



Unified heat kernel regression for diffusion, kernel smoothing and wavelets on manifolds and its application to mandible growth modeling in CT images

Moo K. Chung^{a,b,*}, Anqi Qiu^c, Seongho Seo^d, Houri K. Vorperian^b

^a Department of Biostatistics and Medical Informatics, USA

^b Vocal Tract Development Laboratory, Waisman Center, University of Wisconsin, Madison, USA

^c Department of Biomedical Engineering, National University of Singapore, Singapore

^d Department of Brain and Cognitive Sciences, Seoul National University, Republic of Korea

ARTICLE INFO

Article history:

Received 20 October 2013

Received in revised form 15 February 2015

Accepted 19 February 2015

Available online 2 March 2015

Keywords:

Heat kernel regression

Laplace–Beltrami eigenfunctions

Diffusion wavelets

Surface-based morphometry

Mandible growth

ABSTRACT

We present a novel kernel regression framework for smoothing scalar surface data using the Laplace–Beltrami eigenfunctions. Starting with the heat kernel constructed from the eigenfunctions, we formulate a new bivariate kernel regression framework as a weighted eigenfunction expansion with the heat kernel as the weights. The new kernel method is mathematically equivalent to isotropic heat diffusion, kernel smoothing and recently popular diffusion wavelets. The numerical implementation is validated on a unit sphere using spherical harmonics. As an illustration, the method is applied to characterize the localized growth pattern of mandible surfaces obtained in CT images between ages 0 and 20 by regressing the length of displacement vectors with respect to a surface template.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In medical imaging, anatomical surfaces extracted from MRI and CT are often represented as triangular meshes. The image segmentation and surface extraction processes themselves are likely to introduce noise to the mesh coordinates. It is therefore imperative to reduce the mesh noise while preserving the geometric details of the anatomical structures for various applications.

Diffusion equations have been widely used in image processing as a form of noise reduction since 1990 (Perona and Malik, 1990). Numerous techniques have been developed for surface fairing and mesh regularization (Sochen et al., 1998; Malladi and Ravve, 2002; Tang et al., 1999; Taubin, 2000) and surface data smoothing (Andrade et al., 2001; Chung et al., 2001, 2005; Chung and Taylor, 2004; Cachia et al., 2003a,b; Joshi et al., 2009). Isotropic heat diffusion on surfaces has been introduced in brain imaging for subsequent statistical analysis involving the random field theory (RFT) that assumes an isotropic covariance function as a noise model (Andrade et al., 2001; Chung and Taylor, 2004; Cachia et al.,

2003a,b). Since then, isotropic diffusion has been the standard smoothing technique.

Iterated kernel smoothing has been another widely used method in approximately solving diffusion equations on surfaces (Chung et al., 2005; Han et al., 2006). It is often used in smoothing anatomical surface data including cortical curvatures (Luders et al., 2006b; Gaser et al., 2006), cortical thickness maps (Luders et al., 2006a; Bernal-Rusiel et al., 2008), hippocampus surfaces (Shen et al., 2006; Zhu et al., 2007) and magnetoencephalography (MEG) (Han et al., 2007) and functional-MRI (Hagler et al., 2006; Jo et al., 2007) data on the brain surface. Due to its simplicity, it is the most widely used form of surface data smoothing in brain imaging. In iterated kernel smoothing, kernel weights are spatially adapted to follow the shape of the heat kernel in a discrete fashion along a manifold. In the tangent space of the manifold, the heat kernel with a small bandwidth can be approximated linearly using the Gaussian kernel. The heat kernel with a large bandwidth is then constructed by iteratively applying the Gaussian kernel with the small bandwidth. However, this process compounds the linearization error at each iteration.

We propose a new kernel regression framework that constructs the heat kernel analytically using the eigenfunctions of the Laplace–Beltrami (LB) operator, avoiding the need for the linear approximation used by Chung et al. (2005) and Han et al. (2006).

* Corresponding author at: Medical Science Center 5785, 1300 Highland Ave., Madison, WI 53706, USA.

E-mail address: mkchung@wisc.edu (M.K. Chung).

Although a few studies have introduced the heat kernel in computer vision and machine learning, they mainly used the heat kernels to compute shape descriptors or to define a multi-scale metric (Belkin et al., 2006; Sun et al., 2009; Bronstein and Kokkinos, 2010; de Goes et al., 2008). These studies did not use the heat kernels in regressing functional data on manifolds. This is the first study to use the heat kernel in the form of regression for the subsequent statistical analysis. There have been significant developments in kernel methods in the machine learning community (Schölkopf and Smola, 2002; Nilsson et al., 2007; Shawe-Taylor and Cristianini, 2004; Steinke and Hein, 2008; Yger and Rakotomamonjy, 2011). However, to the best of our knowledge, the heat kernel has never been used in such frameworks. Most kernel methods in machine learning deal with the linear combination of kernels as a solution to penalized regressions. On the other hand, our kernel regression framework does not have a penalized cost function.

Wavelets have recently been popularized for surface and graph data. For instance, spherical wavelets were used on brain surface data already mapped onto a sphere (Nain et al., 2007; Bernal-Rusiel et al., 2008). Since the wavelet basis has local supports in both space and scale, the wavelet coefficients provide shape features that describe local shape variation at a variety of scales and spatial locations. However, spherical wavelets require a smooth mapping from the surface to a unit sphere, thus introducing a serious metric distortion that compounds subsequent statistical parametric maps (SPM). Furthermore, such basis functions are only orthonormal for data defined on the sphere and result in a less parsimonious representation for data defined on other surfaces compared to the intrinsic LB-eigenfunction expansion (Seo and Chung, 2011). To remedy the limitations of spherical wavelets, the diffusion wavelet transform on graph data structures has been proposed (Antoine et al., 2010; Coifman and Maggioni, 2006; Hammond et al., 2011; Kim et al., 2012).

The primary methodological contribution of this study is the establishment of a unified regression framework that combines the diffusion-, kernel- and wavelet-based methods in a coherent mathematical fashion for scalar data defined on manifolds. We unify the apparent differences between the methods while providing detailed theoretical justifications. This paper extends the work by Kim et al. (2011), which introduced heat kernel smoothing to smooth out surface noise in the hippocampus and amygdala. Although the idea of diffusion wavelet transform for surface meshes was explored by Kim et al. (2012), the relationship between the wavelet transform and the proposed kernel regression was not investigated. For the first time, the mathematical equivalence between the two constructs is explained.

The proposed kernel regression framework was subsequently applied in characterizing the growth pattern of the mandible surfaces obtained in CT and identifying the regions of the mandible that show the most significant localized growth. The length of the displacement vector field was regressed over the mandible surface to increase the signal-to-noise ratio (SNR) and hence statistical sensitivity. To our knowledge, this is the first growth modeling of the mandible surface in a continuous fashion without using anatomic landmarks.

2. Methods

2.1. Isotropic diffusion on manifolds

Consider a functional measurement $Y(p)$ observed at each point p on a compact manifold $\mathcal{M} \subset \mathbb{R}^3$. We assume the following linear model on Y :

$$Y(p) = \theta(p) + \epsilon(p), \quad (1)$$

where $\theta(p)$ is the unknown mean signal to be estimated and $\epsilon(p)$ is a zero-mean Gaussian random field. We may assume further $Y \in L^2(\mathcal{M})$, the space of square integrable functions on \mathcal{M} with the inner product:

$$\langle f, g \rangle = \int_{\mathcal{M}} f(p)g(p) d\mu(p), \quad (2)$$

where μ is the Lebesgue measure such that $\mu(\mathcal{M})$ is the total area of \mathcal{M} .

Imaging data such as electroencephalography (EEG), magnetoencephalography (MEG) (Han et al., 2007) and functional-MRI (Hagler et al., 2006; Jo et al., 2007), and anatomical data such as cortical curvatures (Luders et al., 2006b; Gaser et al., 2006), cortical thickness (Luders et al., 2006a; Bernal-Rusiel et al., 2008) and surface coordinates (Chung et al., 2005) can be considered as functional measurements. Functional measurements are expected to be noisy and require filtering to boost signal.

Surface measurements have often been filtered using the isotropic diffusion equation of the form (Andrade et al., 2001; Chung et al., 2001; Cachia et al., 2003a; Rosenberg, 1997)

$$\frac{\partial f}{\partial \sigma} = \Delta f, \quad f(p, \sigma = 0) = Y(p), \quad (3)$$

where Δ is the Laplace–Beltrami operator defined on manifold \mathcal{M} . The diffusion time σ controls the amount of smoothing. It can be shown that the unique solution of (3) is given by a kernel convolution as follows. A Green's function or a fundamental solution of the Cauchy problem (3) is given by the solution of the following equation:

$$\frac{\partial f}{\partial \sigma} = \Delta f, \quad f(p, \sigma = 0) = \delta(p), \quad (4)$$

where δ is the Dirac delta function. The heat kernel K_σ is a Green's function of (4) (Evans, 1998), i.e.,

$$\frac{\partial K_\sigma}{\partial \sigma} = \Delta K_\sigma, \quad K_\sigma(p, \sigma = 0) = \delta(p).$$

Since the differential operators are linear in (4), we can further convolve the terms with the initial data Y such that

$$\frac{\partial}{\partial \sigma}(K_\sigma * Y) = \Delta(K_\sigma * Y), \quad K_\sigma * Y(p, \sigma = 0) = Y(p),$$

where

$$K_\sigma * Y(p) = \int_{\mathcal{M}} K_\sigma(p, q)Y(q) d\mu(q).$$

Hence $K_\sigma * Y$ is a solution of (3).

2.2. Diffusion smoothing

Isotropic diffusion (3) has been solved by various numerical techniques (Chung, 2001; Andrade et al., 2001; Cachia et al., 2003a,b; Chung and Taylor, 2004). For diffusion smoothing, the diffusion equation needs to be discretized using the cotan formulation (Chung, 2001; Chung and Taylor, 2004; Qiu et al., 2006). Since there are many different cotan formulations, we follow the formulation introduced in Chung (2001). The diffusion Eq. (3) is discretized as

$$\frac{\partial \mathbf{f}}{\partial \sigma} = -\mathbf{A}^{-1}\mathbf{C}\mathbf{f}, \quad (5)$$

where $\mathbf{f} = (f(p_1, \sigma), \dots, f(p_n, \sigma))'$ is the vector of measurements over all mesh vertices at time σ . $\mathbf{A} = (A_{ij})$ is the stiffness matrix and $\mathbf{C} = (C_{ij})$ is the global coefficient matrix, which is the assemblage of individual element coefficients. The sparse matrices \mathbf{A} and \mathbf{C} are explicitly given as follows.

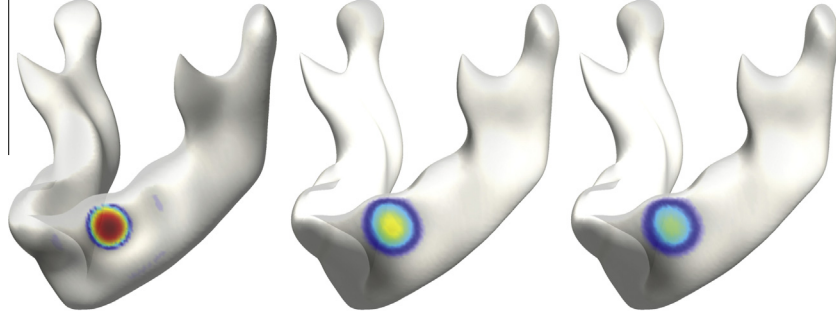


Fig. 1. The heat kernel shape with bandwidths 0.025 (left), 1.25 (middle) and 5 (right) on a mandible surface. The level sets of the heat kernel form geodesic circles.

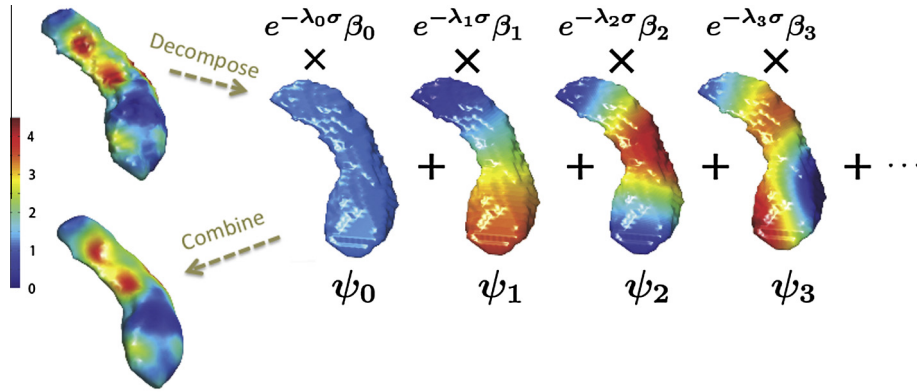


Fig. 2. Schematic of heat kernel smoothing. Given functional data on a surface, we first compute the eigenfunctions ψ_j and the Fourier coefficients β_j . Then, we combine all the terms and reconstruct the functional signal back.

Let T_{ij}^- and T_{ij}^+ denote two triangles sharing the vertex p_i and its neighboring vertex p_j in a mesh. Let the two angles opposite to the edge containing p_i and p_j be ϕ_{ij} and θ_{ij} respectively for T_{ij}^+ and T_{ij}^- . The off-diagonal entries of the stiffness matrix are

$$A_{ij} = \frac{1}{12} (|T_{ij}^+| + |T_{ij}^-|)$$

if p_i and p_j are adjacent and $A_{ij} = 0$ otherwise. $|\cdot|$ denotes the area of a triangle. The diagonal entries are summed as $A_{ii} = \sum_{j=1}^n A_{ij}$. The off-diagonal entries of the global coefficient matrix are

$$C_{ij} = -\frac{1}{2} (\cot \theta_{ij} + \cot \phi_{ij})$$

if p_i and p_j are adjacent and $C_{ij} = 0$ otherwise. The diagonal entries are similarly given as the sum $C_{ii} = -\sum_{j=1}^n C_{ij}$.

The ordinary differential Eq. (5) is then further discretized at each point using the forward finite difference scheme:

$$\mathbf{f}(p_i, \sigma_{n+1}) = \mathbf{f}(p_i, \sigma_n) + (\sigma_{n+1} - \sigma_n) \hat{\Delta} \mathbf{f}(p_i, \sigma_n), \quad (6)$$

where $\hat{\Delta} \mathbf{f}(p_i, \sigma_n)$ is the estimated Laplacian obtained from the i -th row of $-\mathbf{A}^{-1} \mathbf{C} \mathbf{f}$. For the forward Euler scheme to converge, we need to have a sufficiently small step size $\Delta \sigma = \sigma_{n+1} - \sigma_n$ (Chung, 2001).

2.3. Iterated kernel smoothing

The diffusion Eq. (3) can be approximately solved by iteratively performing Gaussian kernel smoothing (Chung et al., 2005). The weights of the kernel are spatially adapted to follow the shape of the heat kernel along a surface mesh. Heat kernel smoothing with

a large bandwidth can be broken into iterated smoothing with smaller bandwidths (Chung et al., 2005):

$$K_{m\sigma} * Y = \underbrace{K_\sigma * \dots * K_\sigma}_{m \text{ times}} * Y. \quad (7)$$

Then using the parametric expansion (Rosenberg, 1997; Wang, 1997), we approximate the heat kernel with the small bandwidth locally using the Gaussian kernel:

$$K_\sigma(p, q) = \frac{1}{\sqrt{4\pi\sigma}} \exp \left[-\frac{d^2(p, q)}{4\sigma} \right] [1 + O(\sigma^2)], \quad (8)$$

where $d(p, q)$ is the geodesic distance between p and q . For sufficiently small bandwidth σ , all kernel weights are concentrated near the center, so the first neighbors of a given mesh vertex are sufficient for approximation. Unfortunately, this approximation compounds error at each iteration. For numerical implementation, we used the normalized truncated kernel given by

$$W_\sigma(p, q_i) = \frac{\exp \left[-\frac{d^2(p, q_i)}{4\sigma} \right]}{\sum_{j=0}^r \exp \left[-\frac{d^2(p, q_j)}{4\sigma} \right]}, \quad (9)$$

where q_1, \dots, q_r are r neighboring vertices of $p = q_0$. Denote the truncated kernel convolution as

$$W_\sigma * Y(p) = \sum_{i=0}^r W_\sigma(p, q_i) Y(q_i). \quad (10)$$

The iterated heat kernel smoothing is then defined as

$$W_{m\sigma} * Y(p) = \underbrace{W_\sigma * \dots * W_\sigma}_{m \text{ times}} * Y(p).$$

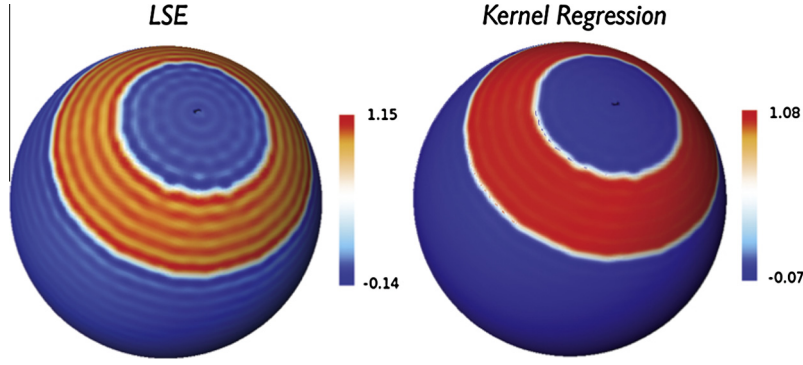


Fig. 3. Gibbs phenomenon (ringing artifacts) is visible in the spherical harmonic series expansion with degree 78 via the least squares estimation (LSE) of the step function defined on a sphere. In contrast, the heat kernel regression with the same degree and bandwidth 0.0001 shows fewer visible artifacts.

2.4. Heat kernel regression

We present a new regression framework for solving the isotropic diffusion Eq. (3). Let Δ be the Laplace–Beltrami operator on \mathcal{M} . Solving the eigenvalue equation

$$\Delta\psi_j = -\lambda_j\psi_j, \quad (11)$$

we order the eigenvalues

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$$

and corresponding eigenfunctions $\psi_0, \psi_1, \psi_2, \dots$ (Rosenberg, 1997; Chung et al., 2005; Lévy, 2006; Shi et al., 2009). The first eigenvalue and eigenfunction are trivially given as $\lambda_0 = 0$ and $\psi_0 = 1/\sqrt{\mu(\mathcal{M})}$. It is possible to have multiple eigenfunctions corresponding to the same eigenvalue.

The eigenfunctions ψ_j form an orthonormal basis in $L^2(\mathcal{M})$. There is extensive literature on the use of eigenvalues and eigenfunctions of the Laplace–Beltrami operator in medical imaging and computer vision (Lévy, 2006; Qiu et al., 2006; Reuter et al., 2009; Reuter, 2010; Zhang et al., 2007, 2010). The eigenvalues have been used in caudate shape discriminators (Niethammer et al., 2007). Qiu et al. used eigenfunctions in constructing splines on cortical surfaces (Qiu et al., 2006). Reuter used the topological features of eigenfunctions (Reuter, 2010). Shi et al. used the Reeb graph of the second eigenfunction in shape characterization and landmark detection in cortical and subcortical structures (Shi et al., 2008, 2009). Lai et al. used the critical points of the second eigenfunction as anatomical landmarks for colon surfaces (Lai et al., 2010). Since the direct application of eigenvalues and eigenfunctions as features of interest is beyond the scope of this paper, we will not pursue the issue in detail here.

Using the eigenfunctions, the *heat kernel* $K_\sigma(p, q)$ is defined as

$$K_\sigma(p, q) = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} \psi_j(p) \psi_j(q), \quad (12)$$

where σ is the bandwidth of the kernel. Fig. 1 shows examples of the heat kernel with different bandwidths. Then the *heat kernel regression* or *heat kernel smoothing* of functional measurement Y is defined as

$$K_\sigma * Y(p) = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} \beta_j \psi_j(p), \quad (13)$$

where $\beta_j = \langle Y, \psi_j \rangle$ are Fourier coefficients (Chung et al., 2005) (Fig. 2). The kernel smoothing $K_\sigma * Y$ is taken as an estimate for the unknown mean signal θ in (1). The degree of truncation of the series expansion can be automatically determined using the forward model selection procedure.

Unlike previous approaches to heat diffusion (Andrade et al., 2001; Chung and Taylor, 2004; Joshi et al., 2009; Tasdizen et al., 2006), the proposed method avoids the direct numerical discretization of the diffusion equation. Instead we discretize the basis functions of the manifold \mathcal{M} by solving for the eigensystem (11) to obtain λ_j and ψ_j .

2.5. Diffusion wavelet transform

We can establish the equivalence between the proposed kernel regression and recently popular diffusion wavelets. This mathematical equivalence eliminates the need for constructing wavelets using complicated computational machinery as has often been done in previous studies (Antoine et al., 2010; Hammond et al., 2011; Kim et al., 2012), and offers a simpler but more unified alternative.

Consider a wavelet basis $W_{\sigma,q}(p)$ obtained from a mother wavelet W with scale and translation parameters σ and q respectively in a Euclidean space:

$$W_{\sigma,q}(p) = \frac{1}{\sigma} W\left(\frac{p-q}{\sigma}\right).$$

Generalizing the idea of scaling a mother wavelet in Euclidean space to a curved surface is trivial. The difficulty arises when one tries to translate a mother wavelet on a curved surface since it is unclear how to define translation along the surface. If one tries to modify the existing spherical wavelet framework to an arbitrary surface (Nain et al., 2007; Bernal-Rusiel et al., 2008), one immediately encounters the problem of establishing regular grids on an arbitrary surface. Recent work based on diffusion wavelets bypass this problem by taking the bivariate kernel as a mother wavelet (Antoine et al., 2010; Hammond et al., 2011; Mahadevan and Maggioni, 2006; Kim et al., 2012).

For some scale function g that satisfies the admissibility conditions, diffusion wavelet $W_{\sigma,q}(p)$ at position p and scale σ is given by

$$W_{\sigma,q}(p) = \sum_{j=0}^k g(\lambda_j \sigma) \psi_j(p) \psi_j(q),$$

where λ_j and ψ_j are eigenvalues and eigenfunctions of the Laplace–Beltrami operator. The wavelet transform is then given by

$$\langle W_{\sigma,q}, Y \rangle = \int_{\mathcal{M}} W_{\sigma,q}(p) Y(p) d\mu(p). \quad (14)$$

By letting $g(\lambda_j \sigma) = \exp(-\lambda_j \sigma)$, we have the heat kernel as the wavelet, i.e.,

$$W_{\sigma,p}(q) = K_\sigma(p, q).$$

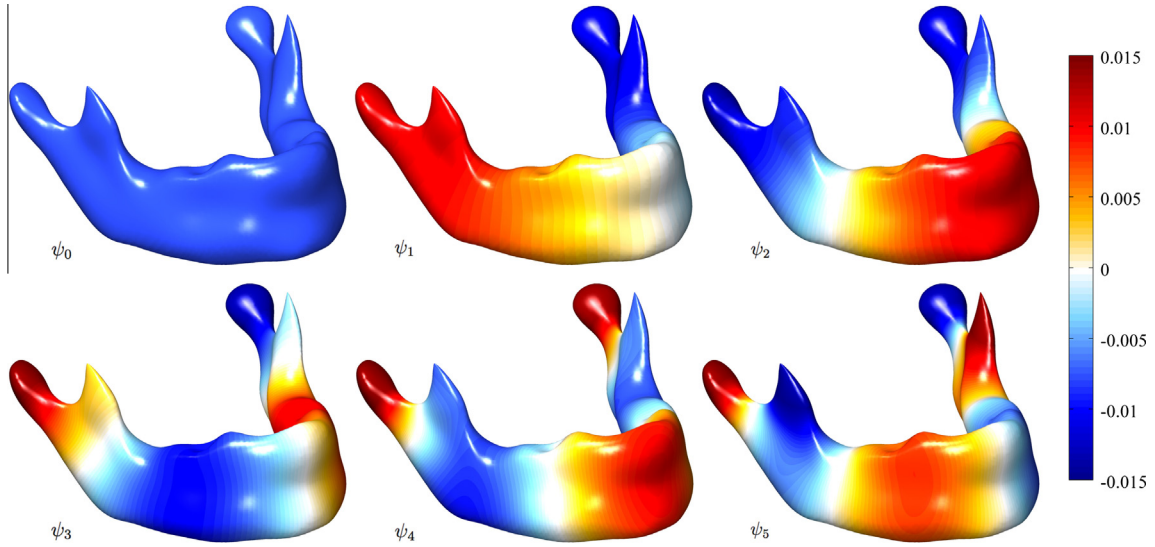


Fig. 4. Eigenfunctions of various degrees for a sample mandible surface. The eigenfunctions are projected on the surface smoothed by the proposed heat kernel smoothing with bandwidth $\sigma = 0.5$ and degree $k = 132$. The smoothed surface is obtained by heat kernel smoothing applied to the coordinates of the surface mesh with the same parameter while preserving the topology of mesh. The first eigenfunction is simply $\psi_0 = 1/\sqrt{\mu(\mathcal{M})}$. The color scale is thresholded for better visualization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The bandwidth σ of the heat kernel is the scale parameter, while the translation is achieved by shifting one argument in the bivariate heat kernel. Subsequently, wavelet transform (14) can then be rewritten as

$$\langle W_{\sigma,p}, Y \rangle = \sum_{j=0}^k e^{-\lambda_j \sigma} \beta_j \psi_j(q), \quad (15)$$

with $\beta_j = \langle Y, \psi_j \rangle$. The expression (15) is the finite truncation of the heat kernel regression in (12). Hence, diffusion wavelet analysis can be simply performed within the proposed heat kernel regression framework without any additional wavelet machinery. We therefore do not distinguish between the heat kernel regression and the diffusion wavelet transform.

Although the heat kernel regression is constructed using the global basis functions ψ_j , surprisingly the kernel regression at each point p coincides with the wavelet transform at that point. Hence, it also inherits the localization property of wavelets at that point. This is clearly demonstrated in the simulation study shown in Fig. 3, where a step function of value 1 in the circular band $1/8 < \theta < 1/4$ (angle from the north pole) and of value 0 outside of the band was constructed. Note that, on a sphere, the Laplace–Beltrami operator is the spherical Laplacian and its eigenfunctions are spherical harmonics Y_{lm} of degree l and order m . Then the step function was reconstructed using the spherical harmonic series expansion

$$Y(p) = \sum_{l=0}^L \sum_{m=-l}^l \beta_{lm} Y_{lm}(p),$$

where the spherical harmonic coefficients $\beta_{lm} = \langle Y, Y_{lm} \rangle$ were obtained by the least squares estimation (LSE). On the unit sphere, we used the heat kernel regression of the form

$$Y(p) = \sum_{l=0}^L \sum_{m=-l}^l e^{-l(l+1)\sigma} \beta_{lm} Y_{lm}(p),$$

with small bandwidth $\sigma = 0.0001$ and degree $L = 78$. The spherical harmonic expansion shows severe ringing artifacts compared to the kernel regression, which inherits the localization power of wavelets. Thus the Gibbs phenomenon was not significantly visible.

2.6. Parameter estimation in heat kernel regression

Since the closed form expression for the eigenfunctions of the Laplace–Beltrami operator on an arbitrary surface is unknown, the eigenfunctions are numerically computed by discretizing the Laplace–Beltrami operator. To solve the eigensystem (11), we need to discretize it on mandible triangular meshes using the cotan formulation (Chung, 2001; Chung and Taylor, 2004; Shi et al., 2009; Qiu et al., 2006; Lévy, 2006; Reuter et al., 2006, 2009; Rostamov, 2007; Zhang et al., 2007; Vallet and Lévy, 2008; Wardetzky, 2008).

Among the many different cotan formulations used in computer vision and medical image analysis, we used the formulation given in Chung (2001) and Qiu et al. (2006). It requires discretizing (11) as the following generalized eigenvalue problem:

$$\mathbf{C}\psi = \lambda \mathbf{A}\psi, \quad (16)$$

where the global coefficient matrix \mathbf{C} is the assemblage of individual element coefficients and \mathbf{A} is the stiffness matrix. We solved (16) using the *Implicitly Restarted Arnoldi Method* (Hernandez et al., 2006; Lehoucq et al., 1998) without consuming large amounts of memory and time for sparse entries. Fig. 4 shows the first few eigenfunctions for a mandible surface.

Once we obtain the eigenfunctions numerically, we estimate the kernel regression parameters β_j using the least squares estimation (LSE) technique. Note $\beta_j = \langle Y, \psi_j \rangle$, the Fourier coefficients with respect to basis ψ_j . $\beta_0, \beta_1, \dots, \beta_k$ are then estimated simultaneously by minimizing the sum of squared residual:

$$\arg \min_{\beta_0, \dots, \beta_k} \left\| Y(p) - \sum_{j=0}^k \beta_j \psi_j(p) \right\|^2. \quad (17)$$

The least squares method is often used in estimating the coefficients in spherical harmonic expansion (Shen et al., 2004; Styner et al., 2006; Chung et al., 2008). Suppose we have n mesh vertices p_1, \dots, p_n . Let

$$\mathbf{Y} = (Y(p_1), \dots, Y(p_n))'$$

be the surface measurements over all n vertices. Denote the j -th eigenfunction evaluated at n vertices as

$$\Psi_j = (\psi_j(p_1), \dots, \psi_j(p_n))'.$$

The minimum in (17) is achieved when

$$\mathbf{Y} = \Psi\beta, \quad (18)$$

where $\Psi = (\Psi_0, \dots, \Psi_k)$ is the matrix of size $n \times (k+1)$. The LSE estimation of coefficients β is then given by

$$\hat{\beta} = (\Psi'\Psi)^{-1}\Psi'\mathbf{Y}. \quad (19)$$

Since it is expected that the number of mesh vertices is substantially larger than the number of eigenfunctions to be used, $\Psi'\Psi$ is well conditioned and invertible. The numerical implementation is available at <http://www.stat.wisc.edu/mchung/mandible> with the full data set used in the study.

2.7. Random field theory (RFT)

Once we have smoothed functional data on a surface, we apply the statistical parametric mapping (SPM) framework for analyzing and visualizing statistical tests performed on the template surface that is often used in structural neuroimaging studies (Andrade et al., 2001; Lerch and Evans, 2005; Wang et al., 2010; Worsley et al., 1995; Yushkevich et al., 2008). Since test statistics are constructed over all mesh vertices on the mandible, multiple comparisons need to be accounted for using the RFT (Taylor and Worsley, 2007; Worsley et al., 1995, 2004). The RFT assumes the measurements to be a smooth Gaussian random field. Heat kernel smoothing will make data smoother and more Gaussian and enhance the SNR (Chung et al., 2005). The proposed kernel regression framework can then be naturally integrated into the RFT-based statistical inference approach (Taylor and Worsley, 2007; Worsley et al., 2004, 1995).

We assume θ is an unknown group level signal and ϵ is a zero-mean unit-variance Gaussian random field in (1). The model assumptions are not as restrictive as it seems since we can always normalize the data in this fashion. We further assume the random field ϵ is the convolution of heat kernel K_s on Gaussian white noise W with bandwidth s , i.e., $\epsilon(p) = K_s * W(p)$. Previously, the smoothness of noise, i.e., kernel bandwidth s , was estimated using the RESEL (resolution element) technique, which requires estimating the quantity $\text{Var}[\partial\epsilon(p)/\partial p]$ along mesh surfaces, which can introduce a bias (Worsley et al., 1999; Hayasaka et al., 2004; Kilner and Friston, 2010). Thus, surface data is often smoothed with bandwidth σ that is sufficiently larger than s so that any high frequency noise smaller than σ is masked out. This provides a motivation for developing the heat kernel regression framework.

In (1), we are interested in determining the significance of θ , i.e.,

$$\begin{aligned} H_0 : \theta(p) &= 0 \quad \text{for all } p \in \mathcal{M}, \\ \text{vs. } H_1 : \theta(p) &> 0 \quad \text{for some } x \in \mathcal{M}. \end{aligned} \quad (20)$$

Note that any point p_0 that gives $\theta(p_0) > 0$ is considered as signal. The hypothesis (20) is an infinite dimensional multiple comparisons problem for continuously indexed hypotheses over the manifold \mathcal{M} . The underlying group level signal h is estimated using the proposed heat kernel regression. Subsequently, a test statistic is often given by a T - or F -field $Y(p)$ (Worsley et al., 1995, 2004).

The multiple comparisons corrected p -value is then computed through by the RFT (Adler, 1981; Cao and Worsley, 2001; Taylor and Worsley, 2007; Worsley, 2003). For the F -field Y with α and β degrees of freedom defined on 2D manifolds \mathcal{M}_F , it is known that

$$P\left(\sup_{p \in \mathcal{M}_F} Y(p) > h\right) \approx \mu_2(\mathcal{M}_T)\rho_2(h) + \mu_0(\mathcal{M}_F)\rho_0(h) \quad (21)$$

for a sufficiently large threshold h , where $\mu_d(\mathcal{M}_F)$ is the d -th Minkowski functional of \mathcal{M}_F and ρ_d is the d -th Euler characteristic

(EC) density of Y (Worsley et al., 1998). The Minkowski functionals are given by

$$\begin{aligned} \mu_2(\mathcal{M}_T) &= \text{area}(\mathcal{M}_T)/2, \\ \mu_0(\mathcal{M}_T) &= \chi(\mathcal{M}_T) = 2. \end{aligned}$$

The EC-density for F -field is then given by

$$\begin{aligned} \rho_2 &= \frac{1}{4\pi\sigma^2} \frac{\Gamma(\frac{\alpha+\beta-2}{2})}{\Gamma(\frac{\alpha}{2})\Gamma(\frac{\beta}{2})} \left(\frac{\alpha h}{\beta}\right)^{\frac{(\alpha-2)}{2}} \left(1 + \frac{\alpha h}{\beta}\right)^{-\frac{(\alpha+\beta-2)}{2}} \left[(\beta-1)\frac{\alpha h}{\beta} - (\alpha-1)\right], \\ \rho_0 &= 1 - P(F_{\alpha,\beta} \leq h), \end{aligned}$$

where $P(F_{\alpha,\beta} \leq h)$ is the cumulative distribution function of F -stat with α and β degrees of freedom. The second order term $\mu_2(\mathcal{M}_T)\rho_2(h)$ dominates the expression (21) and it explicitly has the bandwidth σ of the kernel regression, thus incorporating the proposed kernel framework into the RFT.

3. Experiments

3.1. CT image preprocessing

We applied the proposed smoothing method to CT images of mandibles obtained from several different models of GE multi-slice helical CT scanners. The CT scans were acquired directly in the axial plane with 1.25 mm slice thickness, matrix size of 512×512 and 15–25 cm field of view (FOV). Image resolution varied as voxel size ranged from 0.25 mm^3 to 0.49 mm^3 as determined by the ratio of FOV divided by the matrix. CT scans were converted to DICOM format and Analyze 8.1 software package (AnalyzeDirect, Inc., Overland Park, KS) was then used in segmenting binary mandible structure based on histogram thresholding.

Image acquisition and processing artifacts and partial voluming produce topological defects such as holes and handles in any medical image. In CT images of mandibles, unwanted cavities, holes and handles in the binary segmentation mainly result from differences in CT intensity between relatively low density mandible and teeth and more dense cortical bone and the interior trabecular bone (Andresen et al., 2000; Loubele et al., 2006). In mandibles, these topological noises can appear in thin or cancellous bone, such as in the condylar head and posterior palate (Stratemann et al., 2010). An example is shown in Fig. 5, where the tooth cavity forms a bridge over the mandible. If we apply the isosurface extraction on the topologically defective segmentation results, the resulting surface will have many tiny handles (Wood et al., 2004; Yotter et al., 2009). These handles complicate subsequent surface mesh operations such as smoothing and parameterization. It is thus necessary to correct the topology by filling the holes and removing handles. If we correct such topological defects, it is expected that the resulting isosurface is topologically equivalent to a sphere.

Various topology correction techniques have been proposed in medical image processing. Rather than attempting to repair the topological defects of the already extracted surfaces (Wood et al., 2004; Yotter et al., 2009), we performed the topological simplification on the volume representation directly using morphological operations (Guskov and Wood, 2001; Van Den Boomgaard and Van Balen, 1992; Yotter et al., 2009). The direct correction on surface meshes can possibly cause surfaces to intersect each other (Wood et al., 2004). By checking the Euler characteristic, the holes were automatically filled up using morphological operations to make the mandible binary volume topologically equivalent to a solid sphere. All areas enclosed by the higher density bone included in the mandible definition are morphed into being in the definition of the mandible object. The hole-filled images were then converted to surface meshes via the marching cubes algorithm.

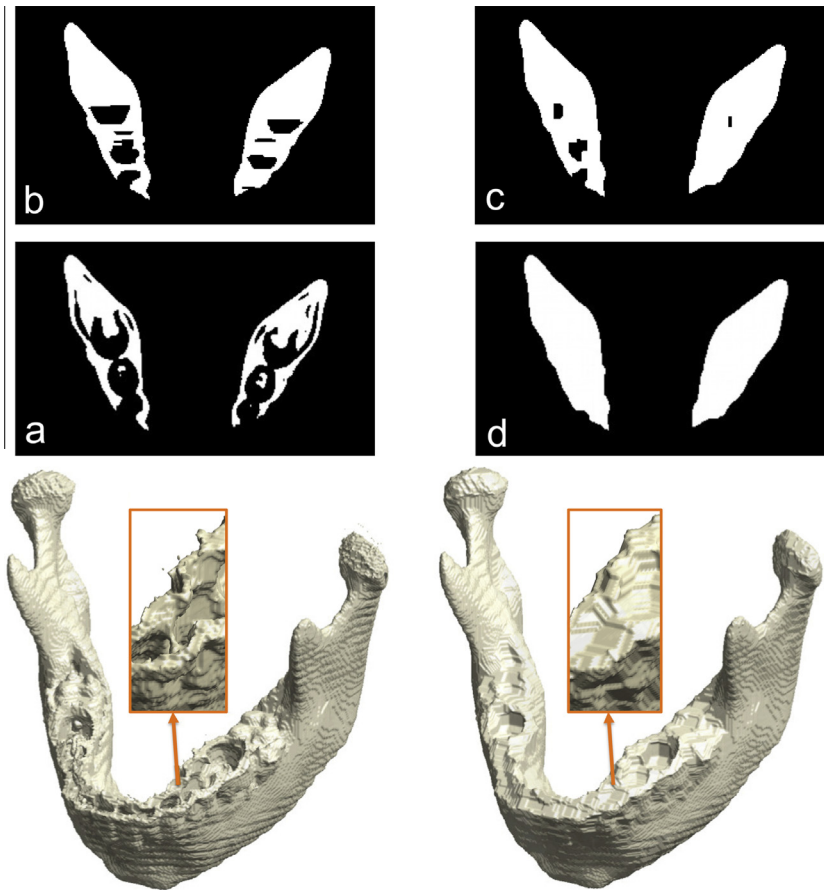


Fig. 5. Upper panel: Topological correction on mandible binary segmentation and surface. Disjointed tiny speckles of noisy components are removed by labeling the largest connected component, and holes and handles are removed by the 2D morphological closing operations applied sequentially to each image dimension (a \rightarrow b \rightarrow c \rightarrow d). Lower panel left: Surface reconstruction showing holes and handles in the teeth regions. The isosurface has Euler characteristic $\chi = 50$. Lower panel right: After the correction with $\chi = 2$.

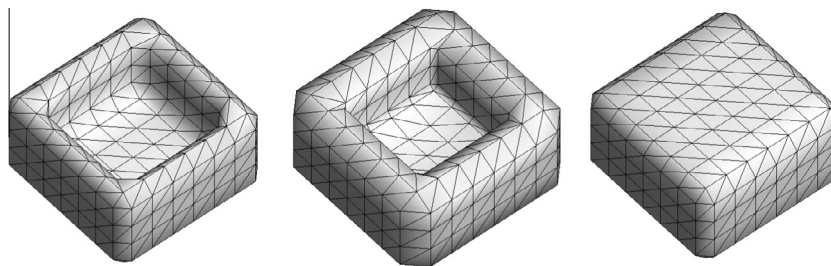


Fig. 6. Cavity patching by topological closing operations. Left: Surface model of the binary volume that simulates a tooth cavity. Middle: The 3D image volume based closing operation does not properly patch the cavity region. Right: The 2D image slice based closing operation patches the cavity region properly.

In our semi-automated algorithm, we first removed the speckles of noise components by identifying the largest connected component in the binary volume. The resulting binary mandible volume was a single connected component with many small holes and handles. Then we applied the morphological closing operation in each 2D slice of CT images one by one in all three axes. Recombining the topology-corrected 2D slices resulted in topologically correct surface meshes (Fig. 5). We used 2D topological closing operations mainly because of better performance and relatively simpler implementation than 3D topological closing operations. In 2D topological operations, we need to consider only 8 neighboring pixels compared to 26 neighboring voxels in a 3D image volume. There are many large concave regions left out by teeth and fillings. These regions may not be closed with a single 3D closing operation

but can be easily patched up with a sequence of 2D closing operations, which put more constraints on the underlying topology. Instead of performing a single 3D closing operation that may not work, we sequentially performed 2D closing operations in each image slice in the x -, y - and z -directions. In Fig. 5 upper panel, (a) is the binary volume before any closing operation, (b) is after the 2D closing operation in the x -direction, (c) is after the 2D closing operation in the y -direction and (d) is after the 2D closing operation in the z -direction. Each time we apply the 2D closing operation, the holes get smaller. Fig. 6 shows a simulated cavity example that was not patched by the 3D closing operation (Van Den Boomgaard and Van Balen, 1992) but was easily patched by the sequential application of 2D closing operations. Note that any 3D object, whose every 2D cross-section is topologically equivalent

to a solid disk, is topologically equivalent to a solid sphere. The problem of 3D topology correction can be thus reduced to a much simpler problem of 2D topology correction of multiple slices. Unfortunately, we cannot perform the closing operations to infinitely many possible 2D cross-sections in 3D image volumes. Therefore, we applied the 2D operations in the three axial directions. So there is a small chance the operation may not work in practice. Therefore, at the end of the processing, we performed a visual inspection of the processed volume. Further, we double checked the Euler characteristic of the resulting surface meshes. Note that for each triangle, there are three edges. For a closed surface topologically equivalent to a sphere, two adjacent triangles share the same edge. The total number of edges E is thus $E = 3F/2$, where F is the number of faces. If V is the total number of vertices, the Euler characteristic of a sphere is given by $\chi = V - E + F = 2$. Thus, we checked if the resulting mesh satisfies the condition $V - F/2 = 2$. 77 binary mandible volumes used in the study produced the topologically correct surfaces without exception. Fig. 5 lower panel shows an example of before and after the topology correction.

3.2. Validation of heat kernel smoothing on mandible surfaces

The accuracy of the heat kernel construction using LB-eigenfunctions on a unit sphere using the ground truth can be found in Kim et al. (2011) so the results are not reproduced here. In this paper, we compared the performance of the proposed kernel regression against iterated kernel smoothing and diffusion smoothing using the real mandible data. The comparison results are similar for all 77 mandible surfaces used in the study, so only the results for one representative mandible surface is shown.

The x , y and z coordinates of a mandible surface are treated as functional measurements on the original surface and smoothed with both methods. For the comparison of performance between the smoothing methods, we calculated the root mean squared errors (RMSE) between them, where the mean of the squared errors is taken over the surface. For the heat kernel regression, we used the bandwidth $\sigma = 0.5$ and eigenfunctions up to $k = 132$ degree. For iterated kernel smoothing, we varied the number of iterations $1 \leq m \leq 200$ with correspondingly smaller bandwidth $0.5/m$, which results in an effective bandwidth of 0.5. For diffusion smoothing, a sufficiently small step size $\Delta\sigma = 0.0025$ was taken for 200 iterations resulting in bandwidth $\sigma = 0.5$. The RMSE between the kernel regression and the iterated kernel smoothing reached up

to 0.5901 (y -coordinate) and did not decrease even when we increased the number of iterations (Fig. 7). The RMSE between the kernel regression and diffusion smoothing was smaller than 0.0046 (y -coordinates). Fig. 7(b) shows the squared differences between the two methods. For the iterated kernel smoothing, the difference is mainly localized in high curvature areas, where the Gaussian kernel used in the iterated kernel smoothing fails to approximate the heat kernel. This comparison clearly demonstrates the limitation of iterated heat kernel smoothing, which does not converge to heat diffusion. However, the heat kernel regression and diffusion smoothing gave almost identical results and there was no discernible difference (Fig. 7(c)).

We also compared the performance of the three smoothing techniques at four different bandwidths $\sigma = 0.5, 20, 50, 100$. For the kernel regression, $k = 132$ was used. For the iterated kernel smoothing and the diffusion smoothing, a fixed step size of $\Delta\sigma = 0.025$ was used with $m = 20, 800, 2000, 4000$ iterations. The diffusion smoothing and heat kernel smoothing gave visually identical results for bandwidths $\sigma = 20, 50, 100$ due to a sufficiently large number of iterations (Fig. 8). However, the iterated kernel smoothing gave a different result. In this experiment, we replaced the original surface coordinates with smoothed ones for the final visualization. However, in the actual computation, we did not replace the original surface coordinates for the three methods. Iterated kernel smoothing compounded the discretization errors over iterations, so it did not converge to the kernel regression and diffusion smoothing. Diffusion smoothing and heat kernel smoothing share the same FEM discretization and converge to each other as the number of iterations increases.

3.3. Simulation studies

Since there is no known ground truth in the imaging data set we are using, it is uncertain how the proposed method will perform with real data. It is therefore necessary to perform simulation studies with ground truths. We performed two simulations with small and large SNR settings on a T-junction shaped surface (Fig. 9), which was chosen because it was a surface with three different curvatures: convex, concave and almost flat regions. Note that surface smoothing methods perform differently under different curvatures. Three black signal regions of different sizes were taken as the ground truth at these regions and 60 independent functional measurements on the T-junction were simulated as $|N(0, \gamma^2)|$, the

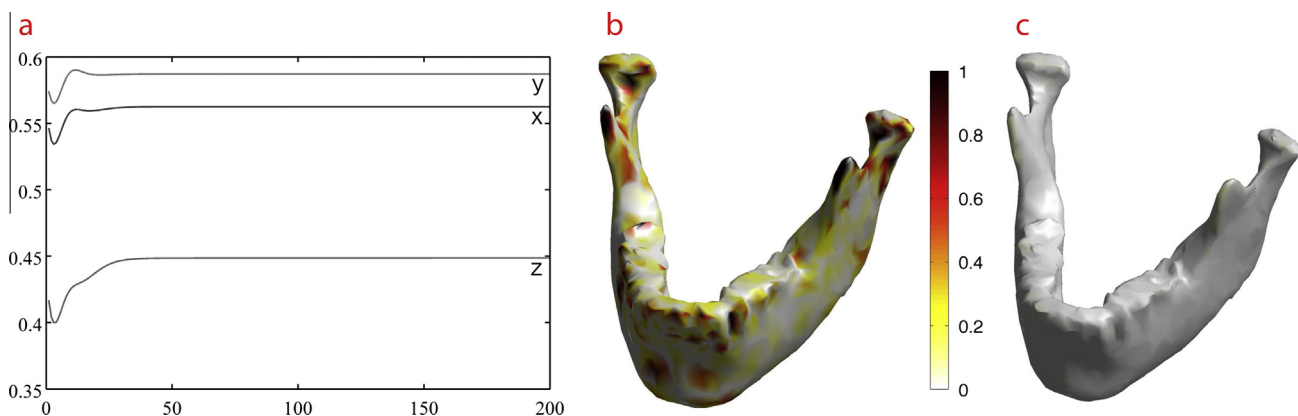


Fig. 7. (a) Plot of the RMSE of the heat kernel regression against iterated kernel smoothing for coordinates x (middle), y (top) and z (bottom) over the number of iterations up to 200. For the heat kernel regression, $\sigma = 0.5$ and $k = 132$ are used. Iterated kernel smoothing does not converge to heat diffusion. RMSE between the kernel regression and the diffusion smoothing is smaller than 0.0046 so they are not displayed in the plot. (b) The squared difference between the kernel regression and the iterated kernel smoothing. The difference is mainly localized in high curvature areas, where the Gaussian kernel used in the iterated kernel smoothing fails to approximate the heat kernel. (c) The squared difference between the kernel regression and the diffusion smoothing. There are almost no visible differences.

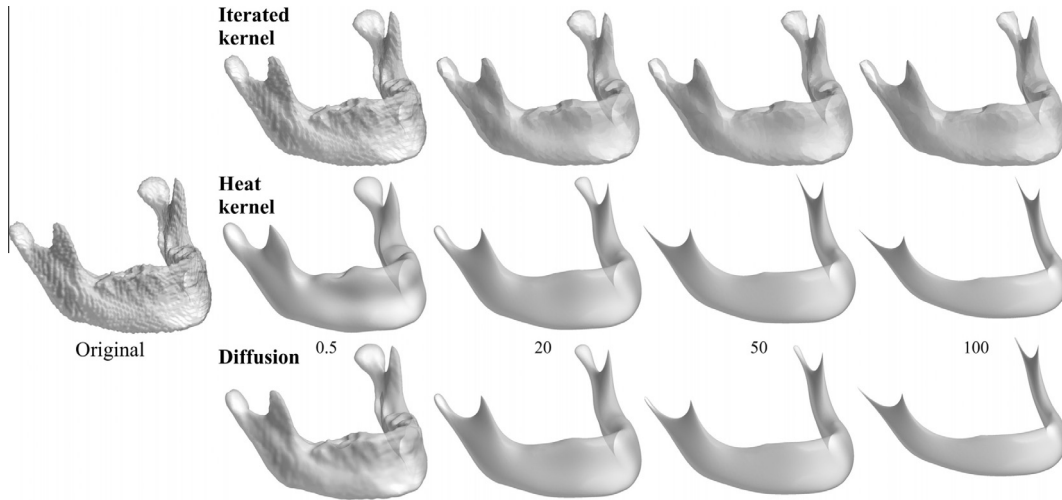


Fig. 8. Smoothed mandible surfaces using three different techniques. The x , y and z surface coordinates are treated as functional measurements on the original surface and smoothed. The proposed heat kernel smoothing is done with bandwidths, $\sigma = 0.5, 20, 50, 100$. Iterated kernel smoothing is done by iteratively approximating the heat kernel linearly with the Gaussian kernel (Chung et al., 2005). Diffusion smoothing directly solves the diffusion equation using the same FEM discretization (Chung and Taylor, 2004). Diffusion smoothing and heat kernel smoothing converge to each other as the bandwidth increases.

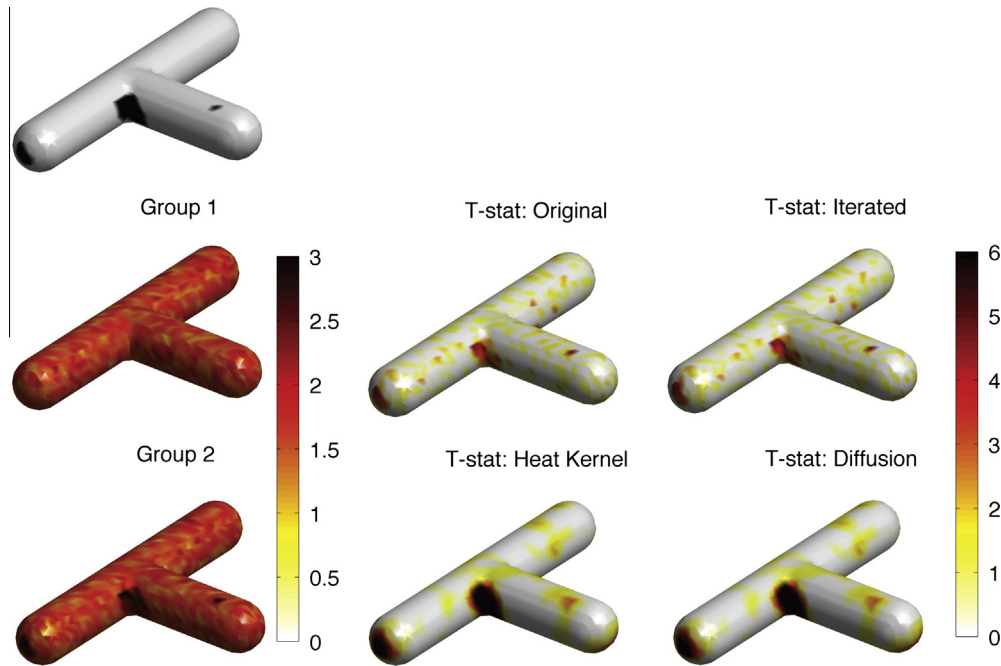


Fig. 9. Simulation study I on a T-junction shaped surface where three black signal regions of different sizes are taken as the ground truth. 60 independent functional measurements on the T-junction were simulated as $|N(0, 2^2)|$ at each mesh vertex. We are only simulating positive numbers to better reflect the positive measurements used in the study. Value 1 was added to the black regions in 30 measurements, which served as group 2 while the other 30 measurements were taken as group 1. T-statistics are shown for these simulations (original) and three techniques with bandwidth 0.5. Heat kernel smoothing performed the best in detecting the ground truth.

absolute value of normal distribution with mean 0 and variance γ^2 , at each mesh vertex. Value 1 was then added to the black regions in 30 of the measurements, which served as group 2, while the other 30 measurements were taken as group 1. Group 1 had distribution $|N(0, \gamma^2)|$ while group 2 had distribution $|N(1, \gamma^2)|$ in the signal regions. Larger variance γ^2 corresponds to smaller SNR.

In Study I, $\gamma^2 = 2^2$ was used to simulate functional measurements with substantially smaller SNR. Fig. 9 shows the simulation results. For iterated kernel and diffusion smoothing, we used bandwidth $\sigma = 0.5$ and 100 iterations. For smaller SNR, it is necessary to smooth with a larger bandwidth, which is determined empirically.

For heat kernel smoothing, the same bandwidth and 1000 eigenfunctions were used. The same number of eigenfunctions was used throughout the study. For all three smoothing techniques, the bandwidth is the main parameter that determines performance. We then performed a two sample T with the RFT-based threshold of 4.90 to detect the group difference at 0.05 level.

Neither the raw data nor iterated smoothing were able to correctly identify any signal region. However, heat kernel and diffusion smoothing correctly identified 94% and 91% of the signal regions respectively. In addition, heat kernel and diffusion smoothing incorrectly identified 0.26% and 0.26% of non-signal regions as

signal. There are no visually discernible differences between the two methods as shown in Fig. 10. The 3% difference in performance is due to the discretization error associated with taking only 100 iterations, which disappears if we take smaller time steps. Alternatively, we can use better time-discretization schemes such as Padé–Chebyshev approximation (Patané, 2015), multi-step methods (Gottlieb et al., 2001) or higher-order Runge–Kutta schemes (Li and Alexiades, 2010).

In Study II, $\gamma^2 = 0.5^2$ was used to simulate functional measurements with substantially larger SNR. Due to the large SNR, the group means showed visible group separations (Fig. 10). For iterated kernel and diffusion smoothing, we used bandwidth $\sigma = 0.1$ and 100 iterations. For heat kernel smoothing, the same bandwidth and 1000 eigenfunctions were used. All the methods detected the signal regions; however, the heat kernel smoothing and diffusion smoothing techniques were more sensitive at large SNR. All the methods correctly identified the signal regions with 100% accuracy. There were no false discoveries in the raw data and iterated kernel smoothing methods. However, due to blurring effects, heat kernel and diffusion smoothing incorrectly identified 0.9% and 0.8% of non-signal regions as signal, which is negligible. For the large SNR setting, all the methods were reasonably able to detect the correct signal regions with minimal error.

In summary, in larger SNR, all three methods performed well. However, in substantially smaller SNR, the kernel regression performed best, closely followed by diffusion smoothing. Neither the raw data nor iterated kernel smoothing performed well in the low SNR setting.

4. Application: mandible growth analysis

As an illustration of the proposed kernel regression technique, we analyzed mandible growth on a CT imaging data set consisting of 77 human subjects between the ages of 0 and 19 years. Subjects were divided into three age categories: 0–6 years (group I, 26 subjects), 7–12 years (group II, 20 subjects), and 13–19 years (group III, 31 subjects). The main biological question of interest was whether there were localized regions of growth between these different age groups. Mandible surface meshes for all subjects were constructed through the image acquisition and processing steps described in the previous section. For surface alignment, diffeomorphic surface registration was used to match mandible surfaces

across subjects (Miller and Qiu, 2009; Vaillant et al., 2007; Qiu and Miller, 2008; Yang et al., 2011).

4.1. Diffeomorphic surface registration

We chose the mandible of a 12-year-old subject identified as F155–12–08, which served as the reference template in previous studies (Seo et al., 2010, 2011), as initial template \mathcal{M}_I and aligned the remaining 76 mandibles to the initial template affinely to remove the overall size variability. Some subjects may have larger mandibles than others, so it is necessary to remove the global size differences in localized shape modeling. From the affine transformed individual mandible surfaces \mathcal{M}_j , we performed an additional nonlinear surface registration to the template using the large deformation diffeomorphic metric mapping (LDDMM) framework (Miller and Qiu, 2009; Vaillant et al., 2007; Qiu and Miller, 2008; Yang et al., 2011).

In the LDDMM framework (Miller and Qiu, 2009; Vaillant et al., 2007; Qiu and Miller, 2008; Yang et al., 2011), the metric space is constructed as an orbit of surface \mathcal{M} under the group of diffeomorphic transformations \mathcal{G} , i.e., $\mathcal{M}_j = \mathcal{G} \cdot \mathcal{M}$. The diffeomorphic transformations (one-to-one, smooth forward and inverse transformation) are introduced as transformations of the coordinates on the background space $\Omega \subset \mathbb{R}^3$. The diffeomorphisms $\phi_t \in \mathcal{G}$ are constructed as a flow of ordinary differential equations (ODE), where ϕ_t , $t \in [0, 1]$ follows

$$\dot{\phi}_t = v_t(\phi_t), \quad \phi_0 = \text{Id}, \quad t \in [0, 1], \quad (22)$$

where Id denotes the identity map and v_t are the associated velocity vector fields. The vector fields v_t are constrained to be sufficiently smooth, so that (22) is integrable and generates diffeomorphic transformations over finite time. The smoothness is ensured by forcing v_t to lie in a smooth vector field V , which is modeled as a reproducing kernel Hilbert space with linear operator L associated with norm $\|u\|_V^2 = \langle Lu, u \rangle_2$ (Dupuis et al., 1998). The group of diffeomorphisms $\mathcal{G}(V)$ is then the solutions of (22) with the vector fields satisfying $\int_0^1 \|v_t\|_V dt < \infty$.

Given the template surface \mathcal{M} and an individual surface \mathcal{M}_j , the geodesic ϕ_t , $t \in [0, 1]$, which lies in the manifold of diffeomorphisms and connects \mathcal{M} and \mathcal{M}_j , is defined as

$$\phi_0 = \text{Id}, \quad \phi_1 \cdot \mathcal{M} = \mathcal{M}_j.$$

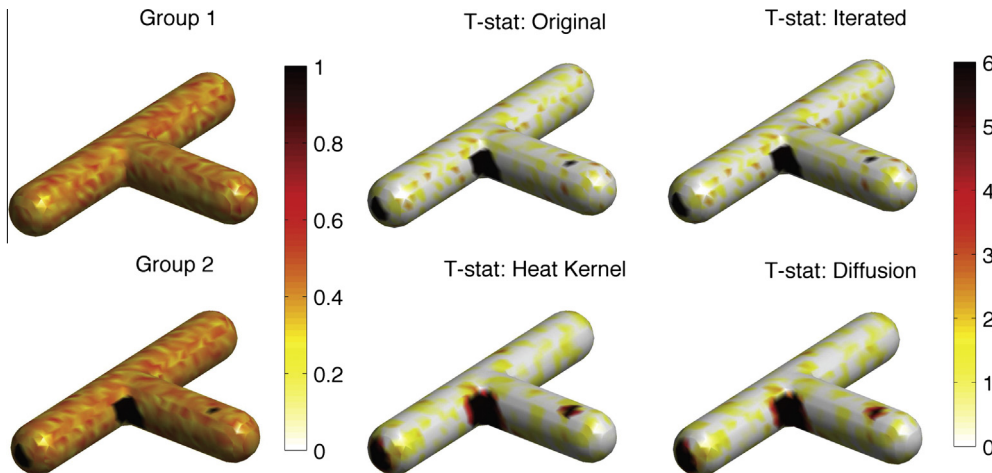


Fig. 10. Simulation study II on a T-junction shaped surface with the same ground truth as simulation study I (Fig. 9). 60 independent functional measurements on the T-junction were simulated as $|N(0, 0.5^2)|$ at each mesh vertex. Value 1 was added to the black regions in 30 of the measurements, which served as group 2, while the remaining 30 measurements are taken as group 1. Due to the large SNR, the group means show visible group separations. All the methods detected the signal regions; however, the heat kernel smoothing and diffusion smoothing techniques were more sensitive at the large SNR.

For our application, we employed the LDDMM approach to estimate the template among all subjects. The estimated template can be simply computed through averaging the initial velocity across all subjects (Zhong and Qiu, 2010), which is similar to the unbiased template estimation approach in Joshi et al. (2004). We then recomputed the displacement fields with respect to the initial template \mathcal{M}_I . We averaged the deformation fields from the initial template \mathcal{M}_I to individual subjects to obtain the final template \mathcal{M}_F . Fig. 11 shows the initial and final templates. Fig. 12 shows the mean displacement differences between groups I and II (top) and II and III (bottom). Each row shows the group differences of the displacement: group II–group I (first row) and group III–group II (second row). The arrows are the growth direction with arrow length being representative of mean displacement differences and colors indicating growth length in mm.

4.2. Statistical analysis

We are interested in determining the significance of the mean displacement differences in Fig. 12. Since the length measurement provides a much easier biological interpretation, we used the length of the displacement vector as a response variable. The RFT assumes the measurements to be a smooth Gaussian random field. Heat kernel smoothing on the length measurement will make the measurement smoother, more Gaussian and increase the SNR

(Chung et al., 2005). Heat kernel smoothing is applied with bandwidth $\sigma = 20$ using 1000 eigenfunctions on the final template \mathcal{M}_F . The number of eigenfunctions used is more than sufficient to guarantee a relative error less than 0.3%. The heat kernel smoothing of the displacement length is taken as the response variable. We constructed the F random field testing the length difference between the age groups I and II, and II and III showing the regions of accelerated growth (Fig. 13).

The comparison of groups I and II is based on an F -field with 1 and 44 degrees of freedom. The comparison of groups II and III is based on an F -field with 1 and 49 degrees of freedom. The multiple comparison corrected F -statistics thresholds corresponding to $\alpha = 0.05$ and 0.01 levels are respectively 8.00 and 10.52 (group II–I) and 8.00 and 10.67 (group III–II). In the F -statistic map shown in Fig. 13, black and red regions are considered as exhibiting growth spurts at 0.01 and 0.05 levels respectively. Our findings are consistent with previous findings of simultaneous forward and downward growth (Scott, 1976; Smartt et al., 2005; Walker and Kowalski, 1972; Lewis et al., 1982; Seo et al., 2011) and bilateral growth (Enlow and Hans, 1996).

We also performed the same statistical analysis to the iterated kernel smoothing and diffusion smoothing results. For the diffusion smoothing, 200 step sizes are used with the fixed time step 0.01, which results in the overall bandwidth $\sigma = 20$. For the iterated kernel smoothing, bandwidth $\sigma = 20$ is split into $m = 200$



Fig. 11. Left: Mandible F155-12-08, which forms an initial template \mathcal{M}_I . All other mandibles were affine registered to F155-12-08. Middle: The superimposition of affine registered mandibles showing local misalignment. Diffeomorphic registration was then performed to warp misaligned affine transformed mandibles. Right: The average of deformation with respect to F155-12-08 provides the final population average template \mathcal{M}_F where statistical parametric maps were constructed.

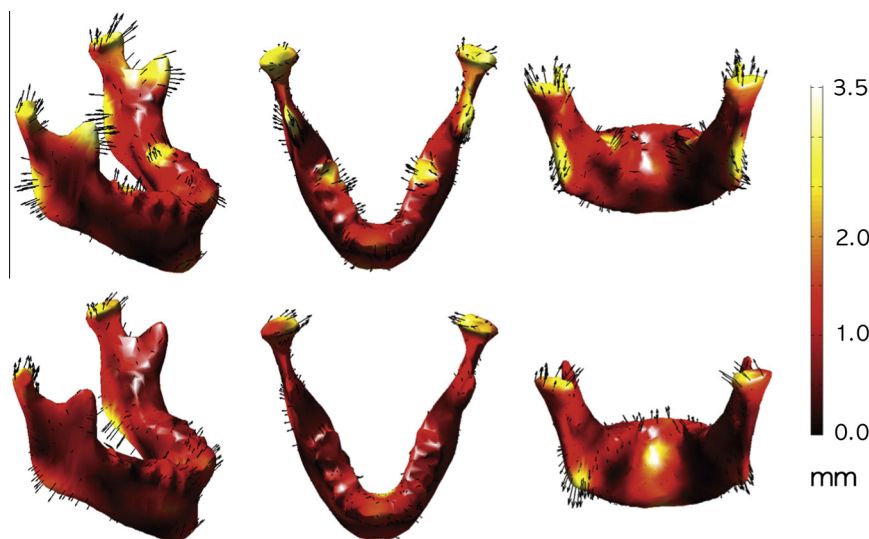


Fig. 12. Mandibles were grouped into three age cohorts: group I (ages 0–6 years), group II (ages 7–12 years) and group III (ages 13–19 years). Each row shows the mean group differences of the displacement: group II–group I (first row) and group III–group II (second row). The arrows are the mean displacement differences and colors indicate their lengths in mm. Longer arrows imply more mean displacement. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

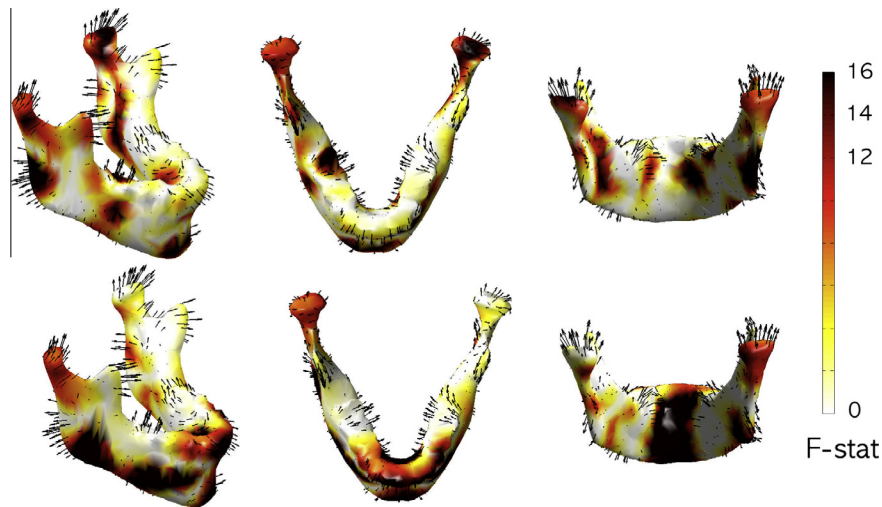


Fig. 13. *F*-statistic map showing the regions of significant growth as measured by mean displacement differences between the groups displayed in Fig. 12. The kernel regression was used to smooth out surface measurements. The top row shows significant growth between groups I and II; and bottom row between groups II and III. The thresholds 10.52 and 10.67 are considered significant at 0.01 level (corrected) for the top and bottom rows.



Fig. 14. *F*-statistic map showing the regions of significant growth as measured by mean displacement differences between the groups displayed in Fig. 12. The iterated kernel smoothing with parameters $\sigma = 20$ and $m = 200$ were used. The top row shows significant growth between groups I and II; and bottom row between groups II and III. The thresholds 10.52 and 10.67 are considered significant at 0.01 level (corrected) for the top and bottom rows.

small bandwidths. The diffusion smoothing results are similar to Fig. 13 so the only iterated kernel smoothing result is shown in Fig. 14. Since this is a high SNR setting, all three methods are expected to perform well and similarly. In the heat kernel regression, 25% of mesh vertices show *F*-statistic value above 8.00 for the comparison of groups I and II (0.05 level). For the iterated kernel smoothing and diffusion smoothing, 24% and 24% of vertices are above 8.00. For the comparison of groups II and III, the numbers are 38%, 36% and 36% respectively. The differences are not significant.

5. Conclusions

This study presents a novel heat kernel regression framework, where functional measurements are expanded analytically using the weighted Laplace–Beltrami eigenfunctions. The weighted eigenfunction expansion is related to isotropic heat diffusion and the diffusion wavelet transform. The method was validated and compared against existing surface-based smoothing methods.

Although the proposed kernel regression and diffusion smoothing share identical FEM discretization, the kernel regression is a parametric model, whereas diffusion smoothing is not. The flexibility of the parametric model enabled us to establish the mathematical equivalence of kernel regression, diffusion smoothing and diffusion wavelets.

The method was subsequently applied to characterize mandible growth. Based on the significant directions of growth identified in Figs. 12 and 13, we quantified the regions, direction and extent of growth during the first two decades of life that contribute to the overall downward and forward growth of the mandible as described in the literature. To quantify mandibular growth using smaller age cohorts, we are currently securing additional CT images.

Acknowledgments

This work was supported by NIH Research Grant R01 DC6282 (MRI and CT Studies of the Developing Vocal Tract), from the

National Institute of Deafness and other Communicative Disorders (NIDCD); core Grant P-30 HD03352 to the Waisman Center from the National Institute of Child Health and Human Development (NICHD); Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS) UL1TR000427, and Vilas Associate Award from University of Wisconsin-Madison. We thank Lindell R. Gentry, Mike S. Schimek, Brian J. Whyms, Reid B. Durtschi and Dongjun Chung at the University of Wisconsin-Madison for assistance with image acquisition and image segmentation. We also thank the anonymous reviewers and Yuan Wang and Jacqueline Houtman for comments on earlier versions of this paper.

References

- Adler, R., 1981. *The Geometry of Random Fields*. John Wiley & Sons.
- Andrade, A., Kherif, F., Mangin, J., Worsley, K., Paradis, A., Simon, O., Dehaene, S., Le Bihan, D., Poline, J.B., 2001. Detection of fMRI activation using cortical surface mapping. *Hum. Brain Mapp.* 12, 79–93.
- Andresen, P.R., Bookstein, F.L., Conradsen, K., Ersbøll, B.K., Marsh, J.L., Kreiborg, S., 2000. Surface-bounded growth modeling applied to human mandibles. *IEEE Trans. Med. Imaging* 19, 1053–1063.
- Antoine, J.P., Roşca, D., Vanderheynt, P., 2010. Wavelet transform on manifolds: old and new approaches. *Appl. Comput. Harm. Anal.* 28, 189–202.
- Belkin, M., Niyogi, P., Sindhwani, V., 2006. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* 7, 2399–2434.
- Bernal-Rusiel, J., Atienza, M., Cantero, J., 2008. Detection of focal changes in human cortical thickness: spherical wavelets versus gaussian smoothing. *NeuroImage* 41, 1278–1292.
- Bronstein, M.M., Kokkinos, I., 2010. Scale-invariant heat kernel signatures for non-rigid shape recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1704–1711.
- Cachia, A., Mangin, J.F., Rivière, D., Kherif, F., Boddaert, N., Andrade, A., Papadopoulos-Orfanos, D., Poline, J.B., Bloch, I., Zilbovicius, M., Sonigo, P., Brunelle, F., Régis, J., 2003a. A primal sketch of the cortex mean curvature: a morphogenesis based approach to study the variability of the folding patterns. *IEEE Trans. Med. Imaging* 22, 754–765.
- Cachia, A., Mangin, J.F., Rivière, D., Papadopoulos-Orfanos, D., Kherif, F., Bloch, I., Régis, J., 2003b. A generic framework for parcellation of the cortical surface into gyri using geodesic Voronoi diagrams. *Image Anal.* 7, 403–416.
- Cao, J., Worsley, K., 2001. Applications of random fields in human brain mapping. *Spatial Stat.: Methodol. Aspects Appl.* 159, 170–182.
- Chung, M., 2001. *Statistical Morphometry in Neuroanatomy*. Ph.D. Thesis, McGill University. <<http://www.stat.wisc.edu/~mchung/papers/thesis.pdf>>.
- Chung, M., Taylor, J., 2004. Diffusion smoothing on brain surface via finite element method. In: *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 432–435.
- Chung, M., Worsley, K., Taylor, J., Ramsay, J., Robbins, S., Evans, A., 2001. Diffusion smoothing on the cortical surface. *NeuroImage* 13, S95.
- Chung, M., Robbins, S., Evans, A., 2005. Unified statistical approach to cortical thickness analysis. *Inform. Process. Med. Imaging (IPMI)*, Lect. Notes Comput. Sci. 3565, 627–638.
- Chung, M., Hartley, R., Dalton, K., Davidson, R., 2008. Encoding cortical surface by spherical harmonics. *Stat. Sinica* 18, 1269–1291.
- Coifman, R., Maggioni, M., 2006. Diffusion wavelets. *Appl. Comput. Harm. Anal.* 21, 53–94.
- de Goes, F., Goldenstein, S., Velho, L., 2008. A hierarchical segmentation of articulated bodies. *Comput. Graph. Forum* 27, 1349–1356.
- Dupuis, P., Grenander, U., Miller, M.L., 1998. Variational problems on flows of diffeomorphisms for image matching. *Quart. Appl. Math.* 56, 587–600.
- Enlow, D.H., Hans, M.G., 1996. *Essentials of Facial Growth*. W.B. Saunders Company, Philadelphia.
- Evans, L., 1998. *Partial Differential Equations*. American Mathematical Society.
- Gaser, C., Luders, E., Thompson, P., Lee, A., Dutton, R., Geaga, J., Hayashi, K., Bellugi, U., Galaburda, A., Korenberg, J., Mills, D., Toga, A., Reiss, A., 2006. Increased local gyrification mapped in Williams syndrome. *NeuroImage* 33, 46–54.
- Gottlieb, S., Shu, C.W., Tadmor, E., 2001. Strong stability-preserving high-order time discretization methods. *SIAM Rev.* 43, 89–112.
- Guskov, I., Wood, Z., 2001. Topological noise removal. In: *Graphics Interface*, pp. 19–26.
- Hagler Jr., D., Saygin, A., Sereno, M., 2006. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *NeuroImage* 33, 1093–1103.
- Hammond, D., Vanderheynt, P., Gribonval, R., 2011. Wavelets on graphs via spectral graph theory. *Appl. Comput. Harm. Anal.* 30, 129–150.
- Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., et al., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *NeuroImage* 32, 180–194.
- Han, J., Kim, J., Chung, C., Park, K., 2007. Evaluation of smoothing in an iterative lp-norm minimization algorithm for surface-based source localization of meg. *Phys. Med. Biol.* 52, 4791–4803.
- Hayasaka, S., Phan, K.L., Liberzon, I., Worsley, K.J., Nichols, T.E., 2004. Nonstationary cluster-size inference with random field and permutation methods. *NeuroImage* 22, 676–687.
- Hernandez, V., Roman, J.E., Tomas, A., Vidal, V., 2006. A Survey of Software for Sparse Eigenvalue Problems. Technical Report STR-6, Universidad Politécnica de Valencia. <<http://www.grycap.upv.es/slep/>>.
- Jo, H., Lee, J.M., Kim, J.H., Shin, Y.W., Kim, I.Y., Kwon, J., Kim, S., 2007. Spatial accuracy of fmri activation influenced by volume- and surface-based spatial smoothing techniques. *NeuroImage* 34, 550–564.
- Joshi, S.C., Davis, B., Jomier, M., Gerig, G., 2004. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* 23, 151–160.
- Joshi, A., Shattuck, D.W., Thompson, P.M., Leahy, R.M., 2009. A parameterization-based numerical method for isotropic and anisotropic diffusion smoothing on non-flat surfaces. *IEEE Trans. Image Process.* 18, 1358–1365.
- Kilner, J., Friston, K., 2010. Topological inference for eeg and meg. *Ann. Appl. Stat.* 4, 1272–1290.
- Kim, S.G., Chung, M., Seo, S., Schaefer, S., van Reekum, C., Davidson, R., 2011. Heat kernel smoothing via Laplace–Beltrami eigenfunctions and its application to subcortical structure modeling. In: *Pacific-Rim Symposium on Image and Video Technology (PSIVT)*. Lecture Notes in Computer Science (LNCS), pp. 36–47.
- Kim, W., Pachauri, D., Hatt, C., Chung, M., Johnson, S., Singh, V., 2012. Wavelet based multi-scale shape features on arbitrary surfaces for cortical thickness discrimination. In: *Advances in Neural Information Processing Systems*, pp. 1250–1258.
- Lai, Z., Hu, J., Liu, C., Taimouri, V., Pai, D., Zhu, J., Xu, J., Hua, J., 2010. Intra-patient supine-prone colon registration in CT colonography using shape spectrum. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, pp. 332–339.
- Lehoucq, R., Sorensen, D., Yang, C., 1998. *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. SIAM Publications, Philadelphia.
- Lerch, J.P., Evans, A., 2005. Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage* 24, 163–173.
- Lévy, B., 2006. Laplace–Beltrami eigenfunctions towards an algorithm that understands geometry. In: *IEEE International Conference on Shape Modeling and Applications*, p. 13.
- Lewis, A., Roche, A., Wagner, B., 1982. Growth of the mandible during pubescence. *Angle Orthod.* 52, 325–342.
- Li, C., Alexiades, V., 2010. Time stepping for the cable equation, Part 1: Serial performance. *Proc. Neural, Parallel Sci. Comput.* 4, 241–246.
- Loubele, M., Maes, F., Schutyser, F., Marchal, G., Jacobs, R., Suetens, P., 2006. Assessment of bone segmentation quality of cone-beam CT versus multislice spiral CT: a pilot study. *Oral Surgery, Oral Med., Oral Pathol., Oral Radiol., Endodontol.* 102, 225–234.
- Luders, E., Narr, K., Thompson, P., Rex, D., Woods, R., DeLuca, H., Jancke, L., Toga, A., 2006a. Gender effects on cortical thickness and the influence of scaling. *Hum. Brain Mapp.* 27, 314–324.
- Luders, E., Thompson, P., Narr, K., Toga, A., Jancke, L., Gaser, C., 2006b. A curvature-based approach to estimate local gyrification on the cortical surface. *NeuroImage* 29, 1224–1230.
- Mahadevan, S., Maggioni, M., 2006. Value function approximation with diffusion wavelets and Laplacian eigenfunctions. *Adv. Neural Inform. Process. Syst.* 18, 843.
- Malladi, R., Ravve, I., 2002. Fast difference schemes for edge enhancing Beltrami flow. In: *Proceedings of Computer Vision–ECCV*, Lecture Notes in Computer Science (LNCS), pp. 343–357.
- Miller, M., Qiu, A., 2009. The emerging discipline of computational functional anatomy. *NeuroImage* 45, S16–S39.
- Nain, D., Styner, M., Niethammer, M., Levitt, J., Shenton, M., Gerig, G., Bobick, A., Tannenbaum, A., 2007. Statistical shape analysis of brain structures using spherical wavelets. In: *IEEE Symposium on Biomedical Imaging ISBI*.
- Niethammer, M., Reuter, M., Wolter, F., Bouix, S., Peinecke, N., Koo, M., Shenton, M., 2007. Global medical shape analysis using the Laplace–Beltrami spectrum. *Lect. Notes Comput. Sci.* 4791, 850.
- Nilsson, J., Sha, F., Jordan, M., 2007. Regression on manifolds using kernel dimension reduction. In: *Proceedings of the 24th International Conference on Machine Learning*. ACM, pp. 697–704.
- Patané, G., 2015. Volumetric heat kernel: Padé–Chebyshev approximation, convergence, and computation. *Comput. Graph.* 46, 64–71.
- Perona, P., Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 629–639.
- Qiu, A., Miller, M., 2008. Multi-structure network shape analysis via normal surface momentum maps. *NeuroImage* 42, 1430–1438.
- Qiu, A., Bitouk, D., Miller, M., 2006. Smooth functional and structural maps on the neocortex via orthonormal bases of the Laplace–Beltrami operator. *IEEE Trans. Med. Imaging* 25, 1296–1396.
- Reuter, M., 2010. Hierarchical shape segmentation and registration via topological features of Laplace–Beltrami eigenfunctions. *Int. J. Comput. Vision* 89, 287–308.
- Reuter, M., Wolter, F.E., Peinecke, N., 2006. Laplace–Beltrami spectra as shape-dna of surfaces and solids. *Comput. Aided Des.* 38, 342–366.
- Reuter, M., Wolter, F.E., Shenton, M., Niethammer, M., 2009. Laplace–Beltrami eigenvalues and topological features of eigenfunctions for statistical shape analysis. *Comput. Aided Des.* 41, 739–755.
- Rosenberg, S., 1997. *The Laplacian on a Riemannian Manifold*. Cambridge University Press.

- Rustamov, R., 2007. Laplace–Beltrami eigenfunctions for deformation invariant shape representation. In: *Proceedings of the Fifth Eurographics Symposium on Geometry Processing*, p. 233.
- Schölkopf, B., Smola, A.J., 2002. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.
- Scott, J.H., 1976. *Dento-facial Development and Growth*. Pergamon Press, Oxford.
- Seo, S., Chung, M., 2011. Laplace–Beltrami eigenfunction expansion of cortical manifolds. In: *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, IEEE*, pp. 372–375.
- Seo, S., Chung, M., Vorperian, H., 2010. Heat kernel smoothing using Laplace–Beltrami eigenfunctions. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, pp. 505–512.
- Seo, S., Chung, M., Voperian, H., 2011. Mandible shape modeling using the second eigenfunction of the Laplace–Beltrami operator. In: *Proceedings of SPIE Medical Imaging*, p. 79620Z.
- Shawe-Taylor, J., Cristianini, N., 2004. *Kernel Methods for Pattern Analysis*. Cambridge University Press.
- Shen, L., Ford, J., Makedon, F., Saykin, A., 2004. Surface-based approach for classification of 3d neuroanatomical structures. *Intell. Data Anal.* 8, 519–542.
- Shen, L., Saykin, A., Chung, M., Huang, H., Ford, J., Makedon, F., McHugh, T., Rhodes, C., 2006. Morphometric analysis of genetic variation in hippocampal shape in mild cognitive impairment: role of an il-6 promoter polymorphism. In: *Life Science Society Computational Systems Bioinformatics Conference*.
- Shi, Y., Lai, R., Kern, K., Sicotte, N., Dinov, I., Toga, A.W., 2008. Harmonic surface mapping with Laplace–Beltrami eigenmaps. In: *11th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2008)*, pp. 147–154.
- Shi, Y., Dinov, I., Toga, A.W., 2009. Cortical shape analysis in the Laplace–Beltrami feature space. In: *12th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2009)*, pp. 208–215.
- Smartt Jr., J.M., Low, D.W., Bartlett, S.P., 2005. The pediatric mandible: I. A primer on growth and development. *Plast. Reconstr. Surg.* 116, 14e–23e.
- Sochen, N., Kimmel, R., Malladi, R., 1998. A general framework for low level vision. *IEEE Trans. Image Process.* 7, 310–318.
- Steinke, F., Hein, M., 2008. Non-parametric regression between manifolds. *Adv. Neural Inform. Process. Syst.*
- Stratemann, S.A., Huang, J.C., Maki, K., Hatcher, D.C., Miller, A.J., 2010. Evaluating the mandible with cone-beam computed tomography. *Am. J. Orthod. Dentofac. Orthop.* 137, S58–S70.
- Styner, M., Oguz, I., Xu, S., Brechbuhler, C., Pantazis, D., Levitt, J., Shenton, M., Gerig, G., 2006. Framework for the statistical shape analysis of brain structures using spharm-pdm. In: *Insight Journal, Special Edition on the Open Science Workshop at MICCAI*.
- Sun, J., Ovsjanikov, M., Guibas, L.J., 2009. A concise and provably informative multi-scale signature based on heat diffusion. *Comput. Graph. Forum* 28, 1383–1392.
- Tang, B., Sapiro, G., Caselles, V., 1999. Direction diffusion. In: *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp. 2:1245–1252.
- Tasdizen, T., Whitaker, R., Burchard, P., Osher, S., 2006. Geometric surface smoothing via anisotropic diffusion of normals. In: *Geometric Modeling and Processing*, pp. 687–693.
- Taubin, G., 2000. Geometric signal processing on polygonal meshes. In: *EUROGRAPHICS*.
- Taylor, J., Worsley, K., 2007. Detecting sparse signals in random fields, with an application to brain mapping. *J. Am. Stat. Assoc.* 102, 913–928.
- Vaillant, M., Qiu, A., Glaunès, J., Miller, M., 2007. Diffeomorphic metric surface mapping in subregion of the superior temporal gyrus. *NeuroImage* 34, 1149–1159.
- Vallet, B., Lévy, B., 2008. Spectral geometry processing with manifold harmonics. *Comput. Graph. Forum* 27, 251–260.
- Van Den Boomgaard, R., Van Balen, R., 1992. Methods for fast morphological image transforms using bitmapped binary images. *CVGIP: Graph. Models Image Process.* 54, 252–258.
- Walker, G.F., Kowalski, C.J., 1972. On the growth of the mandible. *Am. J. Phys. Anthropol.* 36, 111–118.
- Wang, F.Y., 1997. Sharp explicit lower bounds of heat kernels. *Ann. Probab.* 24, 1995–2006.
- Wang, Y., Zhang, J., Gutman, B., Chan, T., Becker, J., Aizenstein, H., Lopez, O., Tamburo, R., Toga, A., Thompson, P., 2010. Multivariate tensor-based morphometry on surfaces: application to mapping ventricular abnormalities in HIV/AIDS. *NeuroImage* 49, 2141–2157.
- Wardetzky, M., 2008. Convergence of the cotangent formula: an overview. In: *Bobenko, A., Sullivan, J., Schröder, P., Ziegler, G. (Eds.), Discrete Differential Geometry, Oberwolfach Seminars*, vol. 38. Birkhäuser Basel, pp. 275–286.
- Wood, Z., Hoppe, H., Desbrun, M., Schröder, P., 2004. Removing excess topology from isosurfaces. *ACM Trans. Graph. (TOG)* 23, 190–208.
- Worsley, K., Poline, J.B., Vandal, A., Friston, K., 1995. Test for distributed, non-focal brain activations. *NeuroImage* 2, 173–181.
- Worsley, K., 2003. Detecting activation in fMRI data. *Stat. Methods Med. Res.* 12, 401–418.
- Worsley, K., Cao, J., Paus, T., Petrides, M., Evans, A., 1998. Applications of random field theory to functional connectivity. *Hum. Brain Mapp.* 6, 364–367.
- Worsley, K., Andermann, M., Koulis, T., MacDonald, D., Evans, A., 1999. Detecting changes in nonisotropic images. *Hum. Brain Mapp.* 8, 98–101.
- Worsley, K., Taylor, J., Tomaiuolo, F., Lerch, J., 2004. Unified univariate and multivariate random field theory. *NeuroImage* 23, S189–S195.
- Yang, X., Goh, A., Qiu, A., 2011. Locally linear diffeomorphic metric embedding (LLDME) for surface-based anatomical shape modeling. *NeuroImage* 56, 149–161.
- Yger, F., Rakotomamonjy, A., 2011. Wavelet kernel learning. *Pattern Recogn.* 44, 2614–2629.
- Yotter, R.A., Dahnke, R., Gaser, C., 2009. Topological correction of brain surface meshes using spherical harmonics. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, pp. 125–132.
- Yushkevich, P., Zhang, H., Simon, T., Gee, J., 2008. Structure-specific statistical mapping of white matter tracts. *NeuroImage* 41, 448–461.
- Zhang, H., van Kaick, O., Dyer, R., 2007. Spectral methods for mesh processing and analysis. In: *EUROGRAPHICS*, pp. 1–22.
- Zhang, H., van Kaick, O., Dyer, R., 2010. Spectral mesh processing. *Comput. Graph. Forum* 29, 1865–1894.
- Zhong, J., Qiu, A., 2010. Multi-manifold diffeomorphic metric mapping for aligning cortical hemispheric surfaces. *Neuroimage* 49, 355–365.
- Zhu, H., Ibrahim, J., Tang, N., Rowe, D., Hao, X., Bansal, R., Peterson, B., 2007. A statistical analysis of brain morphology using wild bootstrapping. *IEEE Trans. Med. Imaging* 26, 954–966.