Statistics on diffeomorphisms via tangent space representations

Laurent Younes

Neuroimage

Related papers

Download a PDF Pack of the best related papers ⏷

Diffeomorphic metric surface mapping in subregion of the superior temporal gyrus
Joan Glaunes

Fast Template-based Shape Analysis using Diffeomorphic Iterative Centroid
Olivier Colliot

Morphometry of anatomical shape complexes with dense deformations and sparse parameters
Qiyong Chen

Statistics on diffeomorphisms via tangent space representations

M. Vaillant,[a,b,]* M.I. Miller,[a] L. Younes,[a,c] and A. Trouvé[d]

[a]*Center for Imaging Science, The Johns Hopkins University, Baltimore, MD 21218, USA*
[b]*Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA*
[c]*Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD 21218, USA*
[d]*CMLA (CNRS, UMR 8536), Ecole Normal Supérieure de Cachan, Cedex, France*

In this paper, we present a linear setting for statistical analysis of shape and an optimization approach based on a recent derivation of a conservation of momentum law for the geodesics of diffeomorphic flow. Once a template is fixed, the space of initial momentum becomes an appropriate space for studying shape via geodesic flow since the flow at any point along the geodesic is completely determined by the momentum at the origin through geodesic shooting equations. The space of initial momentum provides a linear representation of the nonlinear diffeomorphic shape space in which linear statistical analysis can be applied. Specializing to the landmark matching problem of Computational Anatomy, we derive an algorithm for solving the variational problem with respect to the initial momentum and demonstrate principal component analysis (PCA) in this setting with three-dimensional face and hippocampus databases.
© 2004 Elsevier Inc. All rights reserved.

Introduction

In Computational Anatomy, the framework pursued is the deformable template model pioneered by Grenander (1993). A deformable template $\mathcal{I}_\alpha$ corresponds to the orbit under a group of diffeomorphisms $\mathcal{G}$, of one selected and fixed object $I_\alpha \in \mathcal{I}$. The idea is to model comparison between elements in the orbit $\mathcal{I}_\alpha$ via the diffeomorphic transformations in $\mathcal{G}$. The optimal diffeomorphism that matches two arbitrary elements in the orbit is chosen from all curves ϕ_t, $t \in [0,1]$ in $\mathcal{G}$ connecting the two elements via the group action. It is chosen as the curve that minimizes an energy with respect to a measure of infinitesimal variations in $\mathcal{G}$. These energy minimizing paths (geodesics) induce a metric on $\mathcal{I}_\alpha$ (Miller et al., 2002) that provides a natural measure

for comparison of anatomical objects. In this paper, we present applications of a fundamental "conservation of momentum" property of these geodesics that has emerged recently (Miller et al., in press). This property extends the analogous conservation of momentum property in finite-dimensional mechanics and in the infinite-dimensional setting studied by Arnold (1989). We focus on the implications of this fundamental property in providing the powerful capability to represent the entire flow of a geodesic in the orbit by a template configuration and a momentum configuration at a single instant in time. The approach is founded in the Lie group point of view. The tangent space V at the identity id $\in \mathcal{G}$ is considered the "Lie algebra" of the group. The idea will be to use its dual space V^*, the space of momenta to model deformations, given that geodesics can be generated from elements of V or V^*. The power of the approach comes in the dimensionality reduction of geodesic flow to a single representative element, and in the fact that the representative space is linear. Motivated by these implications, we derive the variational problem of landmark matching with respect to initial momentum at time $t = 0$, and we setup the framework for applying linear statistical modeling in this space. The linear statistical setting provides a natural mechanism for coping with the nonlinear nature of the diffeomorphic shape space. Statistics on manifolds, in particular shape manifolds, has been studied in, for example, Bhattacharya and Patrangenaru (2002) and Le and Kume (2000). The representation of shape via their Lie algebra has been applied in the statistical learning setting in Gallivan et al. (2003) and Fletcher et al. (2003). However, it has not been studied in the diffeomorphic setting.

We will first provide background that introduces the diffeomorphic matching setting and the general conservation of momentum principle. We then specialize to the landmark matching problem and derive a new variational problem on the initial momentum as well as a numerical gradient algorithm. We then introduce the initial momentum as a natural setting for linear statistical analysis, in particular we detail the implementation of principal component analysis (PCA). Results of the optimization algorithm and a PCA analysis of three-dimensional face and hippocampus databases are presented in the final section.

---

* Corresponding author. Fax: +1 410 516 4594.
E-mail address: marc@jhu.edu (M. Vaillant).

Background

The group $\mathcal{G}$ is defined as follows. Let Ω be an open-bounded set in $\mathbb{R}^K$ that forms the background space. The fundamental object of construction is a Hilbert space V of vector fields on Ω that contain smooth vector fields with compact support in Ω. For all time-dependent families of elements of V, written $v_t \in V$ for $t \in [0,1]$, such that

$$\int_0^1 \| v_t \|_V dt < \infty,$$

the solution ϕ_t at time $t = 1$ of

$$\frac{\partial \varphi}{\partial t} = v_t \circ \varphi_t, \tag{1}$$

with $\phi_0(x) = x$ is a diffeomorphism (see Dupuis et al., 1998; Trouvé, 1995). The group $\mathcal{G}$ consists of all such solutions. The geodesics of $\mathcal{G}$ provide the transformations that match objects in the orbit and are characterized by extremals of the kinetic energy

$$\frac{1}{2} \int_0^1 \| v_t \|_V^2 dt.$$

We assume that there is an operator L defined on V such that its restriction to sufficiently smooth $v \in V$ gives $Lv \in L^2(\Omega)$ with

$$\langle v, w \rangle_V = \int_\Omega \langle Lv, w \rangle_{\mathbb{R}^K},$$

for all $w \in V$. Lv can be considered as a mapping from V to R through the identification $Lv(w) = \langle v, w \rangle_V$. In particular, for the energy at time t, we have $\|v_t\|_V^2 = Lv_t(v_t)$. The mapping $Lv_t : V \to \mathbb{R}$ is called the momentum of the system at time t. For sufficiently smooth Lv_t, the Euler equation that is satisfied by the extremal curves of the kinetic energy is given by

$$\frac{dLv}{dt} + \operatorname{div}(Lv \otimes v) + dv^* Lv = 0,$$

where $\operatorname{div}(u \otimes v) = duv + \operatorname{div}(v)u$. The Euler equation has been originally derived by Arnold (1966) for the case $L = id$ under the additional constraint $\operatorname{div}(v) = 0$, and by Miller et al. (in press). The Euler equation for more general momenta in the deformable template setting has also been derived in Miller et al. (in press). Also see Holm et al. (1998) and Holm et al. (2004).

In Miller et al. (in press), it is shown that, with respect to a change in variables, momentum is conserved along extremal curves of the kinetic energy. This implies

$$Lv_t(w) = Lv_0\left((d\varphi_t)^{-1} w \circ \varphi_t\right),$$

for all $w \in V$ so in fact the momentum Lv_t at time t is determined by the momentum at time $t = 0$. The important consequence is that the equations for geodesic evolution in the orbit $I_t = \phi_t \cdot I$ depend only on a fixed template and the momentum at time 0. This representation of I_t by Lv_0 offers considerable dimensionality reduction and suggests that the space of initial momenta is an appropriate setting for focusing our modeling effort; in particular for statistical analysis of shape, learning statistical models, and for new optimization procedures that can incorporate these prior models. In this paper, we investigate the diffeomorphic landmark-matching problem to derive a new numerical procedure based on the conservation of momentum property and to demonstrate statistical analysis in the space of initial momenta.

Methods

Landmark matching

The diffeomorphic landmark matching problem has been well formalized by Joshi and Miller (2000). We summarize briefly. $\mathcal{I}$ consists of all N-tuples $(\boldsymbol{q}^1,\ldots,\boldsymbol{q}^N)$ of landmark points in Ω. The orbit $\mathcal{I}_\alpha$ is generated by the action of $\mathcal{G}$ on a template landmark configuration I_α, where the group action is defined to be $\phi \cdot I = (\phi(\boldsymbol{q}^1),\ldots,\phi(\boldsymbol{q}^N))$.

Matching two elements in an orbit is accomplished by solving the following variational problem. Let $(\boldsymbol{x}^1,\ldots,\boldsymbol{x}^N)$ and $(\boldsymbol{y}^1,\ldots,\boldsymbol{y}^N)$ be template and target landmark configurations, respectively. We seek to find the time-varying velocity field v_t in V minimizing the following energy functional

$$\frac{1}{2} \int_0^1 \| v_t \|_V^2 dt + \frac{1}{2\sigma^2} \sum_{i=1}^N \| \varphi_1(\boldsymbol{x}^i) - \boldsymbol{y}^i \|_{\mathbb{R}^k}^2 \tag{2}$$

We assume that V is a reproducing kernel Hilbert space (RKHS), and therefore we define the symmetric matrix valued reproducing kernel K on $\Omega \times \Omega$ where for each $\boldsymbol{x} \in \Omega$, $K(\boldsymbol{x})$ is a map from R^K to V and

$$\langle K(\boldsymbol{x})\alpha, v \rangle_V = \langle v(\boldsymbol{x}), \alpha \rangle_{\mathbb{R}^K}, \tag{3}$$

for all $\alpha \in \mathbb{R}^K$ and $v \in V$. We make the following notational conventions. Bold notation will continue to be used to denote vectors in $\mathbb{R}^K$ so that $q \in \mathcal{I}$ can be written $(\boldsymbol{q}^1,\ldots,\boldsymbol{q}^N)$. We will also use an index notation, writing q^{ik} to denote $q \in \mathcal{I}$ explicitly, where k indexes the kth component of landmark i. When it is not needed to distinguish a landmark from its components, we will treat q as a vector in R^{NK} and abuse the notation by using only one index. That is, we write $q = (q^1,\ldots,q^{NK})$, where $q^1,\ldots,q^K$ denote the coordinates of the first landmark, $q^{K+1},\ldots,q^{2K}$ denote the coordinates of the second landmark, and so forth. We construct the $NK \times NK$ matrix-valued function S on $\mathcal{I}$, consisting of the $K \times K$ blocks $K(\boldsymbol{q}^i, \boldsymbol{q}^j)$ via $S(q) = (K(\boldsymbol{q}^i, \boldsymbol{q}^j), i, j = 1,\ldots,N)$.

Letting $\boldsymbol{q}^i(t) = \phi_t(\boldsymbol{x}^i)$ for $i = 1,\ldots,N$, an application of spline theory (see Joshi and Miller, 2000) from the RKHS viewpoint shows that the energy (Eq. (1)) is equivalent to

$$\frac{1}{2} \int_0^1 \dot{q}(t)^* S(q(t))^{-1} \dot{q}(t) dt + \frac{1}{2\sigma^2} \| q(1) - y \|_{\mathbb{R}^{NK}}^2, \tag{4}$$

where * denotes the matrix transpose. The velocity is interpolated over the entire domain Ω by

$$v_t(x) = \sum_{i=1}^N K(\boldsymbol{q}^i(t), x) \boldsymbol{p}_i(t), \tag{5}$$

where $p(t) = S(q(t))^{-1} \dot{q}(t)$. In the sequel, we will omit the time variable when it is clear from context. An important consequence of the system defined by Eqs. (4) and (5) is that the solution of our variation problem over the entire domain Ω depends only on the trajectories of the landmarks, thereby achieving substantial dimensionality reduction. We also remark that the integrand of the first term in Eq. (4),

$$L(q, \dot{q}, t) = \frac{1}{2} \dot{q}(t)^* S(q(t))^{-1} \dot{q}(t) \qquad (6)$$

is in fact a pure kinetic energy Lagrangian on the configuration space $\mathcal{I}$ (Marsden and Ratiu, 1999). Thus, after reduction to Eq. (4), we find that the original variational problem is connected to the variational principles of mechanics governing a system of NK particles with Lagrangian given in Eq. (6).

Geodesic evolution equations for landmarks

Given a template configuration $(x^1, \ldots, x^N)$ and the initial velocity v_0 of a geodesic in the orbit, as written in Eq. (5), we find

$$Lv_0(w) = \langle v_0, w \rangle_V = \langle \sum_{i=1}^{N} K(x^i) p_i, w \rangle_V$$

$$= \sum_{i=1}^{N} \langle p_i(0), w(x^i) \rangle_{\mathbb{R}^K}.$$

Thus, the elements $p_i(0)$ represent the initial momenta. The evolution equations describing the transport of the template along the geodesic are derived in Miller et al. (in press) and given by

$$\dot{q}^i(t) = \sum_{j=1}^{N} K(q^j(t), q^i(t)) p_j(t)$$

$$-\dot{p}_i(t) = \left(d_{q^i_{(t)}} v_t \right)^* p_i(t), \qquad (7)$$

noting that by Eq. (5), $d_{q^i(t)} v_t$ is a function only of the landmarks in the orbit. We remark that given the connection to mechanics via the Lagrangian (Eq. (6)), the specialization of the conservation of momentum to the landmark setting has produced evolution equations that are in fact recognized as Hamilton's equations of classical mechanics. They can be derived directly by applying the variational principle of Hamilton to Eq. (6) giving the Euler equations of motion and the equivalent Hamiltonian system (Eq. (7)) through a change of variables (see Marsden and Ratiu, 1999). We also recognize, stated above as our motivation for pursing this formulation, that Eq. (7) is an initial value ODE system. So, given initial values $(q^i(0), p_i(0))$, we can solve this system to give the unique solution $(q^i(t), p_i(t))$ for all $t \in [0,1]$ and hence $v_t(x)$ over all $x \in \Omega$ via Eq. (5). We proceed in the following section to further reduce the variation problem (Eq. (2)) with respect to the initial conditions $(q^i(0), p_i(0))$.

Variational problem on initial momentum

Equipped with Eq. (7) as the evolution equations, our strategy is to search for the initial conditions $(q(0), p(0))$ that give rise to the minimizing trajectories $(q(t), p(t))$. Substituting $\dot{q}(t) = S(q(t)) p(t)$, we write Eq. (4) as a function of $q(t)$ and $p(t)$ obtaining

$$\frac{1}{2} \int_0^1 p(t)^* S(q_t) p(t) dt + \frac{1}{2\sigma^2} \| q(1) - y \|_{\mathbb{R}^{NK}}^2.$$

In mechanics, $H(q(t), p(t)) = (1/2) p(t)^* S(qt) p(t)$ is known as the Hamiltonian. An important property along extremals of the energy is that this Hamiltonian energy is conserved. That is, $H(q(t), p(t))$ is a constant function of time. So, if $q(0) = x$ and $p(0) = v$, then $H(q(t), p(t)) = H(x, v)$ for all $t \in [0,1]$. We write $(q(x,v)(t), p(x,v)(t))$ for the mapping $(x,v) \mapsto (q(t), p(t))$ via Eq. (7). With the template x fixed, the energy becomes solely a function of v:

$$J(v) = \frac{1}{2} v^* S(x) v + \frac{1}{2\sigma^2} \| q(x,v)(1) - y \|_{\mathbb{R}^{NK}}^2.$$

In the sequel, q and p will be solutions of Eq. (7) unless otherwise noted, so we will often leave out the variables x and v (in addition to t) when writing $q(x, v)(t)$ and $p(x,v)(t)$.

Gradient

The energy has now been reduced to a functional on $\mathbb{R}^{NK}$, so it may be optimized using standard finite-dimensional nonlinear optimization techniques. We consider gradient-based approaches such as steepest descent and quasi-Newton, and now proceed by computing the gradient of J in $\mathbb{R}^{NK}$. We first write our equations in index notation that will allow us to produce an explicit formulation that can more directly be translated to computer code. Recall that q^{mn} denotes the nth component of the mth landmark. We also index the matrix $S(q)$ by $S^{klmn}(q) = K^{ln}(q^k, q^m)$ in which case the symmetry becomes $S^{klmn} = S^{mnkl}$. We write S_{klmn} for $S(q)^{-1}$. So, in terms of components, the energy becomes

$$J(v) = \frac{1}{2} \sum_{ikjl} S^{ikjl} v_{ik} v_{jl} + \frac{1}{2\sigma^2} \sum_{ik} \left(q^{ik}(x,v)(1) - y^{ik} \right)^2. \qquad (8)$$

A simple computation gives

$$\frac{\partial J}{\partial v_{rs}} = \dot{q}^{rs}(0) + \frac{1}{\sigma^2} \sum_{ik} \frac{\partial q^{ik}(1)}{\partial v_{rs}} \left(q^{ik}(1) - y^{ik} \right), \qquad (9)$$

where we have used the fact $\dot{q}^{rs} = \sum_{jl} S^{rsjl} v_{jl}$. The first term is immediately available and $q^{ik}(1)$ is given via the evolution equations. It is left to compute $(\partial q^{ik}(1)/\partial v_{rs})$, which we obtain by differentiating Eq. (7) to give a second set of differential equations that describe the evolution of $(\partial q^{ik}(t)/\partial v_{rs})$. Defining

$$\partial_{1k} K(x, y) = \frac{\partial K}{\partial x_k}(x, y)$$

$$S_k^{jlmo} = \partial_{ik} K^{lo}(q^j, q^m)$$

$$S_{kp}^{jlmo} = \partial_{1p} \partial_{ik} K^{lo}(q^j, q^m), \qquad (10)$$

we have

$$\dot{q}^{ik} = \sum_{jl} S^{ikjl} p_{jl}$$

$$-\dot{p}_{ik} = \sum_{lmo} S_k^{ilmo} p_{il} p_{mo} \qquad (11)$$

for the evolution equations.

$$\frac{\partial \dot{q}^{ik}}{\partial v_{rs}} = \sum_{jlp} \left(S_p^{ikjl} \frac{\partial q^{ip}}{\partial v_{rs}} p_{jl} + S_p^{jlik} \frac{\partial q^{ip}}{\partial v_{rs}} p_{jl} \right) + \sum_{jl} S^{ikjl} \frac{\partial p_{jl}}{\partial v_{rs}},$$

$$-\frac{\partial p^{ik}}{\partial v_{rs}} = \sum_{lmop} \left(S_{kp}^{ilmo} \frac{\partial q^{ip}}{\partial v_{rs}} p_{il} p_{mo} + S_{kp}^{moil} \frac{\partial q^{mp}}{\partial v_{rs}} p_{il} p_{mo} \right)$$
$$+ \sum_{lmo} \left(S_p^{ilmo} \frac{\partial p_{il}}{\partial v_{rs}} p_{mo} + S_p^{ilmo} \frac{\partial p_{mo}}{\partial v_{rs}} p_{il} \right), \qquad (12)$$

which describe the evolution of $(\partial q^k(t)/\partial v^i)$. In fact, these equations can be further simplified because it is natural to require invariance of $\langle \cdot, \cdot \rangle_V$ under rotations and translations, which results in a simplified form for K. Indeed, K is usually scalar, radial, and diagonal. That is, K is given by a function $g:[0, \infty) \rightarrow [0, \infty)$ that generates a positive definite kernel via

$$K^{lo}(\boldsymbol{x}, \boldsymbol{y}) = g(\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^K}) \delta_0^l.$$

We typically use the Gaussian kernel which corresponds to

$$g(t) = e^{-\frac{t}{2\sigma^2}}.$$

Optimization algorithm

Let x^{ik} and y^{ik} be template and target landmarks, respectively. The evolution Eqs. (11) and (12) are discretized in time and numerically integrated to compute the gradient (Eq. (9)). Letting T be the number of time steps, we have $\Delta t = 1/(T-1)$. A simple Euler integration procedure yields the following algorithm for computing the gradient.

(1) Initialize the initial velocity of the landmark trajectories to be the initial velocity of the straight line path.

$$\dot{q}^{ik}(0) = \Delta t \left(y^{ik} - x^{ik} \right).$$

Then, the remaining initial values become

$$p_{ik}(0) = S_{ikjl} \dot{q}^{jl}(0)$$

$$\frac{\partial p_{ik}(0)}{\partial v_{rs}} = \frac{\partial v_{ik}}{\partial v_{rs}} = \delta_r^i \delta_s^k$$

$$\frac{\partial q^{ik}(0)}{\partial v_{rs}} = 0.$$

(2) For $t = 0$ to $t = T - 2$,

(a) Compute $\dot{q}^{ik}(t)$, $\dot{p}_{ik}(t)$, $(\partial \dot{q}^{ik}(t)/\partial v_{rs})$, and $\partial \dot{p}_{ik}(t)/\partial v_{rs}$ via Eq. (11)

(b) Let

$$p_{ik}(t+1) = p_{ik}(t) + \Delta t \dot{p}_{ik}(t)$$

$$q^{ik}(t+1) = q^{ik}(t) + \Delta t \dot{q}^{ik}(t)$$

$$\frac{\partial p_{ik}(t+1)}{\partial v_{rs}} = \frac{\partial p_{ik}(t)}{\partial v_{rs}} + \Delta t \frac{\partial \dot{p}_{ik}}{\partial v_{rs}}$$

$$\frac{\partial q^{ik}(t+1)}{\partial v_{rs}} = \frac{\partial q^{ik}(t)}{\partial v_{rs}} + \Delta t \frac{\partial \dot{q}^{ik}}{\partial v_{rs}}$$

(3) Compute the gradient

$$\frac{\partial J}{\partial v_{rs}} = \dot{q}^{rs}(0) + \frac{1}{\sigma^2} \sum_{ik} \frac{\partial q^{ik}(T-1)}{\partial v_{rs}} \left(q^{ik}(T-1) - y^{ik} \right)$$

Equipped with this algorithm for computing the gradient, we use an iterative gradient descent scheme to optimize the energy. To obtain transformations that are invariant with respect to rotations R and translations b, we compute the closed form solution to

$$\operatorname*{Arg\,min}_{R,b} \sum_{i=1}^{N} \| \boldsymbol{q}^i(T-1) - R\boldsymbol{y}^i - b \|_{\mathbb{R}^K}^2$$

at each iteration of the descent procedure and replace $\boldsymbol{y}^i$ with $R\boldsymbol{y}^i + b$.

Linear statistics for shape space

Perhaps the most important application of the initial momentum point of view is for the effective linearization of the space of shapes. The shape space of landmark configurations that we are working with is in fact a nonlinear metric space. For example, there is no rigorous way to effectively add two landmark configurations and be guaranteed that the resulting configuration is a meaningful combination of the originals. However, it is desirable to have a good linear approximation to use linear statistical methods for studying shape variability. We have shown that a landmark configuration y^i can be represented by a template configuration x^i and the initial momenta v_i that minimize Eq. (8). The template remains fixed and the momenta belong to a linear space. Therefore, once a template is fixed, it is natural to use the representative space of initial momenta for linear analysis.

A common linear analysis technique for multivariate data, popularized in the stochastic deformable model literature by Cootes et al. (1995), is principal component analysis (PCA). PCA, which is also known as the Karhunen–Loeve decomposition, computes a variance-maximizing basis of the linear space. That is, basis vectors are oriented along directions of maximal variance and ordered such that the first vector spans the one-dimensional subspace of largest variation, the second vector captures the next largest variation, and so forth. It is therefore effective for analyzing variation in the data and for dimensionality reduction. For example, by projecting a data point onto a subspace spanned by the first D basis vectors, we obtain an optimal D dimensional representation of the data. If a
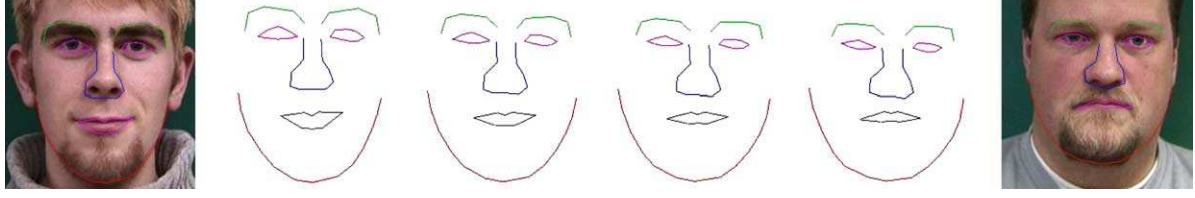
Fig. 1. 2D warping.

large percentage of the variation is captured in this D dimensional subspace, then the approximation can be expected to be quite good. If in addition, D is small, then the subspace may offer considerable dimensionality reduction of the sample space.

PCA is valid only for linear spaces, yet it is commonly used in the analysis of shape via landmark configurations. In particular, in active shape models (ASM) (Cootes et al., 1995), it is used to learn shape variability from a training set of shapes to reduce dimensionality of the shape space and to constrain a deformable shape model in its search for a known object in an image. It is typically implemented by giving the configuration space of landmarks $\mathcal{I}$, the vector space structure of Euclidean space R^{NK}. Given a set of M vectors $x_1^i, \ldots, x_M^i$ in R^{NK}, the PCA procedure is as follows:

(1) Estimate the mean and covariance:

$$\bar{x}^i = \frac{1}{M} \sum_{j=1}^{M} x_j^i$$

$$C^{ij} = \frac{1}{M} \sum_{k=1}^{M} \left(x_k^i - \bar{x}_k^i \right)\left(x_k^j - \bar{x}_k^j \right)$$

The mean is typically estimated in an iterative procedure that couples an alignment procedure called Procrustes alignment (Bookstein, 1993), with the computation above to give what is called the Procrustes mean.

(2) The desired basis is given by the eigenvectors $e_1^i, \ldots, e_{NK}^i$ of the covariance matrix C^{ij}, where the order is determined by the decreasing order of the corresponding eigenvalues, $\lambda_1 \geq \ldots \geq \lambda_{NK}$.

In fact, the eigenvalues estimate the variance along the axis of the corresponding eigenvector. So, if D is chosen such that

$$\sum_{j=1}^{D} \lambda_j \geq .95 \sum_{j=1}^{M} \lambda_j,$$

then the subspace spanned by $e_1^i, \ldots, e_D^i$ retains 95% of the variation in the training set, and we can hope to model the class of shapes described by the training set by D parameters $\lambda_1, \ldots, \lambda_D$ via

$$\bar{x}^i + \sum_{j=1}^{D} \gamma_j e_j^i. \tag{13}$$

For example, a shape y^i is approximated by its orthogonal projection onto the subspace, that is, we set

$$\gamma_j = \sum_{i=1}^{NK} \left(y^i - \bar{x}^i \right) e_j^i,$$

obtaining the approximation $\tilde{y}^i$ via Eq. (13). However, the affine model breaks down because the shape space is nonlinear. This is particularly evident for large coefficients γ_j. Consider three landmark configurations x_1^i, x_2^i, and x_3^i shown in Fig. 1 of two-dimensional face images taken from the AAM database (Stegmann, 2002). Suppose $\bar{x}^i = x_2^i$, and we take $e_1^i = x_1^i - \bar{x}^i$ and $e_3^i = x_3^i - \bar{x}^i$ as two basis elements. Sampling the affine subspace generated by e_1^i and e_3^i by generating random coefficients λ_1 and λ_3 and producing

$$\bar{x}^i + \gamma_1 e_1^i + \gamma_3 e_3^i,$$

we find elements of the subspace such as in Fig. 1d. Clearly, the resulting configuration can no longer be one that represents the "shape" of a face; the mouth appears above the nose and structures run into one another. However, the diffeomorphic landmark matching approach in this paper allows us to overcome the nonlinearity of the shape space in a fundamental way. Considering again the same three configurations in Fig. 1, we compute the initial momentum v_i^1 and v_i^3 that map $\bar{x}^i$ to x_1^i and $\bar{x}^i$ to x_3^i, respectively. Now, if we again generate random coefficients λ_1 and λ_3, set

$$v^i = \gamma_1 v_i^1 + \gamma_3 v_i^3,$$

and apply Eq. (11) to the initial conditions $(\bar{x}, v)$, we obtain a natural combination of the two configurations that are guaranteed to be a diffeomorphism of $\bar{x}^i$. In particular, taking the same coefficients λ_1^i and λ_3^i that produced the pathological

(a) $x^1 = \bar{x} + e_1$ (b) $x^2 = \bar{x}$ (c) $x^3 = \bar{x} + e_3$

(d) $\bar{x} + \boldsymbol{\gamma}_1 e_1 + \boldsymbol{\gamma}_3 e_3$ (e) $q(\bar{x}, \boldsymbol{\gamma}_1 \boldsymbol{u}^1 + \boldsymbol{\gamma}_2 \boldsymbol{u}^2)(1)$
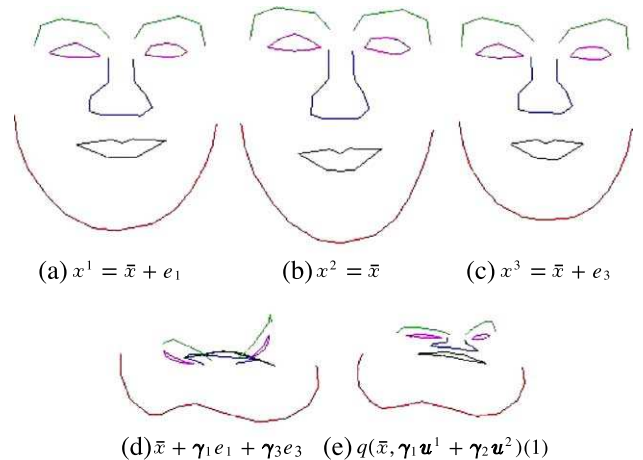
Fig. 2. Non-linearity artifact in (d) overcome by initial approach shown in (e).

model in Fig. 1d, we obtain Fig. 1e that clearly retains the structure of a face.

Mean shape and PCA on initial momentum

Until this point, our PCA has been computed with respect to the usual Euclidean norm on $\mathbb{R}^{NK}$. That is, we compute the basis that minimizes the residual error of the project approximation with respect to the Euclidean metric. In fact, we can do better by obtaining a decomposition with respect to the deformation energy. Recall that we measure the energy or "size" of a deformation by the first term in our energy $v*S(x)v$. Equipping $\mathbb{R}^{NK}$ with this norm $||v||_P^2 := v*S(x)v$ induces a natural metric on the initial momentum through the deformation energy. It is therefore more appropriate to compute PCA with respect to this norm, so that the projected approximation is optimal with respect to the deformation energy. A simple computation shows that the only change to the above procedure is to compute the eigen decomposition of $C_s = CS(x)$ instead of C, where $CS(x)$ is the matrix product of the original matrix C and the matrix $S(x)$.

Equipped with the evolution equations and an optimization procedure for computing diffeomorphic matching, it is straightforward to state PCA of shapes via initial momentum. Consider a set of M landmark shapes $x_1^i, \ldots, x_M^i$, we apply the following procedure to compute the mean shape:

(1) Set $\bar{x}^i$ to an arbitrarily chosen landmark configuration.
(2) Find the initial momenta $v_i^1, \ldots, v_i^{M-1}$ that represent the diffeomorphisms that map $\bar{x}^i$ to each of the $M-1$ remaining configurations.
(3) Compute the mean momentum

$$\bar{v}^i = \sum_{j=1}^{M-1} v_i^j$$

(4) Compute $q(\bar{x}, \bar{v})(t)$ and set $\bar{x} = q(\bar{x}, \bar{v})$ (Eq. (1)).
(5) Return to step 2 until $\|\bar{v}\|_{\mathbb{R}_{NK}}$ converges to 0, that is until the mean configuration does not change.

Because the space of initial momenta is linear, we simply apply PCA to the resulting $v_i^1, \ldots, v_i^M$ that have 0 mean, obtaining the decomposition $e_i^1, \ldots, e_i^{NK}$. Now, applying Eq. (7) with $\bar{x}^i$ and an element of span $\{e_i^1, \ldots, e_i^D\}$ as initial conditions will give a true sample from the space of shapes. That is, span $\{e_i^1, \ldots, e_i^D\}$ represents a true subspace of shapes diffeomorphic to $\bar{x}^i$.

Experimental results

Shown in Fig. 2 is a two-dimensional example in the plane where the landmarks are annotated features in photographs of the face (data are from the AAM database; Stegmann, 2002). The left panel shows the template landmark configuration (straight lines connecting landmarks) overlayed on the corresponding image of the face. The far right panel shows the overlay of the target landmark configuration on its corresponding image, and the center panels show snapshots from the time sequence of deformation of the template configuration. We see a smooth deformation as the template configuration moves close to the target. Shown in the top row of Fig. 3 is a three-dimensional example from the Morphable Faces database (Blanz and Vetter, 1999) of landmarks annotated on a two-dimensional face manifold. Again, the left panel shows the template configuration with the flow sequence in the center panels and the target in the right panels. The landmark deformation interpolates a smooth deformation of the surface, which matches the profile of the target nicely. In the bottom row of Fig. 3 is shown another three-dimensional landmark matching of two hippocampi. The hippocampi data are part of the morphology test bed for the Biomedical Informatics Research Network (BIRN, www.nbirn.net).

Shown in Fig. 4 is PCA via initial momentum of 100 annotated surfaces from the Morphable Faces database. The top row shows the mean configuration in frontal and profile view and the subsequent rows show deformation of the mean configuration along the first three principal directions, respectively. Presented in Fig. 5 is three-dimensional PCA applied to the left hippocampus of 19 normal subjects. The mean is shown in two views in the top row, and the deformation of the mean along the first three eigen modes is shown in the subsequent rows.

Discussion and conclusion

We have applied the conservation of momentum property to illustrate that an appropriate setting for statistical models of
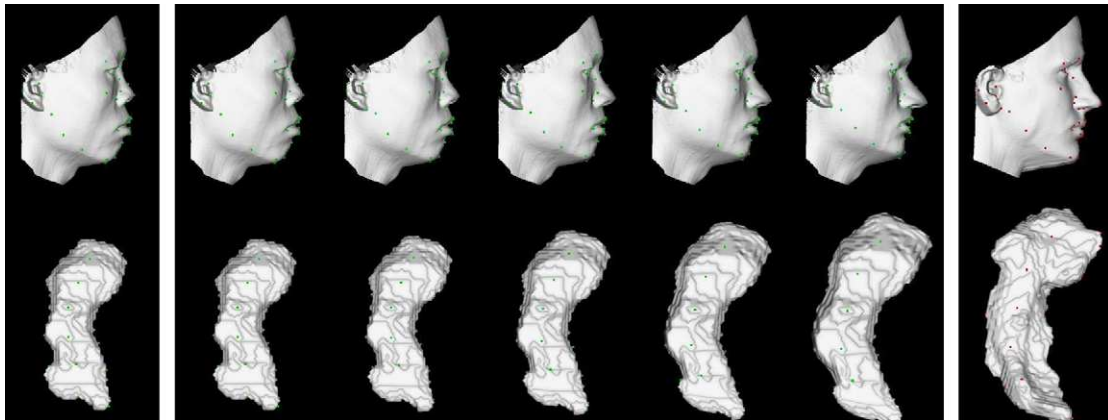

Fig. 3. Three-dimensional deformations.

object variations is the space of initial momentum and to design a new optimization approach for landmark matching. The motivation was to take advantage of the dimensionality reduction and linearity attained by initial momentum representations of geodesics in the orbit. We have emphasized the power of the space of initial momentum in providing a natural setting for shape analysis. Linear combinations of initial momenta always give rise to diffeomorphisms of the template

through the evolution equations, and we showed how this enables us to overcome limitations of other affine representations of shape. In particular, we demonstrated PCA on three-dimensional databases of face and hippocampus surfaces. Although the setting was restricted to landmark matching, the general conservation of momentum property applies to a broad variety of matching problems (Miller et al., in press). We remark that the approach to estimating the mean we have
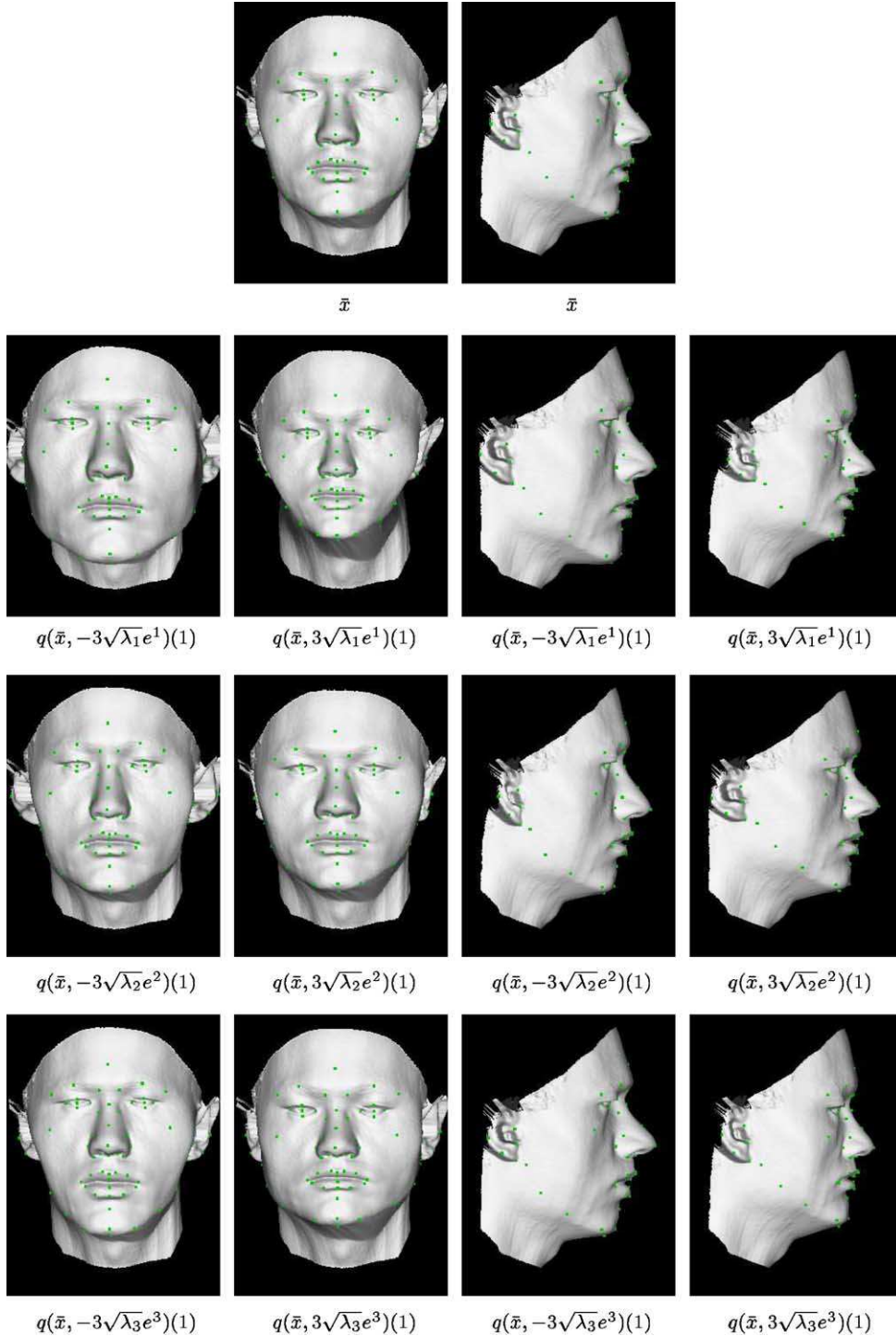

Fig. 4. First three eigen modes of three-dimensional PCA applied to the morphable faces database of 100 faces.

presented is analogous to the Lie group approaches of computing the so-called intrinsic mean of a probability measure on a Riemannian manifold (Miller et al., 2002) that has been applied to shape in, for example, Gallivan et al. (2003) and Fletcher et al. (2003). These other approaches differ in that their Lie algebra representations do not necessarily generate diffeomorphisms of the template. Although the gradient optimization scheme does not appear to achieve lower computational complexity over other optimization approaches (convergence properties have not yet been compared), it has
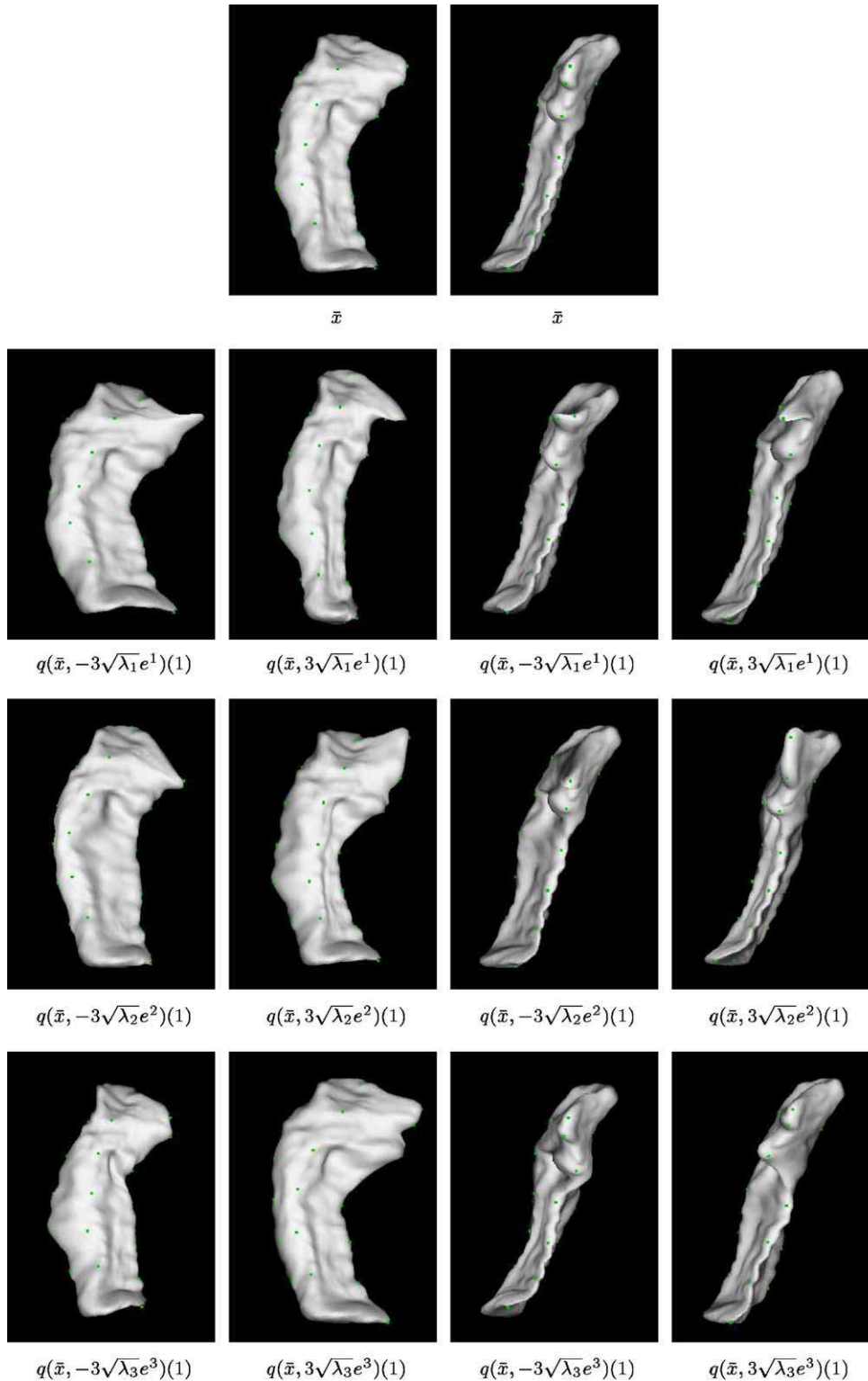

Fig. 5. First three eigen modes of three-dimensional PCA applied to 19 hippocampi from the BIRN project.

several advantages. First, it readily allows statistically learned information in the space of initial momentum to be incorporated in an active shape type approach for finding known objects in images. Second, it is guaranteed to generate a geodesic at every iteration, that is, $q(x,v)(t)$ is always a geodesic connecting x to $q(x,v)$ (1) by definition of the evolution equations. Lastly, the reduction to finite-dimensional nonlinear optimization opens the door to explore well-formalized numerical optimization procedures.

Acknowledgment

Appendix A

In this appendix, we derive the evolution equations for the gradient by differentiating the system (Eq. (11)) with respect to the initial momentum v_{rs}. Using the notational conventions introduced in Eq. (10) we find,

$$\frac{\partial \dot{q}^{ik}}{\partial v_{rs}} = \sum_{jlnp} \frac{\partial S^{ikjl}}{\partial q^{np}} \frac{\partial q^{np}}{\partial v_{rs}} p_{jl} + \sum_{jl} S^{ikjl} \frac{\partial p_{jl}}{\partial v_{rs}}$$

$$-\frac{\partial p^{ik}}{\partial v_{rs}} = \sum_{lmnop} \frac{\partial S_k^{ilmo}}{\partial q^{np}} \frac{\partial q^{np}}{\partial v_{rs}} p_{il} p_{mo}$$
$$+ \sum_{lmo} \left(S_k^{ilmo} \frac{\partial p_{il}}{\partial v_{rs}} p_{mo} + S_k^{ilmo} \frac{\partial p_{mo}}{\partial v_{rs}} p_{il} \right)$$

by the chain rule. We may reduce the computations by exploiting the fact that S is composed of symmetric kernels. We have

$$\frac{\partial S^{ikjl}}{\partial q^{np}} = \frac{\partial k^{kl}}{\partial q^{np}} \left(q^i, q^j \right)$$
$$= \partial_{1p} K^{kl} \left(q^i, q^j \right) \delta_n^i + \partial_{2p} K^{kl} \left(q^i, q^j \right) \delta_n^j$$
$$= \partial_{1p} K^{kl} \left(q^i, q^j \right) \delta_n^i + \partial_{1p} K^{lk} \left(q^j, q^i \right) \delta_n^j$$
$$= S_p^{ikjl} \delta_n^i + S_p^{jlik} \delta_n^j,$$

where the third line follows from the symmetry of K. Also, by equality of mixed partials and symmetry, again we find

$$\frac{\partial S_k^{ilmo}}{\partial q^{np}} = \partial_{1p} \partial_{1k} K^{lo} \left(q^i, q^m \right) \delta_n^i + \partial_{2p} \partial_{1k} K^{lo} \left(q^i, q^m \right) \delta_n^m$$
$$= \partial_{1p} \partial_{1k} K^{lo} \left(q^i, q^m \right) \delta_n^i + \partial_{1p} \partial_{1k} K^{ol} \left(q^m, q^i \right) \delta_n^m$$
$$= S_{kp}^{ilom} \delta_n^i + S_{kp}^{moil} \delta_n^m$$

So now

$$\frac{\partial \dot{q}^{ik}}{\partial v_{rs}} = \sum_{jlp} \left(S_p^{ikjl} \frac{\partial q^{ip}}{\partial v_{rs}} p_{jl} + S_p^{jlik} \frac{\partial q^{ip}}{\partial v_{rs}} p_{jl} \right) + \sum_{jl} S^{ikjl} \frac{\partial p_{jl}}{\partial v_{rs}},$$

and

$$-\frac{\partial p_{ik}}{\partial v_{rs}} = \sum_{lmop} \left(S_{kp}^{ilmo} \frac{\partial q^{ip}}{\partial v_{rs}} p_{il} p_{mo} + S_{kp}^{moil} \frac{\partial q^{mp}}{\partial v_{rs}} p_{il} p_{mo} \right)$$
$$+ \sum_{lmo} \left(S_k^{ilmo} \frac{\partial p_{il}}{\partial v_{rs}} p_{mo} + S_k^{ilmo} \frac{\partial p_{mo}}{\partial v_{rs}} p_{il} \right).$$

References

Arnold, V.I., 1966. Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications á l'hydrodynamique des fluides parfaits. Ann. Inst. Fourier (Grenoble) 16 (1), 319–361.

Arnold, V.I., 1989. Mathematical Methods of Classical Mechanics, 2nd ed. Springer, New York. 1978.

Bhattacharya, R., Patrangenaru, V., 2002. Nonparametric estimation of location and dispersion on Riemannian manifolds. J. Stat. Plan. Inference 108, 23–25.

Blanz, V., Vetter, T.A., 1999. Morphable model for the synthesis of 3D faces. Addison-Wesley: SIGGRAPH '99 Conference Proceedings, pp. 187–194.

Bookstein, F.L., 1993. Morphometric Tools for Landmark Data: Geometry and biology. Cambridge Univ. Press, Cambridge.

Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models: their training and application. Comput. Vis. Image Underst. 61 (1), 38–59.

Dupuis, P., Grenander, U., Miller, M.I., 1998. Variational problems on flows of diffeomorphisms for image matching. Q. Appl. Math. 56, 587–600.

Fletcher, P.T., Lu, C., Joshi, S., 2003. Statistics of shape via principal geodesic analysis on lie groups. Proceedings of CVPR. IEEE, pp. 95–101.

Gallivan, K.A., Srivastava, A., Xiuwen, L., 2003. Efficient algorithms for inferences on Grassmann manifolds. Proceedings of IEEE Conference on Statistical Signal Processing. IEEE, pp. 315–318.

Grenander, U., 1993. General Pattern Theory. Oxford Univ. Press.

Holm, D.D., Marsden, J.E., Ratiu, T.S., 1998. The Euler–Poincaré equations and semidirect products with applications to continuum theories. Adv. Math. 137, 1–81.

Holm, D.D., Ratnanather, J.T., Trouvé, A., Younes, L., 2004. Soliton dynamics in computational anatomy. NeuroImage 23 (Suppl. 1), S170–S178.

Joshi, S.C., Miller, M.I., 2000. Landmark matching via large deformation diffeomorphisms. IEEE Trans. Image Process. 9 (8), 1357–1370.

Le, H., Kume, A., 2000. The fréchet mean shape and the shape of means. Adv. Appl. Probab. (SGSA) 32, 101–113.

Marsden, J.E., Ratiu, T.S., 1999. Introduction to Mechanics and Symmetry, 2nd ed. Springer, New York.

Miller, M.I., Trouvé, A., Younes, L., 2002. On metrics and Euler–Lagrange equations of computational anatomy. Annu. Rev. Biomed. Eng. 4, 375–405.

Miller, M.I., Trouvé, A., Younes, L., 2003. Geodesic shooting for computational anatomy. J. Math. Imaging Vis. (in press).

Stegmann, M.B., 2002. Analysis and segmentation of face images using point annotations and linear subspace techniques. Technical report, Informatics and Mathematical Modelling. Technical University of Denmark, DTU. August.

Trouvé, A., 1995. An infinite dimensional group approach for physics based model. Technical report (electronically available at http://www.cis.jhu.edu).