



HAL
open science

Atlas-based shape analysis and classification of retinal optical coherence tomography images using the functional shape (fshape) framework

Sieun Lee, Nicolas Charon, Benjamin Charlier, Karteek Popuri, Evgeniy Lebed, Marinko V. Sarunic, Alain Trouvé, Mirza Faisal Beg

► To cite this version:

Sieun Lee, Nicolas Charon, Benjamin Charlier, Karteek Popuri, Evgeniy Lebed, et al.. Atlas-based shape analysis and classification of retinal optical coherence tomography images using the functional shape (fshape) framework. *Medical Image Analysis*, 2017, 35, pp.570 - 581. 10.1016/j.media.2016.08.012 . hal-01817551

HAL Id: hal-01817551

<https://hal.science/hal-01817551>

Submitted on 18 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Atlas-based Shape Analysis and Classification of Retinal Optical Coherence Tomography Images using the Functional Shape (fshape) Framework

Sieun Lee^a, Nicolas Charon^b, Benjamin Charlier^c, Karteek Popuri^a, Evgeniy Lebed^a, Marinko V. Sarunic^a, Alain Trounev^d, Mirza Faisal Beg^{a,*}

^a*School of Engineering Science, Simon Fraser University,
Burnaby, British Columbia, V5A 1S6 Canada*

^b*Center for Imaging Sciences (CIS), Johns Hopkins University, Baltimore, MD 21218, USA*

^c*Institut Montpellierain Alexander Grothendieck, Campus Triolet,
34095 Montpellier, France*

^d*CMLA, ENS Cachan, CNRS,
Université Paris-Saclay, 94235 Cachan, France*

Abstract

We propose a novel approach for quantitative shape variability analysis in retinal optical coherence tomography images using the functional shape (fshape) framework. The fshape framework uses surface geometry together with functional measures, such as retinal layer thickness defined on the layer surface, for registration across anatomical shapes. This is used to generate a population mean template of the geometry-function measures from each individual. Shape variability across multiple retinas can be measured by the geometrical deformation and functional residual between the template and each of the observations. To demonstrate the clinical relevance and application of the framework, we generated atlases of the inner layer surface and layer thickness of the Retinal Nerve Fiber Layer (RNFL) of glaucomatous and normal subjects, visualizing detailed spatial pattern of RNFL loss in glaucoma. Additionally, a regularized linear discriminant analysis classifier was used to automatically classify glaucoma, glaucoma-suspect, and control cases based on RNFL fshape metrics.

Keywords: Atlas-based shape analysis, Atlas generation, Computational

*Corresponding author

Email address: mfbeg@sfu.ca (Mirza Faisal Beg)

1. Introduction

In the past two decades, optical coherence tomography (OCT) has been widely adopted in ophthalmology for noninvasive, *in-vivo*, micrometer-resolution imaging of the anterior and posterior segments of the eye, bringing new insights in optic diseases along with more informed clinical decision making. Since the seminal paper in 1991 (Huang et al., 1991), significant progresses in OCT imaging have made possible highly detailed 3D images. Fully utilizing the information in such volumetric data, especially in large population studies, requires robust, quantitative measures of variability in the images. The conventional approach is to define an anatomical parameter, measure it in each image, and compare the values. Examples of OCT image parameters include optic cup and disk measurements and thickness of retinal layers (Medeiros et al., 2005; González-García et al., 2009). However, the spatial and anatomical correspondence across measurements from multiple images is often limited. This can be due to ambiguity in the parameter definition, such as in the optic cup-to-disc ratio, where the boundaries of the cup and disc often do not correspond to a single anatomical structure (Chauhan and Burgoyne, 2013; Young et al., 2014). Poor intra- and inter-subject correspondence can be also due to a lack of clear spatial references, as in the case of retinal layer thickness maps.

In the latter case, comparing or computing a group average from multiple OCT images requires artificially defined regional correspondence. This often relies on gross measures of orientation and distance from an anatomical landmark. In the common sectoral retinal layer thickness analysis of OCT scans, sectors are delimited by superior, inferior, temporal, and nasal orientations, and distance from the foveal pit or optic disc. Such an approach uses limited information, and leaves questions such as how comparable measurements from two eye would be if one eye's optic disc was larger or more elliptical than the other. Averaging over a region may mitigate some of the variability due to unknown

mismatch, but comes at the cost of spatial sensitivity of the measurement.

30 One way to address the issue is to define parameters that are more anatomically consistent. Several groups (Strouthidis et al., 2009; Chauhan and Burgoyne, 2013; Young et al., 2014), suggested Bruch’s Membrane Opening (BMO) based parameters as more reliable structural measurements than the cup-to-disk ratio. Peripapillary sector shapes were adjusted to better align to individual
35 BMO dimensions (Lee et al., 2014), and foveal-BMO axis has been shown to be a better reference for angular orientation than image frame in its anatomical and physiological justification and measurement repeatability (Chauhan and Burgoyne, 2013; He et al., 2014).

Registration of OCT images fundamentally addresses the correspondence
40 issue; however, most of its application has been in averaging multiple images for noise reduction and motion correction (Jørgensen et al., 2007; Young et al., 2011), or rigid alignment of time-course images (Niemeijer et al., 2009). Three previous studies have specifically examined nonrigid registration of OCT data from multiple subjects in the context of shape variability analysis. In Gibson
45 et al. (Gibson et al., 2010), segmented optic cups were registered to a single template optic cup, first by rigid and nonrigid intensity-based volumetric registration followed by spherical mapping and demons algorithm (Yeo et al., 2010). In Chen et al. (Chen et al., 2014), registration of macular OCT scans used rigid alignment of foveae, affine registration of A-scans to match the inner and
50 outer retinal boundaries, and smooth deformation of A-scans using radial basis functions for refined alignment of the retinal layers. Lee et al. (Lee et al., 2015) extended upon (Gibson et al., 2010) and represented segmented retinal surfaces utilizing the framework of mathematical currents. Two surfaces were brought into proximity by minimizing a functional of reproducing kernel Hilbert
55 space (RKHS) norm-based energy and a dissimilarity term. This was followed by spherical demons registration to establish point-to-point correspondence. In above works, target subjects were registered to one subject as the template, which leaves the problem of the bias in template choice.

In this report, we propose to exploit a novel approach to analyzing variability

60 in retinal OCT images using the *functional shape* or fshape framework. Fshapes and some of its abstract mathematical properties were introduced in (Charlier et al., 2015). The fshape framework considers a geometrical surface, such as a retinal layer surface, and functions defined on the surface, such as retinal layer thickness, together as a single object. The “distance”, *i.e.* the shape
65 difference between two retinas is measured by both geometry and one or more physiological or morphological signals defined on the geometrical surface. This is a conceptual departure from previous works in which the goal is to first establish anatomical correspondence between two eyes and then to compare the functional values at resultant corresponding locations, and the first work on
70 atlas-based morphological analysis of the retina. There are several advantages to the fshape approach. First, the mathematical abstraction allows variability analysis, or comparison, beyond the frame of the anatomy; given any one or more surfaces and functional values, the fshape framework can measure the inter-subject or inter-time points difference. This allows for combination of any
75 number of features - for example, Inner Limiting Membrane (ILM) geometry and total retinal thickness, Bruch’s membrane (BM) geometry with RNFL thickness, or ILM and BM surface geometry - and to investigate individually and jointly which features are more or less significant in differentiating between a disease and control group. The core of the fshape framework is generation of a mean
80 template of multiple fshapes. An initial template or hypertemplate is taken as a simple model of a prototype fshape. This template is evolved through an optimization process that simultaneously minimizes i) geometric-functional distance from the current template to the observations, and ii) dissimilarity between the transformed mean template and the observations. This approach
85 eliminates the need to choose one of the existing data as a template, mitigating the issue of template selection and bias. In addition, the fshape framework registration does not rely on specific anatomical or image features based on prior knowledge. Given decent quality segmentation and measurements, the generality and versatility of the algorithm allows it to be applied broadly and
90 robustly.

The fshape framework combines a distance of fshape transformations within the fshape bundle of a template and dissimilarity measures between arbitrary fshapes based on an extension of varifold spaces and norms. By optimization of these metrics, a mean atlas can be created, and serve as the reference to all
95 fshapes from the dataset. In addition, the algorithm simultaneously provides the variations in shape and signal of this template to each observation. These outputs then constitute the basis of inter-subject analysis and group classification through a wide choice of possible statistical analysis tools.

This paper proposes a very first set of applications of the fshape methodology to the OCT dataset while extending the scope of the framework to statistical learning and classification. Namely, after a short application-oriented presentation of fshapes, we propose a formal Bayesian derivation of the template estimation problem studied in (Charlier et al., 2015) and augment the current approach with a statistical analysis module for the estimated geometric
100 and functional features. To demonstrate the clinical relevance and application of the algorithm, we generated a mean atlas and performed automated classification of glaucoma eyes and healthy control eyes based on both geometry and thickness of posterior RNFL surface and RNFL thickness. We also present group averages and t -test of the RNFL thickness over the mean surface between
105 the healthy and glaucomatous groups to emphasize how the fshape approach can reveal morphological and functional patterns in cross-sectional or longitudinal data.

2. Methods

2.1. Image acquisition and processing

115 The OCT images in this study were acquired at the Eye Care Centre at Vancouver General Hospital in Vancouver, British Columbia, Canada, using a custom prototype swept-source OCT machine developed at Simon Fraser University (SFU) with a 1060-nm wavelength light source. Three-dimensional volumetric images centered at optic nerve head were acquired over a 5-8 mm²

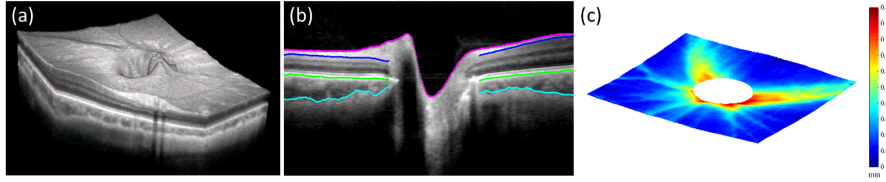


Figure 1: (a) 3D visualization of the ONH scan, (b) segmentation of ILM (magenta), posterior RNFL boundary (blue), BM (green), and posterior choroidal boundary (cyan) of a smoothed volume, (c) a posterior RNFL boundary surface color mapped with RNFL thickness.

120 region with 2.8 mm in depth. The image resolution was approximately $6 \mu\text{m}$ in the axial direction and $12\text{-}21 \mu\text{m}$ in the lateral direction, depending on the subject eye's axial length. Each volume consisted of $1024 \times 400 \times 400$ voxels, with 1024 voxels in axial direction. The images were corrected for axial motion using cross-correlation of adjacent frames. A bounded variation regularization

125 method was used to reduce the effect of speckles and enhance retinal boundaries for segmentation and visualization. Retinal layers were segmented automatically in 3D using a graph-cut based surface segmentation algorithm implemented in MATLAB (Li et al., 2006; Garvin et al., 2008; Lee et al., 2013). In this study four boundaries were segmented for two layers: RNFL and choroid. Left (OS)

130 eyes were flipped so that all eyes were in the right (OD) eye orientation. Layer thickness was measured at each point of the layer's posterior surface as the closest distance to the anterior surface. In addition to the automated segmentation, BMO in each eye was segmented manually, and a best-fit ellipse was generated by Principal Component Analysis (PCA). The BMO ellipse was used to crop

135 the surfaces near the BMO where the layers terminate. Fig. 1 shows an example optic nerve head (ONH) scan, segmentation, and the resulting posterior RNFL surface with RNFL thickness mapped on the surface.

2.2. Atlas estimation

The central goal of the fshape framework is to recover the inter-subject variability both in the geometry of the retina surfaces as well as in their signals (i.e

140 the thickness maps). Following the standard process in computational anatomy

(Joshi et al., 2004; Ma et al., 2010; Zhang et al., 2013), the primary step is to estimate an *atlas* from the population, which serves as a *template* object with the variation of the subjects in the population. The obvious difficulty in this situation is that both geometrical supports and functional maps on the supports vary concurrently. The notion of *functional shape* or *fshape* (Charlier et al., 2015) has the crucial advantage of treating both geometry and function together while providing important flexibility for the atlas generation and registration algorithms. The following sections give a compact and application-oriented exposition of the theoretical presentation presented in (Charlier et al., 2015) and emphasizes the Bayesian interpretation of atlas estimation in the fshapes context.

2.2.1. Background: fshapes and fshape spaces

In the fshape framework, a geometrical structure and its associated scalar field are considered as a single object and processed *jointly*. Thus, in general, an fshape consists of a pair (X, f) where X is the geometrical support, i.e a surface in the 3D ambient space, and f is a function defined on this surface; in our particular application, the pair of a retinal layer surface and a retinal layer thickness mapped on the surface.

Transformation of an fshape (X, f) can be modeled by combining a deformation ϕ of the geometrical support and a signal change ζ . In the simplest setting considered in (Charlier et al., 2015), ϕ is a diffeomorphism of \mathbb{R}^3 and ζ a residual function on X . The combination (ϕ, ζ) then 'acts' on the fshape (X, f) as

$$(\phi, \zeta) \cdot (X, f) = (\phi(X), (f + \zeta) \circ \phi^{-1}) \quad (1)$$

meaning that the surface is transported by deformation ϕ while signal is modified by adding the residual ζ and then mapped onto the deformed surface $\phi(X)$.

Quantifying such transformations is done first by introducing a model of deformation group, our reference model in this paper is the widely studied Large Diffeomorphic Metric Mapping (LDDMM) framework of (Beg et al., 2005), in which diffeomorphisms are constructed as the flow of time-dependent velocity fields $v \in L^2([0, 1], V)$ with V a reproducing kernel Hilbert space (RKHS) of

smooth velocity fields of \mathbb{R}^3 . On the other hand, we shall consider L^2 signals on surfaces, i.e $f, \zeta \in L^2(X)$ with the corresponding surface 2–norm on the shape X . With $\phi = \phi_1^v$ the flow of v up to time $t = 1$, the energy of the fshape transformation (ϕ, ζ) of (X, f) considered in [Charlier et al. \(2015\)](#) is given by:

$$E_X(v, \zeta)^2 = \frac{1}{2\gamma_V^2} \int_0^1 \|v_t\|_V^2 dt + \frac{1}{2\gamma_f^2} \|\zeta\|_{L^2(X)}^2 \quad (2)$$

where σ_V, σ_f are weighting parameters between the geometric and functional energies.

The energy in (2) gives a way to measure a notion of distance between two fshapes but only in the situation where these can be *exactly* mapped on each other in the transformation model introduced above. In practice, this is unrealistic or not desirable since datasets present inter-subject variability caused by noisy irregularity in the shape and signals, variations that are not well represented by the smooth geometric-functional transformations of the previous section.

In registration problems, it is thus common to introduce additional dissimilarity (or data fidelity) terms to the deformation cost, which may be interpreted equivalently as a noise model on the observations as we shall see in the next section. In the situation of functional shapes, unlike usual images, the absence of any explicit correspondence between their geometrical supports does not enable direct comparison of their signals.

A possibility to overcome this issue was examined thoroughly in ([Charon and Trouvé, 2013](#); [Charlier et al., 2015](#)) and consisted in extending previous works on curves and surfaces ([Glaunès et al., 2004](#); [Charon and Trouvé, 2013](#)). In the fvarifold framework that we very briefly sum up, an fshape (X, f) is represented as a distribution on the product space $\mathbb{R}^3 \times P(\mathbb{R}^3) \times \mathbb{R}$, with $P(\mathbb{R}^3)$ being the projective space of all lines in \mathbb{R}^3 . This distribution, written $\mu_{(X, f)}$ is formally a sum of Diracs involving the position of the shape points $x \in X$, the attached (unoriented) direction of the normal vector $\overleftrightarrow{n}(x) \in P(\mathbb{R}^3)$ and signal function $f(x)$.

These distributions are compared via metrics derived from tensor product of

positive kernels of the form $k_g \otimes k_n \otimes k_f$ on $\mathbb{R}^3 \times P(\mathbb{R}^3) \times \mathbb{R}$. The induced Hilbert (pseudo-)metric between any two fshapes (X, f) and (Y, g) writes explicitly:

$$\langle \mu_{(X,f)}, \mu_{(Y,g)} \rangle_{W^*} = \iint_{X \times Y} k_g(x, y) k_n(\overleftarrow{n}_X(x), \overleftarrow{n}_Y(y)) k_f(f(x), g(y)) d\sigma(x) d\sigma(y) \quad (3)$$

185 where $d\sigma$ denotes the surface area measures of X and Y respectively. Then $A \doteq \|\mu_{(X,f)} - \mu_{(Y,g)}\|_{W^*}^2 = \langle \mu_{(X,f)} - \mu_{(Y,g)}, \mu_{(X,f)} - \mu_{(Y,g)} \rangle_{W^*}$ can be taken as a suitable dissimilarity term between (X, f) and (Y, g) since it enforces proximity in the respective geometries and signal functions concurrently.

2.2.2. Heuristics of atlas estimation

Let us now consider a population of N observations $(X^i, f^i)_{i=1, \dots, N}$ where the X^i 's are the retinal surfaces and f^i the corresponding thickness maps. We introduce a *forward generative model* in the line of (Durrleman et al., 2008) for which observations are noisy geometric-functional transformations of a common unknown *template* fshape (X, f) plus additional noise terms:

$$(X^i, f^i) = (\phi^i, \zeta^i) \cdot (X, f) + \varepsilon^i, \quad \text{for all } i = 1, \dots, N. \quad (4)$$

190 Above, ϕ^i is the flow of a vector field $v^i \in L^2([0, 1], V)$ and (v^i, ζ^i) are regarded as hidden latent variables of the transformations from template to subjects, ε^i 's are noise variables.

Considering i.i.d. variables ε^i , we may define a noise model on fshapes based on fvarifold metrics:

$$p(\varepsilon^i) = p((X^i, f^i) | (X, f, v^i, \zeta^i)) \propto e^{-\frac{\|\mu_{(\phi^i, \zeta^i) \cdot (X, f)} - \mu_{(X^i, f^i)}\|_{W^*}^2}{2\gamma_W^2}}$$

which is a Gaussian model with respect to the metric $\|\cdot\|_{W^*}$. Note that this is only formal for the infinite dimensional space of fvarifolds but can be given a rigorous sense if restricted to a predefined discrete grid, similarly to (Gori et al., 2013).

As for the latent variables (v^i, ζ^i) , we take the following prior deriving from the energy (2):

$$p((v^i, \zeta^i) | (X, f)) \propto e^{-E_X(v^i, \zeta^i)^2}$$

which is essentially assuming independent (formal) Gaussian distribution on v and ζ in their respective metric spaces.

Finally, we also model the template (X, f) as a random variable itself. Inspired from the *hypertemplate* model for shape atlases of (Ma et al., 2010), we represent (X, f) as a transformation of a given hypertemplate fshape (X^0, f^0) , i.e $(X, f) = (\phi^0, \zeta^0) \cdot (X^0, \zeta^0)$ for a deformation ϕ^0 and a residual $\zeta^0 \in L^2(X^0)$. As previously, the prior on the template is:

$$p((v^0, \zeta^0)) \propto e^{-E_{X^0}(v^0, \zeta^0)^2}.$$

With an hypertemplate (X_0, f_0) fixed by the user, estimating the template then amounts to computing the maximum a posteriori (MAP) estimate of (v^0, ζ^0) knowing the observations (X^i, f^i) . With Bayes rules this leads to minimizing:

$$-\sum_{i=1}^N \log \left(\int p(X^i, f^i | v^0, \zeta^0, v^i, \zeta^i) p(v^i, \zeta^i) \right) - \log(p(v^0, \zeta^0)).$$

The first term involves the integral with respect to the probability distribution of the latent variables (v^i, ζ^i) . As there is no closed form expression of this integral, we use the standard Fast Approximation with Modes and replace it by $\max_{v^i, \zeta^i} p(X^i, f^i | v^0, \zeta^0, v^i, \zeta^i) p(v^i, \zeta^i)$ leading eventually to the following variational problem:

$$\begin{aligned} (v_*^0, \zeta_*^0, (\phi_*^i)_i, (\zeta_*^i)_i) &= \operatorname{arginf}_{v^0, \zeta^0, (v^i, \zeta^i)_i} \frac{1}{2\gamma_V^2} \int_0^1 \|v_t^0\|_V^2 dt + \frac{1}{2\gamma_f^2} \|\zeta^0\|_{L^2(X^0)}^2 \\ &+ \sum_{i=1}^N \left(\frac{1}{2\gamma_V^2} \int_0^1 \|v_t^i\|_V^2 dt + \frac{1}{2\gamma_f^2} \|\zeta^i\|_{L^2(X^0)}^2 + \frac{1}{2\gamma_W^2} \|\mu_{(\phi^i, \zeta^i) \cdot (X, f)} - \mu_{(X^i, f^i)}\|_{W^*}^2 \right) \end{aligned} \quad (5)$$

where we remind that for $i = 0, \dots, N$ deformations ϕ^i are the flows of the v^i and $X = \phi^0(X_0)$.

The previous paragraphs underline the Bayesian interpretation behind variational problem (5), for which the existence of solutions under certain conditions was addressed in (Charlier et al., 2015). Note that the variances $\gamma_V^2, \gamma_f^2, \gamma_W^2$ act

as weighting coefficients between the different terms. In the rest of the paper,
 205 we will consider these as parameters of the algorithm. However, we wish to
 point out that with a proper discretization of the previous probability densities,
 this could be improved by also estimating the optimal variances, extending ap-
 proaches developed for images in [Zhang et al. \(2013\)](#) and surfaces in [Ma et al.
 \(2010\)](#); [Gori et al. \(2013\)](#).

210 *2.2.3. Discrete fshapes*

In the discrete setting, a functional surface is a textured mesh given as a
 $P \times 3$ matrix \mathbf{x} of coordinates of the P vertices x_k in \mathbb{R}^3 , a $P \times 1$ column vector
 \mathbf{f} of the P values $f_k \doteq f(x_k) \in \mathbb{R}$ of the signal at the vertices and a $T \times 3$
 connectivity matrix \mathbf{C} (a triangulation for simplicity). The L^2 norm of f on
 the surface is here approximated with P0 finite elements

$$\|f\|_{L^2(X)}^2 \approx \sum_{\ell=1}^T \bar{f}_\ell^2 |T_\ell| \quad (6)$$

where for the ℓ^{th} triangle, $|T_\ell|$ is the area of the triangle and \bar{f}_ℓ is the average
 of the three signal values at its vertices.

Optimal deformation fields minimize the metric $\int_0^1 \|v_t\|_V^2 dt$ for given final
 time conditions and are therefore geodesics in the context of diffeomorphism
 groups. It has been shown ([Beg et al., 2005](#); [Arguillere et al., 2015](#)) that such
 geodesic flows are governed by Hamiltonian equations. In the present case of
 discrete set of particles, geodesics are parametrized by initial momenta $\mathbf{p} =$
 $(p_k) \in \mathbb{R}^{P \times 3}$ attached to every vertex x_k and the following *shooting equations*:

$$\begin{cases} \dot{x}_k(t) = v_t(x_k(t)) = \sum_{l=1}^P K_V(x_l(t), x_k(t)) p_l(t) \\ \dot{p}_k(t) = - \sum_{l=1}^P p_l(t) \cdot p_l(t) \partial_1 K_V(x_l(t), x_k(t)) \end{cases} \quad (7)$$

given the initial conditions $\mathbf{x}(0) = \mathbf{x}$ and $\mathbf{p}(0) = \mathbf{p}$ and where K_V is the 3×3
 matrix-valued kernel function associated to the RKHS V .

We thus parametrize geometric-functional transformations of the discrete
 fshape $(\mathbf{x}, \mathbf{f}, \mathbf{C})$ by the momenta vectors \mathbf{p} and a residual discrete signal $\boldsymbol{\zeta}$:

$$(\phi, \boldsymbol{\zeta}) \cdot (\mathbf{x}, \mathbf{f}) = ((x_k(1))_{1 \leq k \leq P}, (f_k + \zeta_k)_{1 \leq k \leq P}). \quad (8)$$

As for the fvarifold-norm fidelity term of (3), we use a similar discrete approximation where each face triangle T_ℓ is approximated by a single Dirac with position \bar{x}_ℓ the vertices' barycenter, \bar{f}_ℓ the mean signal value, and \overleftarrow{n}_ℓ the unit unoriented normal vector to the triangle. With two discrete fshapes $(\mathbf{x}, \mathbf{f}, \mathbf{C}_1)$ and $(\mathbf{y}, \mathbf{g}, \mathbf{C}_2)$,

$$\langle \mu_{(\mathbf{x}, \mathbf{f})}, \mu_{(\mathbf{y}, \mathbf{g})} \rangle_{W^*} = \sum_{k=1}^{T_1} \sum_{\ell=1}^{T_2} |T_k^{\mathbf{x}}| |T_\ell^{\mathbf{y}}| k_g(\bar{x}_k, \bar{y}_\ell) k_n(\overleftarrow{n}_k^x, \overleftarrow{n}_\ell^y) k_f(\bar{f}_k, \bar{g}_\ell). \quad (9)$$

2.2.4. Numerical scheme for atlas estimation

With the above discrete model and a given discrete hypertemplate $(\mathbf{x}^0, \mathbf{f}^0, \mathbf{C}^0)$ the atlas estimation problem of (5) reduces to the minimization of the function:

$$\begin{aligned} \mathbf{J}(\mathbf{p}^0, \boldsymbol{\zeta}^0, (\mathbf{p}^i)_i, (\boldsymbol{\zeta}^i)_i) &\doteq \frac{1}{2\gamma_V^2} (\mathbf{p}^0)^T K_V(\mathbf{x}^0, \mathbf{x}^0) \mathbf{p}^0 + \frac{1}{2\gamma_f^2} \|\boldsymbol{\zeta}^0\|_{L^2}^2 \\ &+ \sum_{i=1}^N \left(\frac{1}{2\gamma_V^2} (\mathbf{p}^i)^* K_V(\mathbf{x}, \mathbf{x}) \mathbf{p}^i + \frac{1}{2\gamma_f^2} \|\boldsymbol{\zeta}^i\|_{L^2}^2 + \frac{1}{2\gamma_W^2} \|\mu_{(\tilde{\mathbf{x}}^i, \tilde{\mathbf{f}}^i)} - \mu_{(\mathbf{x}^i, \mathbf{f}^i)}\|_{W^*}^2 \right) \end{aligned} \quad (10)$$

where $\|\cdot\|_{L^2}$ is given by (6) and the deformed fshapes $(\mathbf{x}, \mathbf{f}) \doteq (\phi^0, \boldsymbol{\zeta}^0) \cdot (\mathbf{x}^0, \mathbf{f}^0)$, $(\tilde{\mathbf{x}}^i, \tilde{\mathbf{f}}^i) \doteq (\phi^i, \boldsymbol{\zeta}^i) \cdot (\mathbf{x}, \mathbf{f})$ are obtained by the shooting equations (7). We will write $(\mathbf{p}_*, \boldsymbol{\zeta}_*, (\mathbf{p}_*^i)_i, (\boldsymbol{\zeta}_*^i)_i)$ for a minimizer of \mathbf{J} . The overall principle is illustrated by Figure 2.

This is now a finite yet very high-dimensional optimization problem since each variable is roughly of dimension P , the number of hypertemplate vertices (which is typically of the order of $P \approx 5000$ in the simulations of this paper) and it is also non-convex. We thus implement a Polak-Ribiere conjugate gradient descent where all variables are simultaneously updated at each step. The algorithm is also coupled with basic line search for each variable to increase the speed of convergence. The gradients with respect to signal variables $\boldsymbol{\zeta}^0$ and $(\boldsymbol{\zeta}^i)$ are easily computable from the expression in (6) and (9). Gradients with respect to initial momenta variables $\mathbf{p}^0, (\mathbf{p}^i)$ are computed based on forward-backward shooting procedure as in usual geometric registration, using Euler

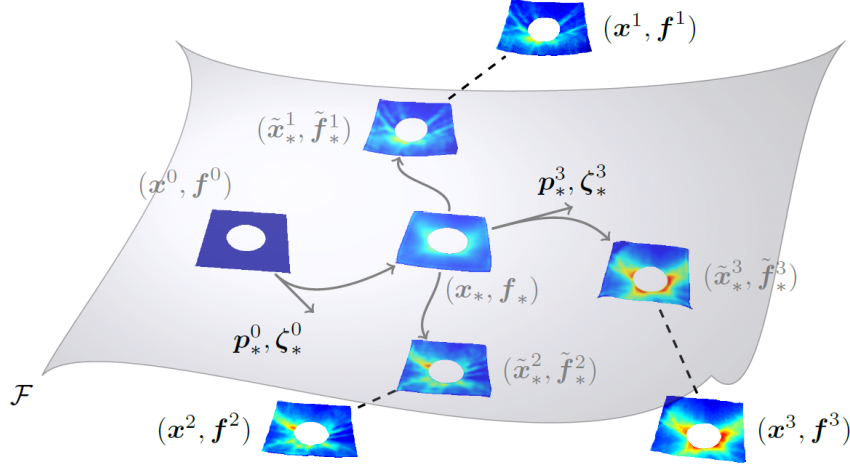


Figure 2: Illustration of the fshape atlas estimation procedure. the estimated mean template $(\mathbf{x}_*, \mathbf{f}_*)$ is a deformation of an initial hypertemplate $(\mathbf{x}^0, \mathbf{f}^0)$. The hypertemplate transformed to each target belongs to the orbit \mathcal{F} of the hypertemplate for example $(\tilde{\mathbf{x}}_*^3, \tilde{\mathbf{f}}_*^3) = (\phi_*^3, \zeta_*^3) \cdot (\mathbf{x}_*, \mathbf{f}_*) \in \mathcal{F}$

230 midpoint scheme for the time-discretization of Hamiltonian and adjoint Hamiltonian systems. We refer to Charlier et al. (2015) for the detailed expressions. The computational burden of the algorithm concentrates in repeated computations of sums of kernels and kernel derivatives in the shooting equations and evaluation of the functional, these specific parts are implemented in CUDA to
 235 achieve competitive speed and precision.

Apart from the construction of a discrete fshape hypertemplate, the rest of the estimation is fully automatic but depends on a few parameters including the weighting coefficients $\gamma_V, \gamma_f, \gamma_W$ between the different energy terms, but also the parameters associated to the kernels that define the metrics on velocity fields
 240 and on fvarifolds. In the applications of this paper, we use isotropic Gaussian kernels for k_g and k_f in (3), which are determined by two variance parameters σ_g and σ_f . Heuristically, σ_g and σ_f measure the sensitivity scale of the fidelity term respectively in the spatial and signal domain. While too big values would typically result in a very imprecise overlap of fshapes, too small values on the

245 other hand could either induce over-fitting or result in no evolution at all in the
case where fshapes are initially too far apart relative to these scales. We address
this issue by completing the approach of Charlier et al. (2015) with a coarse-
to-fine multiscale scheme that allows for successively decreasing σ_g and σ_f in
the minimization. A last important parameter is the scale of the vector kernel
250 defining the space of velocity fields V . Instead of using a single Gaussian kernel,
we also allow for multiscale deformations by incorporating sums of Gaussians
with different scales, as proposed in Bruveris et al. (2012).

2.3. Variability Analysis and Classification

Augmenting the current fshape framework with a classification module is a
255 natural extension with several merits. On experimental data with confirmed
diagnosis, automated classification can demonstrate the mean template-based
variability of retinas has indeed anatomical and clinical relevance. Furthermore,
a classification module can determine whether a particular anatomical feature,
thickness or topology of a layer, is significantly correlated with a disease. This
260 can lead to better understanding of the structural manifestation of the disease
in terms of cause or effect and directly assist in clinical decision making.

The previous section demonstrated how, given a set of observations, in our
case retinal surfaces, the fshape approach can generate an estimated mean atlas
of the dataset. Such a group average retina in itself is useful for purposes such
265 as qualitative group comparisons. To quantify more accurately group differ-
ences and perform statistical classification, it is necessary to rely on the latent
variables (ϕ^i, ζ^i) that characterize deviations in shape and thickness from the
template for each subject. These are automatically estimated in the previous
framework and, as we saw, are optimal or 'geodesic' transformation paths for
270 the energy of (2) which can serve as metrics for inter-eye variability.

Statistical analysis on shape spaces typically exploits the linearization on
diffeomorphism groups given by the initial velocity fields (or momenta) and
extract main shape variance components via dimensionality-reduction methods
like PCA (Vaillant et al., 2004). We easily extend that principle to fshapes by ap-

275 pying PCA to the initial momenta $(\mathbf{p}_*^i)_i$ and functional residuals $(\zeta_*^i)_i$ of Section
2.2.4. However, for classification tasks, PCA directions need not be discrimina-
280 tive with regard to the differences across populations (e.g controls vs glaucoma
in our case). Supervised methods such as a Linear Discriminant Analysis (LDA)
classifier (Hastie et al., 2009) are generally more adequate for pathology detec-
tion (Durrleman et al., 2014). The method we propose in the present work is
a variation of LDA tuned to the specific structure of variables issued by fshape
atlases, which deals with a relatively low number of high-dimensional samples in
particular Hilbert spaces. The general setting is described below and the results
of the classification experiments using this method are presented in section 3.3.

285 2.4. Regularized LDA on Hilbert space valued dataset

LDA is a classic linear classification method that may be viewed as a weighted
principal component analysis (see chapter 4 of (Hastie et al., 2009) for an intro-
duction). We discuss hereafter the general case in which data points x_1, \dots, x_N
belong to a separable Hilbert space H of a possibly infinite dimension. In the
290 context of fshapes and the applications of this paper (cf section 3.3), the x_i 's
are either functional residuals for which $H = L^2(X)$ the space of L^2 functions
on the template (or the discrete equivalent with the metric given by (6)), and
initial momenta in which case $H = V^*$.

2.4.1. Within- and between-class scatter operators

295 We assume that the N observations at hand are divided into K classes
 C_1, \dots, C_k such that $\{1, \dots, N\} = \bigcup_{k=1}^K C_k$. The mean of the k -th class is
 $\bar{x}_k = \frac{1}{N_k} \sum_{i \in C_k} x_i$ where $N_k = \text{Card}(C_k)$. Let us define the within-class scatter
operator $S_w : H \rightarrow H$

$$S_w(\cdot) = \frac{1}{N} \sum_{k=1}^K \sum_{i \in C_k} \langle x_i - \bar{x}_k, \cdot \rangle_H (x_i - \bar{x}_k)$$

and the between-class scatter operator $S_b : H \rightarrow H$

$$S_b(\cdot) = \frac{1}{N} \sum_{k=1}^K N_k \langle \bar{x}_k - \bar{x}, \cdot \rangle_H (\bar{x}_k - \bar{x}).$$

In the case where H is of finite dimension, the matrix associated to the operator
 300 S_w (or S_b) is the standard within-class (or between-class) scatter matrix.

2.4.2. Finite dimensional representation

We denote $H_0 = \text{Span}(x_1, \dots, x_N) \subset H$ and $q = \dim(H_0)$. In our frame-
 work, H is high (or infinite) dimensional, and, in practice, q equals to N .
 We then denote $f_{H_0} : H_0 \rightarrow H_0$ the restriction to H_0 of any linear mapping
 305 $f : H \rightarrow H$ such that $f(H_0) \subset H_0$. In particular, we may consider the restric-
 tions S_{w,H_0} and S_{b,H_0} of the scatter operators S_w and S_b , respectively.

Let us now choose an arbitrary *isometric* linear mapping $L : H_0 \rightarrow \mathbb{R}^q$ in
 order to get a finite representation of the data

$$\tilde{x}_i = Lx_i \in \mathbb{R}^q.$$

By definition we have $\|\tilde{x}_i\|_{\mathbb{R}^q} = \|x_i\|_H$. From now on, we will work with this
 new representation of the data and we may consider their corresponding scatter
 operators $\tilde{S}_w, \tilde{S}_b : \mathbb{R}^q \rightarrow \mathbb{R}^q$ defined as,

$$\tilde{S}_w = LS_{w,H_0}L^\dagger \text{ and } \tilde{S}_b = LS_{b,H_0}L^\dagger,$$

where $L^\dagger : \mathbb{R}^q \rightarrow H_0$ is given by $\langle L^\dagger a, x \rangle_H = \langle a, Lx \rangle_{\mathbb{R}^q}$ for any $x \in H_0$ and
 $a \in \mathbb{R}^q$.

When the x_i 's form a basis of H_0 , an effective way to build an isometric
 310 mapping L is to introduce the mapping $\gamma(x) = (\langle x, x_i \rangle_H)_{i=1}^N$ for any $x \in H_0$
 and the Gram matrix $G = [\langle x_i, x_j \rangle_H]_{i,j=1}^N \in \mathbb{R}^{N \times N}$. It is then easy to check
 that the linear mapping $Lx = G^{-1/2}\gamma(x)$ is isometric.

2.4.3. Discriminant axes in the finite dimensional space

In our framework, the within-class scatter operator \tilde{S}_w may not be invertible
 in general. We consider the following regularization of the within-class scatter

operator

$$\tilde{S}_w^\varepsilon = \tilde{S}_w + \varepsilon Id_{\mathbb{R}^q}, \quad (11)$$

where $Id_{\mathbb{R}^q}$ is the identity matrix and $\varepsilon > 0$ is a regularization parameter which
 315 has to be calibrated by the user as described below.

The discriminant spaces of the LDA are given by the eigendecomposition of

$$\tilde{A}^\varepsilon = (\tilde{S}_w^\varepsilon)^{-1} \tilde{S}_b.$$

The number of non-vanishing eigenvalues is limited by the rank of \tilde{S}_b which is less than $K - 1$. In general, we have the $K - 1$ corresponding unit eigenvectors $\tilde{u}_1, \dots, \tilde{u}_{K-1} \in \mathbb{R}^q$ that will be used to derive the discriminant axes in H .

2.4.4. Classification with regularized LDA

The rationale behind the isometric dimension reduction is that the unit vectors

$$u_\ell = L^\dagger \tilde{u}_\ell \in H_0, \quad \ell = 1, \dots, K - 1$$

320 are the eigenvectors corresponding to the non-vanishing eigenvalues of $A_{H_0}^\varepsilon = (S_{w,H_0} + \varepsilon Id_{H_0})^{-1} S_{b,H_0}$, as in (Friedman, 1989). This classic trick avoids numerical issues as the matrix inversion is performed in the small dimensional space \mathbb{R}^q .

Then, given a new observation $y \in H$ we use directly the discriminant axes
 325 u_1, \dots, u_{K-1} to define a classification rule. For instance, when $K = 2$ (a two-class classifier) we have $u_1 \propto (S_w^\varepsilon)^{-1}(\bar{x}_1 - \bar{x}_2)$ where the \propto symbol means “collinear to”. The discriminant rule is then a threshold on $y \rightarrow \langle u_1, y - \bar{x} \rangle_H$.

2.4.5. Calibration of the regularization parameter

The regularization parameter $\varepsilon > 0$ can be optimized with a leave- p -out
 330 cross-validation (CV) procedure. Note that we do not need to compute the Gram matrix G appearing in the definition of the isometric mapping L at each stage of the CV. This is very helpful since the operation may be costly depending on the magnitude of N . Instead, the $N \times N$ “full” Gram matrix G_f with all the observations is computed once, and the $(N - p) \times (N - p)$ Gram matrices

335 G are obtained by deleting the p rows and columns of G_f corresponding to the
 p left-out observations.

3. Experimental Results

Peripapillary OCT images from 53 eyes with confirmed diagnosis from 10
controls, 10 bilateral glaucoma patients, and 7 unilateral glaucoma patients
340 were included in the experiment. Written consent forms were obtained from all
participants and ethics review was approved by the Office of Research Ethics at
Simon Fraser University (SFU) and the Research Ethics Board of the University
of British Columbia (UBC). In addition to OCT imaging, all participants were
subject to a battery of standard tests, including dilated stereoscopic examination
345 of the optic nerve, stereo disc photography analysis and visual field abnormality
check, to ensure there was no other pathology present. The acquired images were
smoothed, segmented, and measured for retinal layer thickness as described in
Section 2.1.

3.1. Mean template generation

350 The mean templates of retinal nerve fiber layer (RNFL) posterior surfaces
and associated RNFL thickness maps were generated with the atlas estimation
algorithm described above. The surfaces were rigidly registered prior to the
mean template generation, as shown in in Fig. 3 (a). The physical dimensions
of the images vary as the imaging field of view changes depending on the axial
355 length of the eye. We also used a multiscale approach for the different parame-
ters: the deformation kernel is a mixture of Gaussian of scales 2.4, 1.2, 0.6 and
0.3 mm while the algorithm is run successively with three sets of scale parame-
ters $\sigma_g = 0.8, 0.4, 0.2$ and $\sigma_f = 0.3, 0.2, 0.1mm$ for the fidelity term as described
in Section 2.2.4.

360 Fig. 3 (b), (c), and (d) respectively show the mean templates generated from
all eyes ($N = 53$), bilaterally normal eyes ($N = 20$), and glaucomatous eyes
($N = 26$). Qualitatively, the normal mean template in Fig. 3 (c) displays the

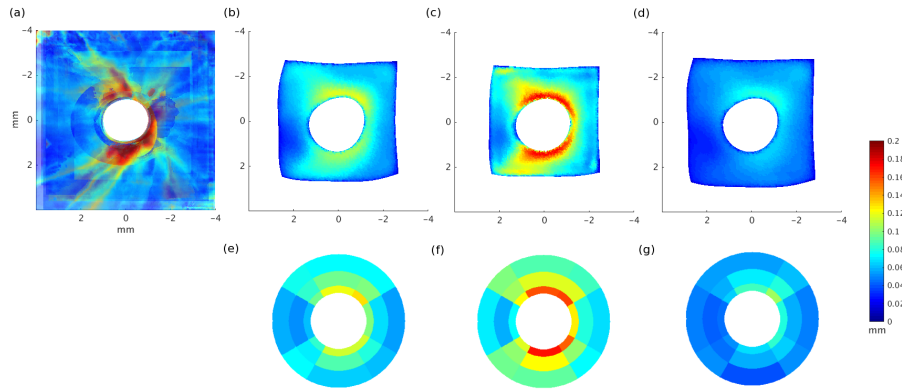


Figure 3: (a) Aligned RNFL surfaces with RNFL thickness mapping, (b) mean template of all RNFLs, (c) mean template of normal RNFLs only, and (d) mean template of glaucomatous RNFLs only. Note the low estimated RNFL thickness of the mean glaucomatous template as compared to that of the mean normal template. (e)-(g) show sectoral thickness averages of all RNFLs, normal RNFLs, and glaucomatous RNFLs, respectively. The sectors in (e)-(g) are not in the physical space as the mean templates in (b)-(d), and plotted for visualization with each sector value given by the average RNFL thickness in the corresponding sector across the eyes in the group.

characteristic hourglass pattern in RNFL thickness, whereas the glaucomatous mean template in Fig. 3 (d) shows much thinner RNFL thickness overall, with superior peripapillary RNFL slightly more preserved compared to the inferior region.

Fig. 3 (e), (f), and (g) show typical sectoral averaging of the RNFL thickness in all eyes, bilaterally normal eyes, and glaucomatous eyes groups. The sectorization process is described in detail in (Lee et al., 2014). Briefly, sectors are drawn in each image by the distance from the BMO (boundaries here are at 0.25, 0.5, 1.0, 1.25 mm from the BMO) and angular sections of superior, inferior, nasal, temporal, and in-between regions determined by the relative position from the reference line extending horizontally from the BMO centroid to the right frame border of the image. Unlike the fshape metrics, sectoral average thickness is measured in the sectors that are defined individually in each eye, and a group average is given by averaging the values in the corresponding sec-

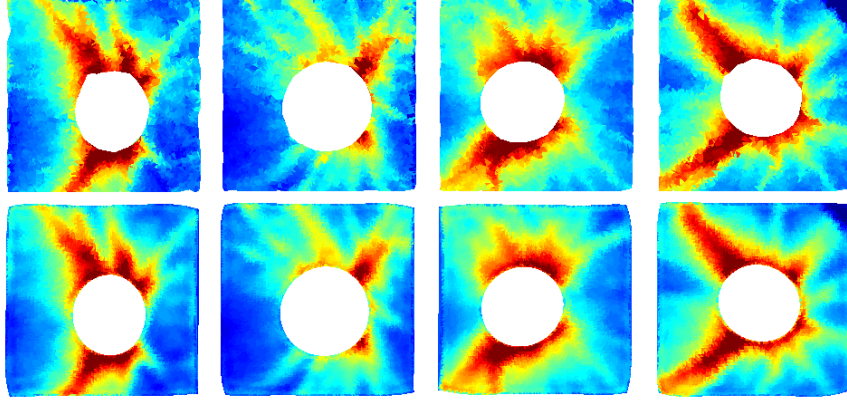


Figure 4: Observed RNFLs (top row) and their reconstructions from a common mean template (bottom row). Note that the reconstruction agrees with the pattern of the original RNFL thickness with an overall smooth and noise-reduced profile.

tors across the eyes in the group. Fig. 3 shows general similarity between the fshape mean templates in (b) - (d) and sectoral averages in (e) - (g). We note, however, the sectoral visualization in (e) - (g) is based on an artificial model and does not live in the physical space as the figures in (a) - (d), as the actual sector dimension in each RNFL is different depending on the BMO dimension and possible cropping of the outer sectors at the image boundaries. Even in the relatively crude sectoral averaging, RNFL thickness is distinctly thicker in the bilateral normal group (f) than in the glaucomatous group (g). However, more detailed features such as the hourglass pattern in RNFL thickness, clearly visualized in Fig. 3 (b) and (c), are lost.

Fig. 4 shows the observed RNFL surfaces (top row) and their approximations (bottom row) from the common mean template of 53 RNFLs in Fig. 3 (b). As described by equation (8), the i^{th} approximation is a deformed version of the template, such that $\tilde{\mathbf{x}}_*^i \doteq \phi_*^i(\mathbf{x}_*)$, $\tilde{\mathbf{f}}_*^i \doteq (\mathbf{f}_* + \boldsymbol{\zeta}_*^i) \circ (\phi_*^i)^{-1}$. The salient shape features are reproduced, in particular the RNFL thickness pattern and the BMO location and size. The level of detail in approximation can be tuned by user specified parameters including the kernel sizes for geometric

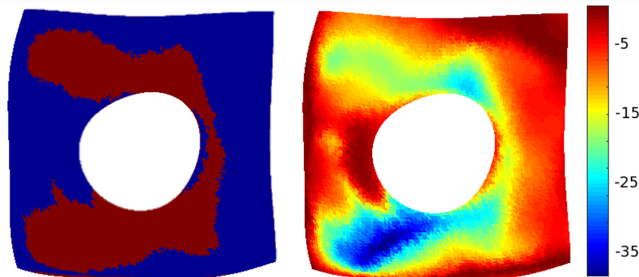


Figure 5: Left: Voxel-wise t -test significance map of retinal nerve fiber layer (RNFL) thickness between normal ($N = 26$) and glaucomatous ($N = 27$) eyes. The number of vertices is 5355, and the red region indicates Bonferroni corrected $p < 0.05/5355$ i. e. $p < 9.34 \times 10^{-6}$. Right: The log of the p -value presented on the entire surface.

deformation and dissimilarity metric, and runtime parameters in the gradient
 395 descent optimization. Although sharper reconstructions can be achieved by adjusting the parameter γ_W , we note that this will not necessarily induce a better discriminative power due to the risk of over-fitting.

3.2. T -test and z -score map between healthy and glaucomatous RNFL thickness residuals

400 The mean templates of the normal (c) and glaucomatous (d) RNFLs in Fig. 3 qualitatively show the difference between the two groups. In order to identify the spatial locations of the thickness difference, a vertex-wise t -test was performed using the functional (thickness) residual ζ^i , the vertex-wise thickness offset between the template and its approximation of the i^{th} observation. The
 405 t -test compared the residuals of 26 normal RNFLs and 27 glaucoma RNFLs at each vertex on the mean template of all RNFLs in Fig. 3 (b).

The result is shown in Fig. 5. On the left, the voxels with p -values less than 0.05 are marked in red and indicate the regions with statistically significant RNFL thickness difference between the normal and glaucomatous RNFLs.
 410 Most of the peripapillary region shows significance, except at the image boundaries, where the fshape correspondence may be less reliable due to the different imaging field of view size, and in the region immediately temporal to BMO.

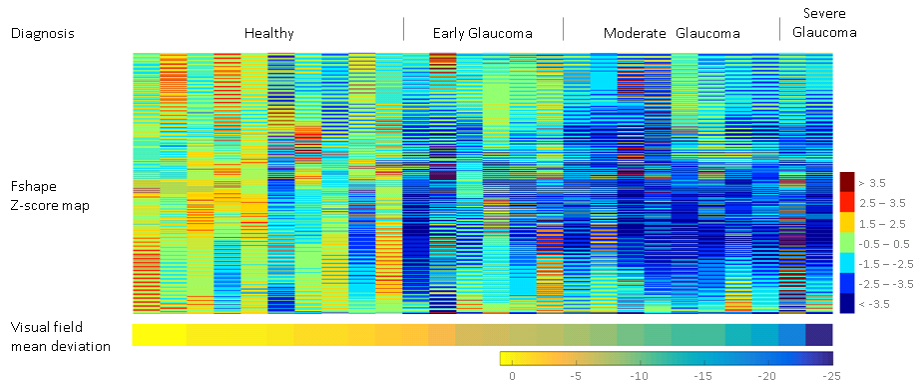


Figure 6: Pointwise Z-score map of retinal nerve fiber layer (RNFL) fshape residuals for age-matched (59.6 ± 6.7) healthy eyes and glaucomatous eyes in early, moderate, and severe stages of the disease, with visual field mean deviation - a measure of glaucomatous vision loss.

More interesting pattern is shown in the right image, which displays the log of the p -value. Cooler colors indicate smaller p -values and greater statistical significance across the multiple RNFLs. The figure shows that glaucomatous thinning occurs most distinctly in the inferior-temporal region of the RNFL, which agrees with previous studies (Morrison and Pollack, 2011; Kanamori et al., 2003) that the inferior peripapillary region is the most distinguishing between normal and glaucomatous eyes, especially in the early stage of the disease. Unlike the previous studies, however, our analysis shows full spatial detail without averaging in sectors, and reveals a clear pattern of statistical significance that resembles the characteristic thickness pattern of a healthy RNFL. The result suggests that the most significant amount of glaucomatous thinning, or the earliest of the thinning, may occur along the RNFL ridges where the RNFL is naturally the thickest.

Fig. 6 plots the pointwise Z-score map of ζ for age-matched (59.6 ± 6.7) healthy eyes and glaucomatous eyes in early, moderate, and severe stages of the disease. Visual field mean deviation (VFMD), a measure of glaucomatous vision loss, is also plotted for each eye. The z-score was computed at each point of the mean template by subtracting the mean ζ of the healthy samples and dividing

by the standard deviation ζ of the healthy samples as a normalized measure. The map shows distinctions between the healthy and early glaucoma groups, and between the early glaucoma and moderate to severe glaucoma groups, along with the individual variability in the relationship between the degrees of RNFL loss and vision loss among the glaucomatous eyes. It is noteworthy that even within
435 the group diagnosed as healthy, the eyes with lower VFMD show generally low ζ values.

3.3. Classification

With the methods detailed in Section 2.3, classification experiments were
440 performed using the LDA classifier on the momenta \mathbf{p}^i , the vertex-wise geometric deformation of the template to the i^{th} observation, and functional residuals ζ^i , the vertex-wise thickness offset between the template and the approximation of the i^{th} retina. The classification experiments are presented as a demonstration of the discriminative power of the fshape metrics and how it captures the
445 anatomical variability in RNFLs.

For each experiment, the LDA classifier was trained by leave-one-out cross-validation, with the best-performing regularization parameter ε (see section 2.4.5) selected by a nested leave-one-out procedure from values between 0.001 to 1.

3.3.1. Healthy vs. Glaucoma

Age-matched (59.6 ± 6.7) normal ($N = 10$) and glaucomatous ($N = 18$) RNFLs were classified with the result in Table 1. The result indicates that the fshape metrics of the peripapillary RNFL posterior surface and RNFL thickness can predict the clinical diagnosis of glaucoma with high accuracy, and confirms
455 the connection between vision loss that bases glaucoma diagnosis, and characteristic morphological changes in the RNFL.

3.3.2. Healthy vs. Suspect

Glaucoma is generally bilateral, affecting both eyes of the patient, and often asymmetric, such that the affected fellow eyes exhibit different degrees of

460 severity. In unilateral primary glaucoma in which only one eye is diagnosed with glaucoma, the healthy fellow eye is at a greater risk of developing glaucoma in the future than healthy eyes of bilaterally normal subjects (Kass et al., 1976; Susanna et al., 1978). Based on this, we labeled 7 healthy fellow eyes from unilateral glaucoma cases as *suspect*, and attempted to detect these from 465 healthy eyes of bilaterally nonglaucomatous subjects. The classification result of the bilaterally healthy eyes ($N = 19$, mean age: 43.4 ± 14.8) and suspect eyes ($N = 7$, mean age: 57.1 ± 12.4) are shown in Table 1.

That the suspect eyes are distinguished from the bilaterally healthy eyes is noteworthy, considering that the suspect eyes are pre-diagnosis and without 470 functional loss or other conventional clinical features of glaucoma. As expected, the classification rates are lower than those between confirmed glaucomatous eyes and bilaterally healthy eyes in the previous experiment, but still relatively high. The result suggests the fshape metrics may capture some morphological changes in the RNFL that precedes vision loss in glaucoma.

	Accu. (%)	Sens. (%)	Spec. (%)
Healthy vs. Glaucoma	92.9	94.4	90.0
Healthy vs. Suspect	88.5	71.4	94.7

Table 1: Accuracies, sensitivities, and specificities of the classification of healthy, glaucomatous, and suspect RNFLs based on the fshape metrics of RNFL posterior surface geometry and RNFL thickness. Both Healthy vs. Glaucoma and Healthy vs. Suspect show high classification success rates.

475 In LDA classification, an intuitive way to understand the result is to visualize the classifier, or the direction that yields the maximum between-class variance. The classifier for the Healthy vs. Glaucoma dataset is visualized in Fig. 7 on the mean template RNFL, in functional residual (RNFL thickness), and x , y , and z coordinates of initial momenta representing the template deformation. 480 The relative magnitude can be interpreted as the degree of contribution to the classification, and it can be seen that RNFL thickness was a more decisive

factor in the classification of healthy and glaucomatous eyes than posterior RNFL surface geometry. The map in Fig. 7 A, which is the spatial pattern of classification contribution of RNFL thickness residual, is similar to the left image of Fig. 5, which is the statistical significance map of the group difference between the two classes. This confirms that the classification was the most influenced by the regions where there are the most significant difference between the healthy and glaucomatous eyes.

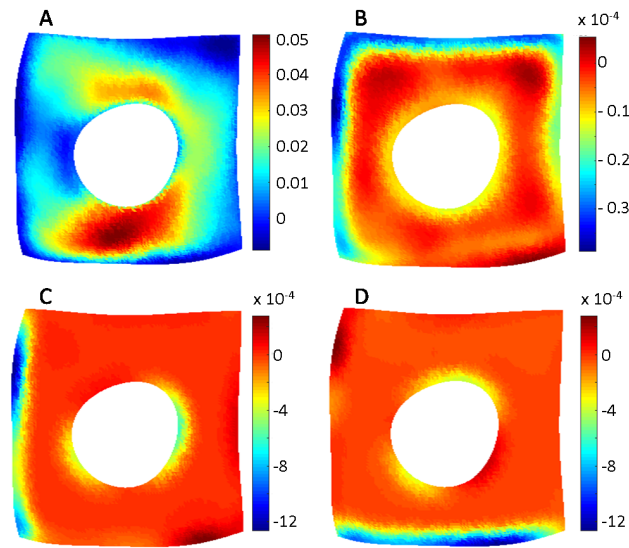


Figure 7: LDA classifier for Healthy vs. Glaucoma data in A) functional residual, B) initial momenta in x-direction, C) initial momenta in y-direction, and D) initial momenta in z-direction.

3.3.3. Classification with RNFL thickness sectoral averages

The above classifications for Healthy vs. Glaucoma and Healthy vs. Suspect were repeated with the same eyes with the RNFL thickness sectoral averages as described in Section 3.1. In this case, correspondence is assumed between the corresponding sectors across the subjects. Comparing the results in Table 2 with that in Table 1 of the fshape metrics, the classifier performance is significantly reduced. The higher classification success rate between the healthy and

glaucomatous groups than between the healthy and suspect groups is consistent with the fshape metrics result in 1 and can be attributed to the large RNFL thickness difference between the two groups, visualized in the sectoral group averages in Fig. 3. The thinning in the glaucomatous RNFLs is distinguish-
500 able even after sectoral averaging. However, any difference between healthy and suspect RNFLs is likely more subtle, as the eyes in the suspect groups are nonglaucomatous with normal visual function and without clinically observable structural degradation. Such fine distinctions may be smoothed out by the sectoral averaging. The result of this experiment, along with the results presented above, suggests that spatially more detailed comparison and analysis are
505 possible with the fshape metrics than the conventional sectorization approach. Finer sectorization for more localized comparison will be limited by decreasing confidence in the correspondence between the same sectors from different eyes.

	Accu. (%)	Sens. (%)	Spec. (%)
Healthy vs. Glaucoma	66.7	64.7	70.0
Healthy vs. Suspect	59.2	57.1	60.0

Table 2: Accuracies, sensitivities, and specificities of the classification of healthy, glaucomatous, and suspect RNFLs based on RNFL thickness sectoral averages. In comparison with the results in Table 1 with RNFL fshape metrics, the classification performance is lower.

3.3.4. RNFL thickness vs. RNFL surface geometry

510 In order to compare the discriminating power of RNFL thickness against RNFL posterior surface geometry, and to confirm what was shown in Fig. 7, we repeated the Healthy vs. Glaucomatous classification with RNFL functional residual (RNFL thickness) and initial momenta (RNFL posterior surface geometry) separately. The result is summarized in Table 3. As expected from
515 Fig. 7, the classification result with RNFL thickness only is comparable to that with both RNFL thickness and geometry, and superior to the result with RNFL geometry only.

	Accu. (%)	Sens. (%)	Spec. (%)
Functional residual	92.9	94.4	90.0
Geometrical momenta	67.9	66.6	70.0

Table 3: Accuracies, sensitivities, and specificities of the classification of healthy and glaucomatous RNFLs by RNFL thickness only (functional residual, top row) and RNFL posterior surface geometry only (geometrical momenta, bottom row) (BMO-based rigid pre-registration)

Prior to the fshape template generation, the RNFL surfaces were rigidly registered by the BMO centroids. The fshape geometrical deformation then contains information for both RNFL surface topology, and the distance between
520 RNFL and BMO. The experiment was repeated with rigid pre-registration by aligning the central opening of the RNFL surface instead of the BMO centroid, and the result is summarized in Table 4. The functional-based classification is comparable; however, the geometry-based classification is worse in that most of
525 the glaucoma cases are misclassified. This suggests that the factors that determine the distance between the RNFL and BMO, such as the optic canal skew or post-RNFL retinal thickness, may be affected by glaucomatous structural change, more so than RNFL posterior surface topology alone.

	Accu. (%)	Sens. (%)	Spec. (%)
Functional residual	96.4	94.4	100.0
Geometrical momenta	53.6	33.3	90.0

Table 4: Accuracies, sensitivities, and specificities of the classification of healthy and glaucomatous RNFLs by RNFL thickness only (functional residual, top row) and RNFL posterior surface geometry only (geometrical momenta, bottom row) (RNFL opening-based rigid pre-registration)

3.3.5. Comparison with standard surface registration

530 To illustrate the influence of functions on the estimated deformation and residuals, we performed a simple matching experiment between two surfaces,

by a standard surface LDDMM (Lee et al., 2015) and by the fshape approach. In the former, the residuals were retrieved by closest point projection from the target to the mapped surface. As shown in second column of Figure 8, the
535 ssLDDMM algorithm returns a geometric map close to identity between the given source and target, whereas the fshapes returns a geometric map further from the identity indicating that it was driven by the function signal for creating overlap. This has a direct impact on the error remaining after mapping as well (as shown in the third column). The geometry-based transformation of
540 source signal to target derived from ssLDDMM shows high values of remaining error after mapping - the error profile shows the presence of two unregistered function signals as the transformed source function is subtracted from the target signal. In comparison, by incorporating the function signal in the fshapes registration, the function profiles of the source and target are matched leading
545 to a lower error profile obtained by fshape approach. A typical effect on the atlas estimation is shown in Figure 9 in which mean templates and group difference significance maps for the older normals and older glaucoma groups are compared. The mean function estimated using fshapes is sharper and retains more high frequency features of the population since the functions are aligned,
550 whereas the ssLDDMM mean is more blurry and retains less of the function profiles since the geometry-only registration does not incorporate the alignment of functional signals. An important point to note is that the geometry-only-based ssLDDMM benefits from the pre-alignment that occurs in the imaging system by using the chin and forehead restraint providing a vertical orientation of the
555 head for retinal imaging. This normalizes the pose for the superior-inferior crescent shaped nerve fiber thickness profile as it is consistently located across individuals albeit the individual variability is still present. Hence, before ssLDDMM was utilized here, the pose of the source and target were already a-priori aligned, which aligns the function signal and benefits ssLDDMM. Without this
560 benefit, if the original pose were to be arbitrary, and absent surface geometry features for registration, as is the case here, the ssLDDMM will return close to identity warps. These warps which would align the geometry but fail to align

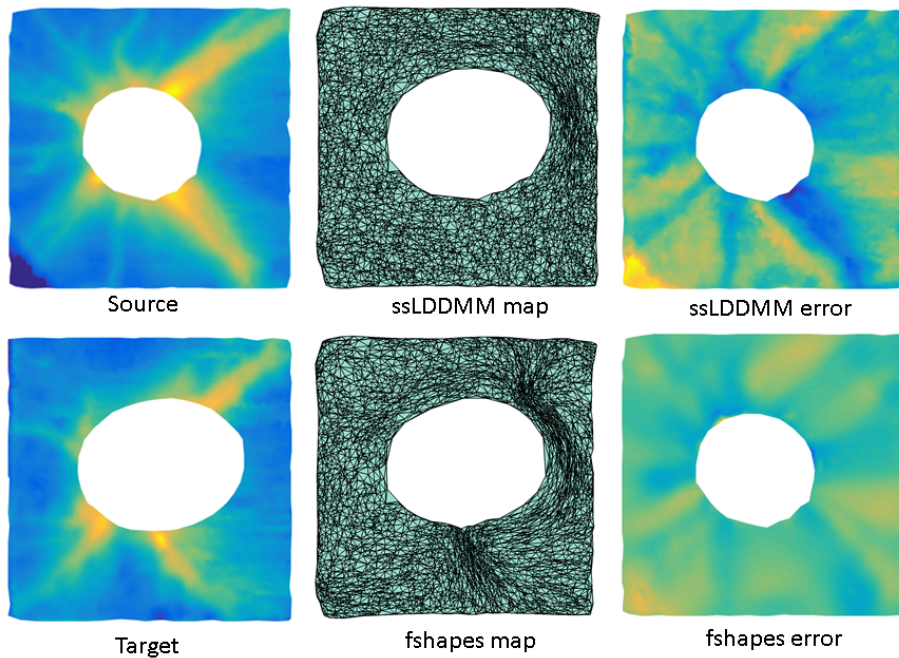


Figure 8: The first column shows a source RNFL thickness function and a target RNFL thickness function. Top row shows results with standard-surface LDDMM (ssLDDMM) and bottom row shows results using the fshape algorithm. The ssLDDMM algorithm returns a geometric map close to identity to map the given source and target, and the remainder after transforming the functions from this map show considerable residue that this mapping is unable to register. The bottom row show the same experiments with fshapes, where the residue left after mapping the source to the target function is smaller and more diffuse since the overall functions have been matched.

the function signal on the geometry as this aspect is not taken into account in ssLDDMM registration. Hence, the competing state-of-art of ssLDDMM will fail as compared to the fshape approach.

4. Discussion and future work

We presented a novel application of the fshape framework for variability analysis of shape and associated signals in retinal optical coherence tomography (OCT) images, comparing healthy, glaucomatous, and suspect peripapillary

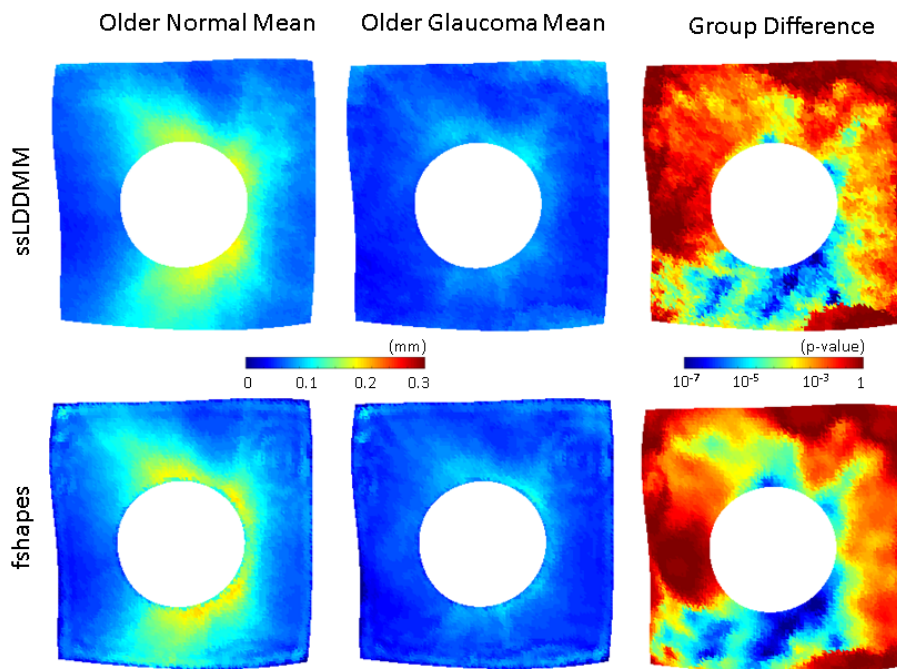


Figure 9: Top row shows results with standard-surface LDDMM (ssLDDMM) and bottom row shows results using the fshape algorithm. The first column is the group average of healthy individuals, the second is group average of individuals with Glaucoma and the third column shows t-statistic from group difference using these two methods. Note that the ssLDDMM results are more blurry due to mis-registration of the functional signal, and the group differences are missing the superior part of the crescent shape of pathological changes as well as overall being more diffuse and less strong everywhere.

570 retinal nerve fiber layers (RNFL). The fshape framework generated mean tem-
 plate RNFL surfaces and thickness maps and fshape metrics for each RNFL.
 Compared to the conventional sectoral measures, fshape metrics capture the
 variability in the anatomy and anatomically oriented signals jointly, and can be
 used to identify features that are important in distinguishing pathological cases
 575 from healthy cohort. There is potential for broad application of the framework
 in both longitudinal and cross-sectional studies of retinal OCT images.

In retinal OCT imaging, factors such as varying field-of-view sizes and the
 position of the retina within the image frame can contribute to the geometrical

variability in the data. In order to remove such artefacts, the RNFL surfaces in
580 our experiments were pre-aligned but indirectly by aligning the BMOs instead
of the surfaces themselves. This resulted in the relative positions of the RNFL
surfaces containing the extra information of the RNFL-BMO distance, which
reflects the optic canal skew and total retinal thickness. Directly aligning the
RNFL surfaces with each other worsened the geometry-based classification, as
585 described in 3.3.4. This showed that in our data the gross features such as
the optic cup skew or total retinal thickness were indicative of glaucoma, and
demonstrated the utility of fshape-based analysis in testing the significance of
a specific shape feature.

RNFL within 0.25 mm from BMO was excluded, because the layer boundary
590 in the region is often ambiguous. The input surfaces were then concentric with
respect to BMO and centrally overlapping, but the outer edges were mismatched
due to the varying image sizes. Whereas BMO is an anatomical structure, the
image boundary is artificial and depends on the subject’s axial length. We were
wary that the estimated momenta at the boundary region would be affected by
595 this mismatch. This could also have confounded the classification as the metrics
at the boundary region have less spatial correspondence across eyes. However,
the regions near the outer boundary also made the least contribution to the
classification, as it can be seen in Fig. 5 and 7. One way to possibly mitigate
the boundary effect is to crop the images, for example, to the largest commonly
600 overlapping region. This however can waste a significant amount of data, and
the cropping must be done in a way that does not introduce additional bias.

Choosing a good initial template is important in the computation and quality
of the final mean template. The initial template must be topologically equiv-
alent to the observations. An implicit assumption is that the observations are
605 topologically similar. Retinal layer surfaces in OCT images can be modeled rel-
atively simply, as a planar surface in the macular region, and as a planar surface
with a hole in the center in the peripapillary region. The initial template can
incorporate the general dimensions of the observations. A future work would
address the question of the sensitivity of the method to the choice of the initial

610 template. This will be an important issue with more complex structures than
retinal surfaces, like lamina cribrosa.

Future work will include creation of population normative atlases and statis-
tics using larger cohorts, and a comprehensive investigation of the optic nerve
and macular morphology in a larger data using the fshape and classification
615 modules. The goal is to better detect and understand the shape changes or
differences correlated with diseases such as glaucoma. In this report only RNFL
was examined, but any other retinal layers can be similarly analyzed, in any
combination of layers, surface topology, and signal types. Shape and signal
from multiple eyes will be made comparable by a common atlas, and signifi-
620 cance of a metric, region, or anatomical structure can be tested by statistical
analysis including the classification modules above. In this work the mean tem-
plate generation and classification process yielded convincing results despite of
the mismatch between the RNFL sizes and relative BMO locations. Another
direction of future research is to apply the fshape framework to structures like
625 lamina cribrosa, which is not only more topologically complex but also much
less consistently visible across eyes than retinal layers. The combined frame-
work may be used to extract common shape information from multiple lamina
cribrosa images with varying and limited visibility. Lastly, it will be of value
and interest to comparatively investigate different classification methods and
630 optimization techniques for the retinal fshape metrics.

Acknowledgment

This work was supported by Canadian Institute of Health Research(CIHR),
Natural Science and Engineering Research Council of Canada (NSERC), Michael
Smith Foundation for Health Research (MSFHR), Alzheimer Society Research
635 Program (ASRP), and Brain Canada.

References

- Arguillere, S., Trélat, E., Trouvé, A., Younes, L., 2015. Shape deformation analysis from the optimal control viewpoint. *Journal de Mathématiques Pures et Appliquées* 104, 139–178.
- 640 Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision* 61.
- Bruveris, M., Risser, L., Vialard, F., 2012. Mixture of Kernels and Iterated Semidirect Product of Diffeomorphisms Groups. *Multiscale Modeling and Simulation* 10, 1344–1368. URL: <http://epubs.siam.org/doi/abs/10.1137/110846324>, doi:10.1137/110846324, arXiv:<http://epubs.siam.org/doi/pdf/10.1137/110846324>.
- 645 Charlier, B., Charon, N., Trouvé, A., 2015. The Fshape Framework for the Variability Analysis of Functional Shapes. *Foundations of Computational Mathematics* , 1–71.
- 650 Charon, N., Trouvé, A., 2013. Functional currents : a new mathematical tool to model and analyse functional shapes. *JMIV* 48, 413–431.
- Charon, N., Trouvé, A., 2013. The varifold representation of non-oriented shapes for diffeomorphic registration. *SIAM journal of Imaging Science* 6, 2547–2580.
- 655 Chauhan, B.C., Burgoyne, C.F., 2013. From clinical examination of the optic disc to clinical assessment of the optic nerve head: A paradigm change. *Am. J. Ophthalmol.* 156, 218–227.
- Chen, M., Lang, A., Ying, H.S., Calabresi, P.A., Prince, J.L., Carass, A., 2014. Analysis of macular OCT images using deformable registration. *Biomed. Opt. Express* 5, 2196–2214. URL: <http://www.opticsinfobase.org/boe/abstract.cfm?URI=boe-5-7-2196>, doi:10.1364/BOE.5.002196.
- 660

- Durrleman, S., Pennec, X., Trouvé, A., Ayache, N., 2008. A forward model to build unbiased atlases from curves and surfaces. Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy .
- 665 Durrleman, S., Prastawa, M., Charon, N., Korenberg, J., Joshi, S., Gerig, G., Trouvé, A., 2014. Deformetrics : morphometry of shape complexes with space deformations. Neuroimage 101, 35–49.
- Friedman, J.H., 1989. Regularized discriminant analysis. Journal of the American Statistical Association 84, pp. 165–175.
- 670 Garvin, M., Abramoff, M., Kardon, R., Russell, S., Wu, X., Sonka, M., 2008. Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-D graph search. Medical Imaging, IEEE Transactions on 27, 1495–1505. doi:[10.1109/TMI.2008.923966](https://doi.org/10.1109/TMI.2008.923966).
- Gibson, E., Young, M., Sarunic, M., Beg, M., 2010. Optic nerve head registration via hemispherical surface and volume registration. Biomedical Engineering, IEEE Transactions on 57, 2592–2595. doi:[10.1109/TBME.2010.2060337](https://doi.org/10.1109/TBME.2010.2060337).
- Glaunès, J., Trouvé, A., Younes, L., 2004. Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2, 712–718. doi:<http://doi.ieeecomputersociety.org/10.1109/CVPR.2004.81>.
- 680 González-García, A.O., Vizzeri, G., Bowd, C., Medeiros, F.A., Zangwill, L.M., Weinreb, R.N., 2009. Reproducibility of RTVue retinal nerve fiber layer thickness and optic disc measurements and agreement with Stratus optical coherence tomography measurements. Am. J. Ophthalmol. 147, 1067–1074.
- 685 Gori, P., Colliot, O., Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Poupon, C., Hartmann, A., Ayache, N., Durrleman, S., 2013. Bayesian Atlas Estimation for the Variability Analysis of Shape Complexes. MICCAI , 267–274.

- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical
690 Learning (2nd edition). Springer Series in Statistics, Springer New York Inc.
- He, L., Ren, R., Yang, H., Hardin, C., Reyes, L., Reynaud, J., Gardiner, S.K.,
Fortune, B., Demirel, S., Burgoyne, C.F., 2014. Anatomic vs. acquired image
frame discordance in spectral domain optical coherence tomography minimum
rim measurements. PLoS ONE 9, e92225.
- 695 Huang, D., Swanson, E.A., Lin, C.P., Schuman, J.S., Stinson, W.G., Chang,
W., Hee, M.R., Flotte, T., Gregory, K., Puliafito, C.A., Fujimoto, J.G., 1991.
Optical coherence tomography. Science 254, 1178–1181.
- Jørgensen, T.M., Thomadsen, J., Christensen, U., Soliman, W., Sander, B.,
2007. Enhancing the signal-to-noise ratio in ophthalmic optical coherence
700 tomography by image registration—method and clinical examples. J. Biomed.
Opt. 12, 041208.
- Joshi, S., Davis, B., Jomier, M., Gerig, G., 2004. Unbiased diffeomorphic atlas
construction for computational anatomy. NeuroImage 23, S151–S160.
- Kanamori, A., Nakamura, M., Escano, M.F., Seya, R., Maeda, H., Negi, A.,
705 2003. Evaluation of the glaucomatous damage on retinal nerve fiber layer
thickness measured by optical coherence tomography. Am. J. Ophthalmol.
135, 513–520.
- Kass, M., Kolker, A., Becker, B., 1976. Prognostic factors in
glaucomatous visual field loss. Archives of Ophthalmology 94,
710 1274–1276. URL: [+http://dx.doi.org/10.1001/archophth.1976.03910040146002](http://dx.doi.org/10.1001/archophth.1976.03910040146002),
doi:10.1001/archophth.1976.03910040146002,
[arXiv:/data/Journals/OPHTH/17855/archophth948002.pdf](https://arxiv.org/data/Journals/OPHTH/17855/archophth948002.pdf).
- Lee, S., Fallah, N., Forooghian, F., Ko, A., Pakzad-Vaezi, K., Merkur, A.B.,
Kirker, A.W., Albiani, D.A., Young, M., Sarunic, M.V., Beg, M.F., 2013.
715 Comparative analysis of repeatability of manual and automated choroidal

- thickness measurements in nonneovascular age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* 54, 2864–2871. URL: <http://www.iovs.org/content/54/4/2864.abstract>, doi:10.1167/iovs.12-11521, arXiv:<http://www.iovs.org/content/54/4/2864.full.pdf+html>.
- 720 Lee, S., Han, S.X., Young, M., Beg, M.F., Sarunic, M.V., Mackenzie, P.J., 2014. Optic nerve head and peripapillary morphometrics in myopic glaucoma. *Invest. Ophthalmol. Vis. Sci.* 55, 4378–4393.
- Lee, S., Lebed, E., Sarunic, M.V., Beg, M.F., 2015. Exact surface registration of retinal surfaces from 3-d optical coherence tomography images. *IEEE transactions on bio-medical engineering* 62, 609–17. URL: <http://www.ncbi.nlm.nih.gov/pubmed/25312906>, doi:10.1109/TBME.2014.2361778.
- 725 Li, K., Wu, X., Chen, D., Sonka, M., 2006. Optimal surface segmentation in volumetric images - a graph-theoretic approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28, 119–134. doi:10.1109/TPAMI.2006.19.
- 730 Ma, J., Miller, M.I., Younes, L., 2010. A Bayesian generative model for surface template estimation. *Journal of Biomedical Imaging* 2010, 16.
- Medeiros, F.A., Zangwill, L.M., Bowd, C., Vessani, R.M., Jr, R.S., Weinreb, R.N., 2005. Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography. *Am. J. Ophthalmol.* 139, 44–55.
- 735 Morrison, J., Pollack, I., 2011. *Glaucoma: Science and Practice*. Thieme.
- Niemeijer, M., Garvin, M.K., Lee, K., van Ginneken, B., Abramoff, M.D., Sonka, M., 2009. Registration of 3D spectral OCT volumes using 3D SIFT feature point matching, in: *SPIE Medical Imaging, International Society for Optics and Photonics*. pp. 72591I–72591I.
- 740 Strouthidis, N.G., Yang, H., Reynaud, J.F., Grimm, J.L., Gardiner, S.K., Fortune, B., Burgoyne, C.F., 2009. Comparison of clinical and spectral domain

- optical coherence tomography optic disc margin anatomy. *Invest. Ophthalmol. Vis. Sci.* 50, 4709–4718.
- 745
- Susanna, R., Drance, S., Douglas, G., 1978. The visual prognosis of the fellow eye in uniocular chronic open-angle glaucoma. *The British Journal of Ophthalmology* 62, 327–329.
- Vaillant, M., Miller, M.I., Trouvé, A., Younes, L., 2004. Statistics of diffeomorphisms via tangent space representation. *Neuroimage* 23.
- 750
- Yeo, B., Sabuncu, M., Vercauteren, T., Ayache, N., Fischl, B., Golland, P., 2010. Spherical demons: Fast diffeomorphic landmark-free surface registration. *Medical Imaging, IEEE Transactions on* 29, 650–668. doi:[10.1109/TMI.2009.2030797](https://doi.org/10.1109/TMI.2009.2030797).
- 755
- Young, M., Lebed, E., Jian, Y., Mackenzie, P.J., Beg, M.F., Sarunic, M.V., 2011. Real-time high-speed volumetric imaging using compressive sampling optical coherence tomography. *Biomed. Opt. Express* 2, 2690–2697. URL: <http://www.opticsinfobase.org/boe/abstract.cfm?URI=boe-2-9-2690>, doi:[10.1364/BOE.2.002690](https://doi.org/10.1364/BOE.2.002690).
- 760
- Young, M., Lee, S., Rateb, M., Beg, M.F., Sarunic, M.V., Mackenzie, P.J., 2014. Comparison of the clinical disc margin seen in stereo disc photographs with neural canal opening seen in optical coherence tomography images. *J. Glaucoma* 23, 360–367.
- Zhang, M., Singh, N., Fletcher, P.T., 2013. Bayesian Estimation of Regularization and Atlas Building in Diffeomorphic Image Registration. *Information Processing in Medical Imaging* , 37–48URL: http://dx.doi.org/10.1007/978-3-642-38868-2_4, doi:[10.1007/978-3-642-38868-2_4](https://doi.org/10.1007/978-3-642-38868-2_4).
- 765