aging,* vol. 1, pp. 143-154, 2006.

[15] J. Ross, P. M. Thompson, P. Tariot, E. M. Reiman, L. Schneider, E. Frigerio, F. Fiorentini, L. Giardino, L. Calzà, D. Norris, H. Cicirello, D. Casula, and B. P. Imbimbo, "Primary and Secondary Prevention Trials in Subjects at Risk of Developing Alzheimer's Disease: the GEPARD-AD (Genetically Enriched Population At Risk of Developing

Alzheimer’s Disease) Studies," presented at the CTAD conference, Monte Carlo, Monaco, 2012.

[16] J. L. Cummings, "Integrating ADNI results into Alzheimer's disease drug development programs," *Neurobiology of aging,* vol. 31, pp. 1481-92, Aug 2010.

[17] B. T. Wyman, D. J. Harvey, K. Crawford, M. A. Bernstein, O. Carmichael, P. E. Cole, P. K. Crane, C. Decarli, N. C. Fox, J. L. Gunter, D. Hill, R. J. Killiany, C. Pachai, A. J. Schwarz, N. Schuff, M. L. Senjem, J. Suhy, P. M. Thompson, M. Weiner, and C. R. Jack, Jr., "Standardization of analysis sets for reporting results from ADNI MRI data," *Alzheimer's & dementia : the journal of the Alzheimer's Association,* Oct 27 2012.

[18] B. A. Gutman, X. Hua, P. Rajagopalan, Y.-Y. Chou, Y. Wang, I. Yanovsky, A. W. Toga, C. R. Jack Jr, M. W. Weiner, and P. M. Thompson, "Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features," *Neuroimage,* vol. 70, pp. 386-401, 2013.

[19] B. Gutman, R. McComb, J. Sung, W. Moon, and P. Thompson, "Robust Shape Correspondence via Spherical Patch Matching for Atlases of Partial Skull Models," in *Mesh Processing in Medical Image Analysis 2012*. vol. 7599, J. Levine, R. Paulsen, and Y. Zhang, Eds., ed: Springer Berlin Heidelberg, 2012, pp. 89-100.

[20] Y. Shi, P. M. Thompson, G. Zubicaray, and Se, "Direct mapping of hippocampal surfaces with intrinsic shape context," *Neuroimage,* vol. 37, pp. 792-807, 2007.

[21] D. Tosun and J. L. Prince, "Cortical surface alignment using geometry driven multispectral optical flow," presented at the Proceedings of the 19th international conference on Information Processing in Medical Imaging, Glenwood Springs, CO, 2005.

[22] B. Fischl, M. Sereno, R. Tootell, and A. Dale, "High-resolution intersubject averaging and a coordinate system for the cortical surface," *Human Brain Mapping,* vol. 8, pp. 272-284, 1999.

[23] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor, "A minimum description length approach to statistical shape modeling," *Medical Imaging, IEEE Transactions on,* vol. 21, pp. 525-537, 2002.

[24] G. Gerig, M. Styner, M. E. Shenton, and J. A. Lieberman, "Shape versus Size: Improved Understanding of the Morphology of Brain Structures," presented at the Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention, 2001.

[25] L. Shen, H. Huang, F. Makedon, and A. J. Saykin, "Efficient Registration of 3D SPHARM Surfaces," presented at the Proceedings of the Fourth Canadian Conference on Computer and Robot Vision, 2007.

[26] J. Talman, "Special functions: A group-theoretic approach," *W.A.Benjamin, Inc.,* 1968.

[27] L. Sorgi and K. Daniilidis, "Template gradient matching in spherical images," *Image Processing: Algorithms and Systems III. Proceedings of the SPIE,* vol. 5298, 2004.

[28] X. Gu, Y. 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,* vol. 23, pp. 949-58, Aug 2004.

[29] D. M. Healy, D. N. Rockmore, P. J. Kostelec, and S. Moore, "FFTs for the 2-sphere-improvements and variations," *Journal of Fourier Analysis and Applications,* vol. 9, pp. 341-385, 2003.

[30] M. Frigo, "A fast Fourier transform compiler," *SIGPLAN Not.*, vol. 34, pp. 169-180, 1999.

[31] P. A. Yushkevich, H. Zhang, and J. C. Gee, "Parametric medial shape representation in 3-D via the Poisson partial differential equation with non-linear boundary conditions," *Inf Process Med Imaging*, vol. 19, pp. 162-73, 2005.

[32] M. Niethammer, M. Reuter, F. E. Wolter, S. Bouix, N. Peinecke, M. S. Koo, and M. E. Shenton, "Global medical shape analysis using the Laplace-Beltrami spectrum," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 10, pp. 850-7, 2007.

[33] B. Gutman, W. Yalin, L. Lok Ming, T. F. Chan, and P. M. Thompson, "Hippocampal surface discrimination via invariant descriptors of spherical conformals maps," in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, 2007, pp. 1316-1319.

[34] J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson, "Automatic subcortical segmentation using a contextual model," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 11, pp. 194-201, 2008.

[35] H. Xiao, X. Chenyang, and J. L. Prince, "A topology preserving level set method for geometric deformable models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, pp. 755-768, 2003.

[36] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," presented at the Proceedings of the 24th annual conference on Computer graphics and interactive techniques, 1997.

[37] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, pp. 273-297, 1995.

[38] T. Joachims, "Training linear SVMs in linear time," presented at the Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia, PA, USA, 2006.

[39] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157-1182, 2003.

[40] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. Jack, Jr., "Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies," *Neuroimage*, vol. 39, pp. 1186-97, Feb 1 2008.

[41] J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, C. Avedissian, S. K. Madsen, N. Parikshak, X. Hua, A. W. Toga, C. R. Jack, Jr., M. W. Weiner, and P. M. Thompson, "Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls," *Neuroimage*, vol. 43, pp. 59-68, Oct 15 2008.

[42] S. Kloppel, C. M. Stonnington, J. Barnes, F. Chen, C. Chu, C. D. Good, I. Mader, L. A. Mitchell, A. C. Patel, C. C. Roberts, N. C. Fox, C. R. Jack, Jr., J. Ashburner, and R. S. Frackowiak, "Accuracy of dementia diagnosis: a direct comparison between radiologists

and a computerized method," *Brain : a journal of neurology,* vol. 131, pp. 2969-74, Nov 2008.

[43] D. Silverman and P. Thompson, "Structural and Functional Neuroimaging: Focussing on Mild Cognitive Impairment," *Applied Neurology,* vol. 2, pp. 10-22, 2006.

[44] J.-S. Kim, Y.-G. Kim, S.-M. Choi, and M.-H. Kim, "Morphometry of the hippocampus based on a deformable model and support vector machines," presented at the Proceedings of the 10th conference on Artificial Intelligence in Medicine, Aberdeen, UK, 2005.

[45] S. Li, F. Shi, F. Pu, X. Li, T. Jiang, S. Xie, and Y. Wang, "Hippocampal shape analysis of Alzheimer disease based on machine learning methods," *AJNR. American journal of neuroradiology,* vol. 28, pp. 1339-45, Aug 2007.

[46] R. H. Davies, C. J. Twining, P. D. Allen, T. F. Cootes, and C. J. Taylor, "Shape discrimination in the hippocampus using an MDL model," *Information processing in medical imaging : proceedings of the ... conference,* vol. 18, pp. 38-50, Jul 2003.

[47] K. Gorczowski, M. Styner, J. Ja-Yeon, J. S. Marron, J. Piven, H. C. Hazlett, S. M. Pizer, and G. Gerig, "Statistical Shape Analysis of Multi-Object Complexes," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on,* 2007, pp. 1-8.

[48] M. E. Shenton, G. Gerig, R. W. McCarley, G. Szekely, and R. Kikinis, "Amygdala-hippocampal shape differences in schizophrenia: the application of 3D shape models to volumetric MR data," *Psychiatry research,* vol. 115, pp. 15-35, Aug 20 2002.

[49] M. Chupin, E. Gerardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehericy, H. Benali, L. Garnero, O. Colliot, and A. s. D. N. Initi, "Fully Automatic Hippocampus Segmentation and Classification in Alzheimer's Disease and Mild Cognitive Impairment Applied on Data From ADNI," *Hippocampus,* vol. 19, pp. 579-587, 2009.

[50] L. Wang, J. S. Swank, I. E. Glick, M. H. Gado, M. I. Miller, J. C. Morris, and J. G. Csernansky, "Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging," *Neuroimage,* vol. 20, pp. 667-82, Oct 2003.

[51] S. M. Pizer, P. T. Fletcher, S. Joshi, A. G. Gash, J. Stough, A. Thall, G. Tracton, and E. L. Chaney, "A method and software for segmentation of anatomic object ensembles by deformable m-reps," *Med Phys,* vol. 32, pp. 1335-45, May 2005.

[52] Y. Wang, Y. Song, P. Rajagopalan, T. An, K. Liu, Y. Y. Chou, B. Gutman, A. W. Toga, and P. M. Thompson, "Surface-based TBM boosts power to detect disease effects on the brain: an N=804 ADNI study," *Neuroimage,* vol. 56, pp. 1993-2010, Jun 15 2011.

[53] Y. Shi, J. H. Morra, P. M. Thompson, and A. W. Toga, "Inverse-consistent surface mapping with Laplace-Beltrami eigen-features," *Information processing in medical imaging : proceedings of the ... conference,* vol. 21, pp. 467-78, 2009.

[54] S. Kurtek, E. Klassen, D. Zhaohua, S. W. Jacobson, J. B. Jacobson, M. J. Avison, and A. Srivastava, "Parameterization-Invariant Shape Comparisons of Anatomical Surfaces," *Medical Imaging, IEEE Transactions on,* vol. 30, pp. 849-858, 2011.

[55] I. Friedel, P. Schroeder, and M. Desbrun, "Unconstrained spherical parameterization," presented at the ACM SIGGRAPH 2005 Sketches, Los Angeles, California, 2005.

[56] Y. Wang, J. Zhang, B. Gutman, T. F. Chan, J. T. Becker, H. J. Aizenstein, O. L. Lopez, R. J. Tamburo, A. W. Toga, and P. M. Thompson, "Multivariate tensor-based morphometry on surfaces: application to mapping ventricular abnormalities in HIV/AIDS," *Neuroimage,* vol. 49, pp. 2141-57, Feb 1 2010.

[57] Y. Wang, T. F. Chan, A. W. Toga, and P. M. Thompson, "Multivariate tensor-based brain anatomical surface morphometry via holomorphic one-forms," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 12, pp. 337-44, 2009.

[58] J. Glaunes, M. Vaillant, and M. I. Miller, "Landmark matching via large deformation diffeomorphisms on the sphere," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 179-200, Jan-Mar 2004.

[59] J. Shi, P. M. Thompson, B. Gutman, and Y. Wang, "Surface fluid registration of conformal representation: Application to detect disease burden and genetic influence on hippocampus," *Neuroimage*, vol. 78, pp. 111-134, 2013.

[60] Lepore, N., A. Leow, and P. Thompson, "Landmark matching on the sphere using distance functions," in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, 2006, pp. 450-453.

[61] M. I. Miller and L. Younes, "Group Actions, Homeomorphisms, and Matching: A General Framework," *Int. J. Comput. Vision*, vol. 41, pp. 61-84, 2001.

[62] J. P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Medical image analysis*, vol. 2, pp. 243-60, Sep 1998.

[63] 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.

[64] M. Bro-Nielsen and C. Gramkow, "Fast Fluid Registration of Medical Images," presented at the Proceedings of the 4th International Conference on Visualization in Biomedical Computing, 1996.

[65] P. J. Kostelec, D. K. Maslen, D. M. Healy, and D. N. Rockmore, "Computational harmonic analysis for tensor fields on the two-sphere," *Journal of Computational Physics,* vol. 162, pp. 514-535, Aug 10 2000.

[66] R. Courant and D. Hilbert, *Methods of mathematical physics,* 1st English ed. New York,: Interscience Publishers, 1953.

[67] M. K. Chung, R. Hartley, K. M. Dalton, and R. J. Davidson, "Encoding Cortical Surface by Spherical Harmonics," *Statistica Sinica,* vol. 18, pp. 1269-1291, Oct 2008.

[68] P. Cachier and N. Ayache, "Isotropic Energies, Filters and Splines for Vector Field Regularization," *J. Math. Imaging Vis.,* vol. 20, pp. 251-265, 2004.

[69] B. Gutman, Y. Wang, L. M. Lui, T. F. Chan, P. M. Thompson, and A. W. Toga, "Shape Registration with Spherical Cross Correlation," *MICCAI Workshop on Mathematical Foundations in Computational Anatomy (MFCA '08),* 2008.

[70] L. Yuan, Y. Wang, P. M. Thompson, V. A. Narayan, and J. Ye, "Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data," *Neuroimage,* vol. 61, pp. 622-32, Jul 2 2012.

[71] X. Hua, D. P. Hibar, C. R. Ching, C. P. Boyle, P. Rajagopalan, B. A. Gutman, A. D. Leow, A. W. Toga, C. R. Jack, Jr., D. Harvey, M. W. Weiner, and P. M. Thompson, "Unbiased tensor-based morphometry: Improved robustness and sample size estimates for Alzheimer's disease clinical trials," *Neuroimage,* vol. 66C, pp. 648-661, Feb 1 2013.

[72] N. C. Fox, G. R. Ridgway, and J. M. Schott, "Algorithms, atrophy and Alzheimer's disease: cautionary tales for clinical trials," *Neuroimage,* vol. 57, pp. 15-8, Jul 1 2011.

[73] D. Holland, L. K. McEvoy, and A. M. Dale, "Unbiased comparison of sample size estimates from longitudinal structural measures in ADNI," *Hum Brain Mapp,* Aug 9 2011.

[74] J. M. Schott, J. W. Bartlett, J. Barnes, K. K. Leung, S. Ourselin, and N. C. Fox, "Reduced sample sizes for atrophy outcomes in Alzheimer's disease trials: baseline adjustment," *Neurobiol Aging,* vol. 31, pp. 1452-62, 1462 e1-2, Aug 2010.

[75] K. K. Leung, G. R. Ridgway, S. Ourselin, N. C. Fox, and A. s. D. Neuroimaging, "Consistent multi-time-point brain atrophy estimation from the boundary shift integral," *Neuroimage,* vol. 59, pp. 3995-4005, Feb 15 2012.

[76] D. Holland and A. M. Dale, "Nonlinear registration of longitudinal images and measurement of change in regions of interest," *Medical image analysis,* vol. 15, pp. 489-97, Aug 2011.

[77] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *Proc Natl Acad Sci U S A,* vol. 97, pp. 11050-5, Sep 26 2000.

[78] S. M. Smith, Y. Zhang, M. Jenkinson, J. Chen, P. M. Matthews, A. Federico, and N. De Stefano, "Accurate, robust, and automated longitudinal and cross-sectional brain change analysis," *Neuroimage,* vol. 17, pp. 479-89, Sep 2002.

[79] M. Reuter, N. J. Schmansky, H. D. Rosas, and B. Fischl, "Within-subject template estimation for unbiased longitudinal image analysis," *Neuroimage,* vol. 61, pp. 1402-1418, Jul 16 2012.

[80] L. A. Beckett, "Community-based studies of Alzheimer's disease: statistical challenges in design and analysis," *Stat Med,* vol. 19, pp. 1469-80, Jun 15-30 2000.

[81] A. D. Leow, I. Yanovsky, M. C. Chiang, A. D. Lee, A. D. Klunder, A. Lu, J. T. Becker, S. W. Davis, A. W. Toga, and P. M. Thompson, "Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration," *IEEE Trans Med Imaging,* vol. 26, pp. 822-32, Jun 2007.

[82] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Med Image Anal,* vol. 12, pp. 26-41, Feb 2008.

[83] X. Hua, S. Lee, I. Yanovsky, A. D. Leow, Y. Y. Chou, A. J. Ho, B. Gutman, A. W. Toga, C. R. Jack, Jr., M. A. Bernstein, E. M. Reiman, D. J. Harvey, J. Kornak, N. Schuff, G. E. Alexander, M. W. Weiner, and P. M. Thompson, "Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects," *Neuroimage,* vol. 48, pp. 668-81, Dec 2009.

[84] K. Chen, J. B. Langbaum, A. S. Fleisher, N. Ayutyanont, C. Reschke, W. Lee, X. Liu, D. Bandy, G. E. Alexander, P. M. Thompson, N. L. Foster, D. J. Harvey, M. J. de Leon, R. A. Koeppe, W. J. Jagust, M. W. Weiner, and E. M. Reiman, "Twelve-month metabolic declines in probable Alzheimer's disease and amnestic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the Alzheimer's Disease Neuroimaging Initiative," *Neuroimage,* vol. 51, pp. 654-64, Jun 2010.

[85] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification,* 2nd ed. New York: Wiley, 2001.

[86] O. Kohannim, X. Hua, D. P. Hibar, S. Lee, Y. Y. Chou, A. W. Toga, C. R. Jack, Jr., M. W. Weiner, and P. M. Thompson, "Boosting power for clinical trials using classifiers based on multiple biomarkers," *Neurobiol Aging,* vol. 31, pp. 1429-42, Aug 2010.

[87] S. Kloppel, A. Abdulkadir, C. R. Jack, Jr., N. Koutsouleris, J. Mourao-Miranda, and P. Vemuri, "Diagnostic neuroimaging across diseases," *Neuroimage,* vol. 61, pp. 457-63, Jun 2012.

[88] N. Z. Hobbs, S. M. D. Henley, G. R. Ridgway, E. J. Wild, R. A. Barker, R. I. Scahill, J. Barnes, N. C. Fox, and S. J. Tabrizi, "The progression of regional atrophy in premanifest and early Huntington's disease: a longitudinal voxel-based morphometry study," *Journal of Neurology Neurosurgery and Psychiatry,* vol. 81, pp. 756-763, Jul 2010.

[89] X. Hua, B. Gutman, C. P. Boyle, P. Rajagopalan, A. D. Leow, I. Yanovsky, A. R. Kumar, A. W. Toga, C. R. Jack, Jr., N. Schuff, G. E. Alexander, K. Chen, E. M. Reiman, M. W. Weiner, and P. M. Thompson, "Accurate measurement of brain changes in longitudinal MRI scans using tensor-based morphometry," *Neuroimage,* vol. 57, pp. 5-14, Jul 1 2011.

[90] W. K. Thompson and D. Holland, "Bias in tensor based morphometry Stat-ROI measures may result in unrealistic power estimates," *Neuroimage,* vol. 57, pp. 1-4, Jul 1 2011.

[91] B. A. Gutman, W. Yalin, P. Rajagopalan, A. W. Toga, and P. M. Thompson, "Shape matching with medial curves and 1-D group-wise registration," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, 2012, pp. 716-719.

[92] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology,* vol. 34, pp. 939-44, Jul 1984.

[93] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "The Alzheimer's disease neuroimaging initiative," *Neuroimaging Clin N Am,* vol. 15, pp. 869-77, xi-xii, Nov 2005.

[94] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. R. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI)," *Alzheimers Dement,* vol. 1, pp. 55-66, Jul 2005.

[95] C. R. Jack, Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," *J Magn Reson Imaging,* vol. 27, pp. 685-91, Apr 2008.

[96] J. Jovicich, S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, D. Kennedy, F. Schmitt, G. Brown, J. Macfall, B. Fischl, and A. Dale, "Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data," *Neuroimage,* vol. 30, pp. 436-43, Apr 1 2006.

[97] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE transactions on medical imaging,* vol. 17, pp. 87-97, Feb 1998.

[98] J. Gunter, M. Bernstein, B. Borowski, J. Felmlee, D. Blezek, R. Mallozzi, J. Levy, N. Schuff, and C. R. Jack Jr, "Validation testing of the MRI calibration phantom for the

Alzheimer's Disease Neuroimaging Initiative Study," *ISMRM 14th Scientific Meeting and Exhibition,* 2006.

[99] Y. Y. Chou, N. Lepore, G. I. de Zubicaray, O. T. Carmichael, J. T. Becker, A. W. Toga, and P. M. Thompson, "Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease," *Neuroimage,* vol. 40, pp. 615-30, Apr 1 2008.

[100] N. D. Cornea, D. Silver, X. S. Yuan, and R. Balasubramanian, "Computing hierarchical curve-skeletons of 3D objects," *Visual Computer,* vol. 21, pp. 945-955, Oct 2005.

[101] B. R. Ott, R. A. Cohen, A. Gongvatana, O. C. Okonkwo, C. E. Johanson, E. G. Stopa, J. E. Donahue, G. D. Silverberg, and I. Alzheimer's Disease Neuroimaging, "Brain ventricular volume and cerebrospinal fluid biomarkers of Alzheimer's disease," *Journal of Alzheimer's disease : JAD,* vol. 20, pp. 647-57, 2010.

[102] S. M. Nestor, R. Rupsingh, M. Borrie, M. Smith, V. Accomazzi, J. L. Wells, J. Fogarty, and R. Bartha, "Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database," *Brain : a journal of neurology,* vol. 131, pp. 2443-54, Sep 2008.

[103] Y. Y. Chou, N. Lepore, C. Avedissian, S. K. Madsen, N. Parikshak, X. Hua, L. M. Shaw, J. Q. Trojanowski, M. W. Weiner, A. W. Toga, and P. M. Thompson, "Mapping correlations between ventricular expansion and CSF amyloid and tau biomarkers in 240 subjects with Alzheimer's disease, mild cognitive impairment and elderly controls," *Neuroimage,* vol. 46, pp. 394-410, Jun 2009.

[104] O. T. Carmichael, L. H. Kuller, O. L. Lopez, P. M. Thompson, R. A. Dutton, A. Lu, S. E. Lee, J. Y. Lee, H. J. Aizenstein, C. C. Meltzer, Y. Liu, A. W. Toga, and J. T. Becker,

"Ventricular volume and dementia progression in the Cardiovascular Health Study," *Neurobiology of aging,* vol. 28, pp. 389-97, Mar 2007.

[105] M. W. Weiner, "Expanding ventricles may detect preclinical Alzheimer disease," *Neurology,* vol. 70, pp. 824-5, Mar 11 2008.

[106] T. J. DiCicio and B. Efron, "Bootstrap Confidence Intervals," *Statistical Science,* vol. 11, p. 10, 1996.

[107] P. Golland, W. E. L. Grimson, and R. Kikinis, "Statistical shape analysis using fixed topology skeletons: Corpus callosum study," *Information Processing in Medical Imaging, Proceedings,* vol. 1613, pp. 382-387, 1999.

[108] D. A. Alcantara, O. Carmichael, W. Harcourt-Smith, K. Sterner, S. R. Frost, R. Dutton, P. Thompson, E. Delson, and N. Amenta, "Exploration of Shape Variation Using Localized Components Analysis," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 31, pp. 1510-1516, Aug 2009.

[109] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of Computational and Graphical Statistics,* vol. 15, pp. 265-286, Jun 2006.

[110] O. Carmichael, J. Xie, E. Fletcher, B. Singh, C. DeCarli, and A. s. D. Neuroimaging, "Localized hippocampus measures are associated with Alzheimer pathology and cognition independent of total hippocampal volume," *Neurobiology of Aging,* vol. 33, Jun 2012.

[111] W. Yang, X. Chen, X. Hong, and X. Huang, "ICA-Based Automatic Classification of Magnetic Resonance Images from ADNI Data," *Proceedings of the 2010 international conference on Life system modeling and simulation and intelligent computing, and 2010*

international conference on Intelligent computing for sustainable energy and environment: Part III, p. 8, 2010.

[112] Y. Fan, N. Batmanghelich, C. M. Clark, and C. Davatzikos, "Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline," *Neuroimage*, vol. 39, pp. 1731-43, Feb 15 2008.

[113] C. D. Good, R. I. Scahill, N. C. Fox, J. Ashburner, K. J. Friston, D. Chan, W. R. Crum, M. N. Rossor, and R. S. Frackowiak, "Automatic differentiation of anatomical patterns in the human brain: validation with studies of degenerative dementias," *Neuroimage*, vol. 17, pp. 29-46, Sep 2002.

[114] R. Cuingnet, M. Chupin, H. Benali, and O. Colliot, "Spatial prior in SVM-based classification of brain images," *Proc. SPIE 7624, Medical Imaging 2010: Computer-Aided Diagnosis*, vol. 7624, 2010.

[115] B. Gutman, Y. Wang, J. Morra, A. W. Toga, and P. M. Thompson, "Disease classification with hippocampal shape invariants," *Hippocampus*, vol. 19, pp. 572-8, Jun 2009.

[116] Y. Cho, J. K. Seong, Y. Jeong, and S. Y. Shin, "Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data," *Neuroimage*, vol. 59, pp. 2217-30, Feb 1 2012.

[117] E. Gerardin, G. Chetelat, M. Chupin, R. Cuingnet, B. Desgranges, H. S. Kim, M. Niethammer, B. Dubois, S. Lehericy, L. Garnero, F. Eustache, O. Colliot, and A. s. D. N. Initi, "Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging," *Neuroimage*, vol. 47, pp. 1476-1486, Oct 1 2009.

[118] M. Styner, J. A. Lieberman, R. K. McClure, D. R. Weinberger, D. W. Jones, and G. Gerig, "Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, pp. 4872-7, Mar 29 2005.

[119] L. Shen, Y. Qi, S. Kim, K. Nho, J. Wan, S. L. Risacher, and A. J. Saykin, "Sparse bayesian learning for identifying imaging biomarkers in AD prediction," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention,* vol. 13, pp. 611-8, 2010.

[120] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *Neuroimage,* vol. 55, pp. 856-67, Apr 1 2011.

[121] A. A. Qazi, D. R. Jorgensen, M. Lillholm, M. Loog, M. Nielsen, and E. B. Dam, "A framework for optimizing measurement weight maps to minimize the required sample size," *Medical Image Analysis,* vol. 14, pp. 255-264, Jun 2010.

[122] L. Ferrarini, W. M. Palm, H. Olofsen, M. A. van Buchem, J. H. Reiber, and F. Admiraal-Behloul, "Shape differences of the brain ventricles in Alzheimer's disease," *Neuroimage,* vol. 32, pp. 1060-9, Sep 2006.

[123] K. S. Cover, R. A. van Schijndel, B. W. van Dijk, A. Redolfi, D. L. Knol, G. B. Frisoni, F. Barkhof, and H. Vrenken, "Assessing the reproducibility of the SienaX and Siena brain atrophy measures using the ADNI back-to-back MP-RAGE MRI scans," *Psychiatry research,* vol. 193, pp. 182-90, Sep 30 2011.

[124] P. A. Yushkevich, B. B. Avants, S. R. Das, J. Pluta, M. Altinay, and C. Craige, "Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from

asymmetric global normalization: an illustration in ADNI 3 T MRI data," *Neuroimage,* vol. 50, pp. 434-45, Apr 1 2010.

[125] K. Zhang and T. J. Sejnowski, "A universal scaling law between gray matter and white matter of cerebral cortex," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 97, pp. 5621-5626, May 9 2000.

[126] C. C. Brun, N. Lepore, E. Luders, Y. Y. Chou, S. K. Madsen, A. W. Toga, and P. M. Thompson, "Sex differences in brain structure in auditory and cingulate regions," *Neuroreport,* vol. 20, pp. 930-5, Jul 1 2009.

[127] H. D. Tagare, D. Groisser, and O. Skrinjar, "Symmetric Non-rigid Registration: A Geometric Theory and Some Numerical Techniques," *Journal of Mathematical Imaging and Vision,* vol. 34, pp. 61-88, 2009.

[128] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Boosting linear discriminant analysis for face recognition," *ICIP,* pp. 657-660, 2003.

[129] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, pp. 1168-72, Jan 31 2006.

[130] A. M. Bronstein, M. M. Bronstein, R. Kimmel, M. Mahmoudi, and G. Sapiro, "A Gromov-Hausdorff Framework with Diffusion Geometry for Topologically-Robust Non-rigid Shape Matching," *Int. J. Comput. Vision,* vol. 89, pp. 266-286, 2010.

[131] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models\—their training and application," *Comput. Vis. Image Underst.,* vol. 61, pp. 38-59, 1995.

[132] G. Zou, J. Hua, and O. Muzik, "Non-rigid surface registration using spherical thin-plate splines," presented at the Proceedings of the 10th international conference on Medical image computing and computer-assisted intervention - Volume Part I, Brisbane, Australia, 2007.

[133] M. K. Chung, "Heat kernel smoothing on unit sphere," in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, 2006, pp. 992-995.

[134] B. Huhle, T. Schairer, and W. Strasser, "Normalized Cross-Correlation using SOFT," in *Local and Non-Local Approximation in Image Processing, 2009. LNLA 2009. International Workshop on*, 2009, pp. 82-86.

[135] L. Sorgi and K. Daniilidis, "Normalized Cross-Correlation for Spherical Images," in *Computer Vision - ECCV 2004*. vol. 3022, T. Pajdla and J. Matas, Eds., ed: Springer Berlin Heidelberg, 2004, pp. 542-553.