

# Variational methods in shape analysis

Martin Rumpf<sup>1</sup> and Benedikt Wirth<sup>1</sup>

Bonn University, 53113 Bonn, Germany,  
{martin.rumpf,benedikt.wirth}@ins.uni-bonn.de,  
WWW home page: <http://www.ins.uni-bonn.de>

The analysis of shapes as elements in a frequently infinite-dimensional space of shapes has attracted increasing attention over the last decade. There are pioneering contributions in the theoretical foundation of shape space as a Riemannian manifold as well as path-breaking applications to quantitative shape comparison, shape recognition, and shape statistics. The aim of this chapter is to adopt a primarily physical perspective on the space of shapes and to relate this to the prevailing geometric perspective. Indeed, we here consider shapes given as boundary contours of volumetric objects, which consist either of a viscous fluid or an elastic solid.

In the first case, shapes are transformed into each other via viscous transport of fluid material, and the flow naturally generates a connecting *path* in the space of shapes. The viscous dissipation rate—the rate at which energy is converted into heat due to friction—can be defined as a metric on an associated Riemannian manifold. Hence, via the computation of shortest transport paths one defines a distance measure between shapes.

In the second case, shapes are transformed via elastic deformations, where the associated elastic energy only depends on the final *state* of the deformation and not on the path along which the deformation is generated. The minimal elastic energy required to deform an object into another one can be considered as a dissimilarity measure between the corresponding shapes.

In what follows we discuss and extensively compare the *path*-based and the *state*-based approach. As applications of the elastic shape model we consider shape averages and a principal component analysis of shapes. The viscous flow model is used to exemplarily cluster 2D and 3D shapes and to construct a flow type nonlinear interpolation scheme. Furthermore, we show how to approximate the viscous, path-based approach with a time-discrete sequence of state-based variational problems.

## 1 A review of different shape space concepts

The structure of shape spaces and statistical analyses of shapes have been examined in various settings, and applications range from the computation of priors for segmentation [1,2,3] and shape classification [4,5,6,7] to the construction of standardized anatomical atlases [8,9,10]. Among all existing approaches, a number of different concepts of a shape are employed, including landmark vectors [11,1], planar curves [12,13,14], surfaces in  $\mathbb{R}^3$  [4,15,16], boundary contours of

objects [7,17,18], multiphase objects [19] as well as the morphologies of images [20].

The analysis of a shape space is typically based on a notion of a distance or dissimilarity measure  $d(\cdot, \cdot)$  between shapes [21,5,18,17,22,23], whose definition frequently takes a variational form. This distance can be used to define an average [24,18] or a median [25,26]  $\mathcal{S}$  of given shapes  $\mathcal{S}_1, \dots, \mathcal{S}_n$  according to  $\mathcal{S} = \operatorname{argmin}_{\tilde{\mathcal{S}}} \sum_{i=1}^n d(\tilde{\mathcal{S}}, \mathcal{S}_i)^p$  for  $p = 1$  and  $p = 2$ , respectively (cf. Sec. 4.1). Likewise, shape variations can be obtained by a principal component analysis (PCA, cf. Sec. 4.2) or a more general covariance analysis in a way which is consistent with the dissimilarity measure between shapes [1,27,24,28]. From the conceptual point of view one can distinguish two types of these dissimilarity or distance measures which may be characterized as rather state-based or path-based, respectively. While the first approach is independent of the notion of paths of shapes, the latter distance definition requires the computation of an optimal, connecting path in shape space. In some cases, both concepts coincide: The Euclidean distance between two points, for example, can equivalently be interpreted in a state-based manner as the norm of the difference vector or as the length of the shortest connecting path (we shall provide a physical interpretation for each case in Sec. 2).

The notion of a shape space was already introduced by Kendall in 1984 [11], who considers shapes as  $k$ -tuples of points in  $\mathbb{R}^d$ , endowed with the quotient metric of  $\mathbb{R}^{kd}$  with respect to similarity transforms. Often, however, a shape space is just modelled as a linear vector space which is not invariant with respect to shift or rotation a priori. In the simplest case, such a shape space is made up of vectors of landmark positions, and distances between shapes can be evaluated in a state-based manner as the Euclidean norm of their difference. Chen and Parent [29] investigated averages of 2D contours already in 1989. Cootes et al. perform a PCA on training shapes with consistently placed landmarks to obtain priors for edge-based image segmentation [1]. Hafner et al. use a PCA of position vectors covering the proximal tibia to reconstruct the tibia surface just from six dominant modes [30]. Perperidis et al. automatically assign consistent landmarks to training shapes by a non-rigid registration as a preprocessing step for a PCA of the cardiac anatomy [31]. Söhn et al. compute dominant eigenmodes of landmark displacement on human organs, also using registration for preprocessing [32].

As an infinite-dimensional vector space, the Lebesgue-space  $L^2$  has served as shape space, where again shape alignment is a necessary preprocessing step. Leventon et al. identify shapes with their signed distance functions and impose the Hilbert space structure of  $L^2$  on them to compute an average and dominant modes of variation [2]. Tsai et al. apply the same technique to 3D prostate images [33]. Dambreville et al. also compute shape priors, but using characteristic instead of signed distance functions [34].

A more sophisticated state-based shape space is obtained by considering shapes as subsets of an ambient space with a metric  $d(\cdot, \cdot)$  and endowing them with the Hausdorff distance

$$d_H(\mathcal{S}_1, \mathcal{S}_2) = \max\left\{ \sup_{x \in \mathcal{S}_1} \inf_{y \in \mathcal{S}_2} d(x, y), \sup_{y \in \mathcal{S}_1} \inf_{x \in \mathcal{S}_2} d(x, y) \right\}$$

between any two shapes  $\mathcal{S}_1, \mathcal{S}_2$ . Charpiat et al. employ smooth approximations of the Hausdorff distance based on a comparison of the signed distance functions of shapes [21]. For a given set of shapes, the gradient of the shape distance functional at the average shape is regarded as shape variation of the average and used to analyse its dominant modes of variation [27]. Frame indifference is mimicked by an inner product that weights rotations, shifts, scalings, and the orthogonal complement to these transformations differently. Charpiat et al. also consider gradient flow morphing from one shape onto another one which can be regarded as a means to obtain meaningful paths even in shape spaces with state-based distance measures.

An isometrically invariant distance measure between shapes (or more general metric spaces) that is also not based on connecting paths is provided by the Gromov–Hausdorff distance, which can be defined variationally as

$$d_{\text{GH}}(\mathcal{S}_1, \mathcal{S}_2) = \frac{1}{2} \inf_{\substack{\phi: \mathcal{S}_1 \rightarrow \mathcal{S}_2 \\ \psi: \mathcal{S}_2 \rightarrow \mathcal{S}_1}} \sup_{\substack{y_i = \phi(x_i) \\ \psi(y_i) = x_i}} |d_{\mathcal{S}_1}(x_1, x_2) - d_{\mathcal{S}_2}(y_1, y_2)|,$$

where  $d_{\mathcal{S}_i}(\cdot, \cdot)$  is a distance measure between points in  $\mathcal{S}_i$ . The Gromov–Hausdorff distance represents a global, supremum-type measure of the lack of isometry between two shapes. Memoli and Sapiro use this distance for clustering shapes described by point clouds, and they discuss efficient numerical algorithms to compute Gromov–Hausdorff distances based on a robust notion of intrinsic distances  $d_{\mathcal{S}}(\cdot, \cdot)$  on the shapes [5]. Bronstein et al. incorporate the Gromov–Hausdorff distance concept in various classification and modelling approaches in geometry processing [35]. Memoli investigates the relation between the Gromov–Hausdorff distance and the Hausdorff distance under action of Euclidean isometries as well as  $L^p$ -type variants of the Gromov–Hausdorff distance [36].

In [37], Manay et al. define shape distances via integral invariants of shapes and demonstrate the robustness of this approach with respect to noise.

Another distance which also measures the lack of isometry between shapes can be obtained by interpreting shapes as boundaries of physical objects and measuring the (possibly nonlinear) deformation energy of an elastic matching deformation  $\phi$  between two objects [38,18]. Since, by the axiom of elasticity, this energy solely depends on the original and the final configuration of the deformed object but not on the deformation path, the elastic distance measure can clearly be classified as state-based (as will be detailed in Sec. 3.2). This physical approach comes along with a natural linearization of shapes via boundary stresses to perform a covariance analysis [28] and will be presented in Sec. 4. Pennec et al. define a nonlinear elastic energy as the integral over the ambient space of an energy density that depends on the logarithm of the Cauchy–Green strain tensor  $\mathcal{D}\phi^T \mathcal{D}\phi$  [39,40], which induces a symmetric state-based distance.

Typical path-based shape spaces have the structure of a Riemannian manifold. Here, the strength of a shape variation is measured by a Riemannian metric, and the square root of the Riemannian metric evaluated on the temporal shape variation is integrated along a path of shapes to yield the path length. The

length of the shortest path between two shapes represents their geodesic distance  $d(\cdot, \cdot)$ . Averages are obtained via the Fréchet mean [41], which was further analysed by Karcher [42]. There is also a natural linear representation of shapes in the tangent space at the Fréchet mean via the logarithmic map, which enables a PCA.

A Riemannian shape space which might still be regarded as rather state-than path-oriented is given by the space of polygonal medial axis representations, where each shape is described by a polygonal lattice and spheres around each vertex [43]: Here, the Lie group structure of the medial representation space can be exploited to approximate the Fréchet mean as exponential map of the average of the logarithmic maps of the input. Fletcher et al. perform a PCA on these log-maps to obtain the dominant geometric variations of kidney shapes [24] and brain ventricles [44]. Fuchs and Scherzer use the PCA on log-maps to obtain the covariance of medial representations, and they use a covariance-based Mahalanobis distance to impose a new metric on the shape manifold. This metric is employed to obtain priors for edge based image segmentation [45,46].

Kilian et al. compute and extrapolate geodesics between triangulated surfaces of fixed mesh topology, using isometry invariant Riemannian metrics that measure the local distortion of the grid [15]. Eckstein et al. employ different metrics in combination with a smooth approximation to the Hausdorff distance to perform gradient flows for shape matching [16]. Liu et al. use a discrete exterior calculus approach on simplicial complexes to compute geodesics and geodesic distances in the space of triangulated shapes, in particular taking care of higher genus surfaces [47].

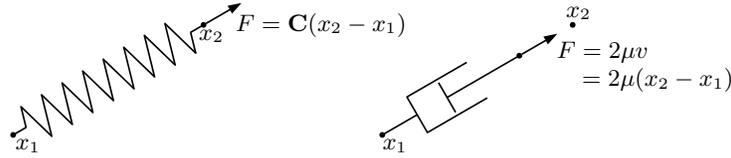
An infinite-dimensional Riemannian shape space has been developed for planar curves. Klassen et al. propose to use as a Riemannian metric the  $L^2$ -metric on variations of the direction or curvature functions of arclength parameterized curves. They implement a shooting method to find geodesics [12], while Schmidt and Cremers present an alternative variational approach [48]. Srivastava et al. assign different weights to the  $L^2$ -metric on stretching and on bending variations and obtain an elastic model of curves [49]. Michor and Mumford examine Riemannian metrics on the manifold of smooth regular curves [23]. They show the standard  $L^2$ -metric in tangent space leading to arbitrarily short geodesics and hence employ a curvature-weighted  $L^2$ -metric instead. Yezzi and Mennucci resolved the problem taking into account the conformal factor in the metric [13]. Sundaramoorthi et al. use Sobolev metrics in the tangent space of planar curves to perform gradient flows for image segmentation via active contours [50]. Michor et al. discuss a specific metric on planar curves, for which geodesics can be described explicitly [14]. In particular, they demonstrate that the sectional curvature on the underlying shape space is bounded from below by zero which points out a close relation to conjugate points in shape space and thus to only locally shortest geodesics. Finally, Younes considers a left-invariant Riemannian distance between planar curves by identifying shapes with elements of a Lie group acting on one reference shape [51].

When warping objects bounded by shapes in  $\mathbb{R}^d$ , a shape tube in  $\mathbb{R}^{d+1}$  is formed. Delfour and Zolésio [52] rigorously develop the notion of a Courant metric in this context. A further generalization to classes of non-smooth shapes and the derivation of the Euler–Lagrange equations for a geodesic in terms of a shortest shape tube is investigated by Zolésio in [53].

Dupuis et al. [54] and Miller et al. [22,55] define the distance between shapes based on a flow formulation in the embedding space. They exploit the fact that in case of sufficient Sobolev regularity for the motion field  $v$  on the whole surrounding domain  $\Omega$ , the induced flow consists of a family of diffeomorphisms. This regularity is ensured by a functional  $\int_0^1 \int_{\Omega} Lv \cdot v \, dx \, dt$ , where  $L$  is a higher order elliptic operator [50,51]. Geometrically,  $\int_{\Omega} Lv \cdot v \, dx$  is the underlying Riemannian metric, and we will discuss related, path-based concepts in Sec. 3.1. Under sufficient smoothness assumptions Beg et al. derive the Euler–Lagrange equations for the diffeomorphic flow field [56]. To compute geodesics between hypersurfaces in the flow of diffeomorphism framework, a penalty functional measures the distance between the transported initial shape and the given end shape. Vaillant and Glaunès [57] identify hypersurfaces with naturally associated two forms and used the Hilbert space structures on the space of these forms to define a mismatch functional. The case of planar curves is investigated under the same perspective by Glaunès et al. in [58]. To enable the statistical analysis of shape structures, parallel transport along geodesics is proposed by Younes et al. [59] as the suitable tool to transfer structural information from subject-dependent shape representations to a single template shape.

In most applications, shapes are boundary contours of physical objects. Fletcher and Whitaker adopt this viewpoint to develop a model for geodesics in shape space which avoids overfolding [60]. Fuchs et al. [17] propose a Riemannian metric on a space of shape contours, motivated by linearized elasticity. This metric can be interpreted as the rate of physical dissipation during the deformation of a viscous liquid object [61,19] and will be elaborated in Sec. 5.

Finally, a shape space is sometimes understood as a manifold, learnt from training shapes and embedded in a higher-dimensional (often linear) space. Many related approaches are based on kernel density estimation in feature space. Here, the manifold is described by a probability distribution in the embedding space, which is computed by mapping points of the embedding space into a higher-dimensional feature space and assuming a Gaussian distribution there. In general, points in feature space have no exact preimage in shape space so that approximate preimages have to be obtained via a variational formulation [62]. Cremers et al. use this technique to obtain 2D silhouettes of 3D objects as priors for image segmentation [3]. Rathi et al. provide a comparison between kernel PCA, local linear embedding (LLE), and kernel LLE (kernel PCA only on the nearest neighbours) [63]. Thorstensen et al. approximate the shape manifold using weighted Karcher means of nearest neighbour shapes obtained by diffusion maps [64].

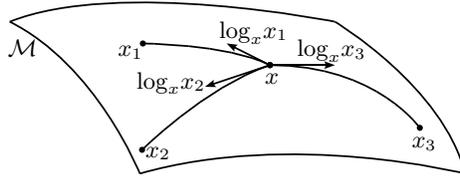


**Fig. 1.** The force  $F$  of an elastic spring between  $x_1$  and  $x_2$  is proportional to  $(x_2 - x_1)$ , as well as the force  $F$  of a dashpot which is extended from  $x_1$  to  $x_2$  within time 1 at constant velocity  $v$ . The spring energy reads  $\mathcal{W} = \int F dx = \frac{1}{2}\mathbf{C}\|x_2 - x_1\|_2^2$  and the dashpot dissipation  $\mathbf{Diss} = \int F \cdot v dt = 2\mu\|x_2 - x_1\|_2^2$ .

## 2 Recalling the finite-dimensional case

At first, let us investigate distances and their relation to concepts from physics in the simple case of Euclidian space. In Euclidean space, shortest paths are straight lines, and they are unique so that the distance computation involves only the states of the two end points: The geodesic distance between any two points  $x_1, x_2 \in \mathbb{R}^d$  is given by the norm of the difference,  $\|x_2 - x_1\|_2$ , which implies the equivalence of the state-based and the path-based perspective. A corresponding physical view might be the following. Considering that—by Hooke’s law—the stored elastic energy of an elastic spring extended from  $x_1$  to  $x_2$  is given by  $\mathcal{W} = \frac{1}{2}\mathbf{C}\|x_2 - x_1\|_2^2$  for the spring constant  $\mathbf{C}$ , the distance can be interpreted in a state-based manner as the square root of the elastic spring energy (Fig. 1). Likewise, from a path-based point of view, the minimum dissipated energy of a dashpot which is extended from  $x_1$  to  $x_2$  at constant speed within the fixed time interval  $[0, 1]$  reads  $\mathbf{Diss} = \int_0^1 2\mu\|v\|_2^2 dt = 2\mu\|x_2 - x_1\|_2^2$ , where  $2\mu$  is the dashpot parameter and the velocity is given by  $v = x_2 - x_1$ . Using this physical interpretation, we can express for instance the arithmetic mean  $x = \frac{1}{n} \sum_{i=1}^n x_i = \operatorname{argmin}_{\tilde{x}} \sum_{i=1}^n \|x_i - \tilde{x}\|_2^2$  of a given set of points  $x_1, \dots, x_n \in \mathbb{R}^d$  either as the minimizer of the total elastic deformation energy in a system where the average  $x$  is connected to each  $x_i$  by elastic springs or as the minimizer of the total viscous dissipation when extending dashpots from  $x_i$  to  $x$ .

Before we investigate the same concepts on more general Riemannian manifolds, let us briefly recall some basic notation. A Riemannian manifold is a set  $\mathcal{M}$  that is locally diffeomorphic to Euclidean space. Given a smooth path  $x(t) \in \mathcal{M}, t \in [0, 1]$ , we can define its derivative  $\dot{x}(t)$  at time  $t$  as a tangent vector to  $\mathcal{M}$  at  $x(t)$ . The vector space of all such tangent vectors makes up the tangent space  $T_{x(t)}\mathcal{M}$ , and it is equipped with the metric  $g_{x(t)}(\cdot, \cdot)$  as the inner product. The length of a path  $x(t) \in \mathcal{M}, t \in [0, 1]$ , is defined as  $\int_0^1 \sqrt{g_{x(t)}(\dot{x}(t), \dot{x}(t))} dt$ , and locally shortest paths are denoted geodesics. They can be shown to minimize  $\int_0^1 g_{x(t)}(\dot{x}(t), \dot{x}(t)) dt$  [65, Lemma 2.3]. Let us emphasize that a general geodesic is only locally the shortest curve. In particular there might be multiple geodesics of different length connecting the same end points. The geodesic distance between two points is the length of the shortest connecting path. Finally, for a given  $x \in \mathcal{M}$  there is a bijection  $\exp_x : T_x\mathcal{M} \rightarrow \mathcal{M}$  of a neighborhood of  $0 \in T_x\mathcal{M}$



**Fig. 2.** The logarithmic map assigns each point  $x_i$  on the manifold  $\mathcal{M}$  a vector in the tangent space  $T_x\mathcal{M}$ , which may be seen as a linear representative.

into a neighborhood of  $x \in \mathcal{M}$  that assigns to each tangent vector  $v \in T_x\mathcal{M}$  the end point of the geodesic emanating from  $x$  with initial velocity  $v$  and running over the time interval  $[0, 1]$  ([66, Thm. 1.6.12] or [67, Chp. 9, Thm. 14]).

We can now define the (possibly non-unique, cf. Sec. 6) mean  $x$  of a number of  $n$  points  $x_1, \dots, x_n \in \mathcal{M}$  in analogy to the Euclidian case as  $x = \operatorname{argmin}_{\tilde{x}} \sum_{i=1}^n d(x_i, \tilde{x})^2$ , where  $d(\cdot, \cdot)$  is the Riemannian distance on  $\mathcal{M}$ . This average is uniquely defined as long as the geodesics involved in the distance computation are unique, and it has been investigated in differential geometry by Karcher [42]. Furthermore, on a Riemannian manifold  $\mathcal{M}$ , the inverse exponential map  $\log_x = \exp_x^{-1}$  provides a method to obtain representatives  $\log_x(x_i) \in T_x\mathcal{M}$  of given input points  $x_i \in \mathcal{M}$  in the (linear) vector space  $T_x\mathcal{M}$  (Fig. 2). On these we can perform a PCA, which is by definition a linear statistical tool.

In a Riemannian space  $\mathcal{M}$ , the path-based approach can immediately be applied by exploiting the Riemannian structure, and  $\int_0^1 g_{x(t)}(\dot{x}(t), \dot{x}(t)) dt$  can be considered as the energy dissipation spent to move a point from  $x(0)$  to  $x(1)$  along a geodesic. The logarithms  $\log_x(x_i)$  in this model correspond to the initial velocities of the transport process leading from  $x$  to  $x_i$ . When applying the state-based elastic model in  $\mathcal{M}$ , however, there is no mechanically motivated notion of paths and thus also no logarithmic map. Only if we suppose that the Riemannian structure of the space  $\mathcal{M}$  is not induced by changes in the inner structure of our objects, the physical model based on elastic springs still coincides with the viscous model: We consider elastic springs stretched on the surface  $\mathcal{M}$  and connecting the points  $x$  and  $x_i$  with a stored energy  $\frac{1}{2}\mathbf{C}d(x, x_i)^2$ . Then, as before in the Euclidian case a state-based average  $x$  of input points  $x_1, \dots, x_n$  can be defined. Furthermore, interpreting spring forces acting on  $x$  and pointing towards  $x_i$  as linear representatives of the input points  $x_i$  one can run a PCA on these forces as well. However, for any reasonable (even finite-dimensional) model of shape space, objects are not rigid, and the inner relation between points as subunits (such as the vertex points of polygonal shapes) essentially defines the Riemannian (and thus the path-based) structure of the space  $\mathcal{M}$ : The rate of dissipation along a path in shape space depends on the interaction of object points. Physically, the corresponding point interaction energy is converted into thermal energy via friction. This dissipation depends significantly on the path in shape space traversed from one shape to the other. In contrast, when applying

the state-based approach to the same shape space, we directly compare the inner relations between the subunits, i. e. we have no history of these relations. This comparison can be quantified based on a stored (elastic) interaction energy which is then a quantitative measure of the dissimilarity of the two objects but in general no metric distance.

### 3 Path-based viscous dissipation versus state-based elastic deformation for non-rigid objects

In the following, we will especially consider two different physically motivated perspectives on a shape space of non-rigid volumetric objects in more detail. In the first case we will adopt a path-based view, motivated by the theory of viscous fluids, while the second, state-based approach will be motivated by elasticity.

We will regard shapes  $\mathcal{S}$  as boundaries  $\mathcal{S} = \partial\mathcal{O}$  of domains  $\mathcal{O} \subset \mathbb{R}^d$  which will be interpreted as physical objects. The resulting shape space structure depends on the particular type of physical objects  $\mathcal{O}$ : An interpretation of  $\mathcal{O}$  as a blob of a viscous fluid will yield an actually Riemannian, path-based shape space, while the interpretation as an elastic solid results in a state-based perspective, which will turn out to be non-Riemannian by construction.

#### 3.1 Path-based, viscous Riemannian setup

Shapes will be modeled as the boundary contour of a physical object that is made of a viscous fluid. The object might be surrounded by a different fluid (e. g. with much lower viscosity and compression modulus), nevertheless, without any restriction we will assume void outside the object in the derivation of our model. Here, *viscosity* describes the internal resistance in a fluid and is a macroscopic measure of the friction between fluid particles, e.g. the viscosity of honey is significantly larger than that of water. The friction is described in terms of the stress tensor  $\sigma = (\sigma_{ij})_{ij=1,\dots,d}$ , whose entries describe a force per area element. By definition,  $\sigma_{ij}$  is the force component along the  $i$ th coordinate direction acting on the area element with a normal pointing in the  $j$ th coordinate direction. Hence, the diagonal entries of the stress tensor  $\sigma$  refer to normal stresses, e. g. due to compression, and the off-diagonal entries represent tangential (shear) stresses. The Cauchy stress law states that due to the preservation of angular momentum the stress tensor  $\sigma$  is symmetric [68].

In a *Newtonian fluid* the stress tensor is assumed to depend linearly on the gradient  $\mathcal{D}v := (\frac{\partial v_i}{\partial x_j})_{ij=1,\dots,d}$  of the velocity  $v$ . In case of a rigid body motion the stress vanishes. A rotational component of the local motion is generated by the antisymmetric part  $\frac{1}{2}(\mathcal{D}v - (\mathcal{D}v)^T)$  of the velocity gradient, and it has the local rotation axis  $\nabla \times v$  and local angular velocity  $|\nabla \times v|$  [69]. Thus, as rotations are rigid body motions, the stress only depends on the symmetric part  $\epsilon[v] := \frac{1}{2}(\mathcal{D}v + (\mathcal{D}v)^T)$  of the velocity gradient. For an isotropic Newtonian fluid we get  $\sigma_{ij} = \lambda \delta_{ij} \sum_k (\epsilon[v])_{kk} + 2\mu (\epsilon[v])_{ij}$ , or in matrix notation  $\sigma = \lambda \text{tr}(\epsilon[v]) \mathbb{1} + 2\mu \epsilon[v]$ , where  $\mathbb{1}$  is the identity matrix. The parameter  $\lambda$  is denoted Lamé's first

coefficient. The local rate of viscous dissipation—the rate at which mechanical energy is locally converted into heat due to friction—can now be computed as

$$\mathbf{diss}[v] = \frac{\lambda}{2}(\mathrm{tr}\epsilon[v])^2 + \mu\mathrm{tr}(\epsilon[v]^2). \quad (1)$$

This is in direct correspondence to the mechanical definition of the stress tensor  $\sigma$  as the first variation of the local dissipation rate with respect to the velocity gradient, i. e.  $\sigma = \delta_{Dv}\mathbf{diss}$ . Indeed, by a straightforward computation we obtain  $\delta_{(Dv)_{ij}}\mathbf{diss} = \lambda\mathrm{tr}\epsilon[v]\delta_{ij} + 2\mu(\epsilon[v])_{ij} = \sigma_{ij}$ . Here  $\mathrm{tr}(\epsilon[v]^2)$  measures the averaged local change of length and  $(\mathrm{tr}\epsilon[v])^2$  the local change of volume induced by the transport. Obviously  $\mathrm{div}v = \mathrm{tr}(\epsilon[v]) = 0$  characterizes an incompressible fluid.

Now, let us consider a path  $(\mathcal{O}(t))_{t \in [0,1]}$  of objects connecting  $\mathcal{O}(0)$  with  $\mathcal{O}(1)$  and generated by a time-continuous deformation. If each point  $x \in \mathcal{O}(t)$  of the object  $\mathcal{O}(t)$  at time  $t \in [0, 1]$  moves in an Eulerian framework at the velocity  $v(t, x)$  ( $\dot{x} = v(t, x)$ ) so that the total deformation of  $\mathcal{O}(0)$  into  $\mathcal{O}(t)$  can be obtained by integrating the velocity field  $v$  in time, then the accumulated global dissipation of the motion field  $v$  in the time interval  $[0, 1]$  takes the form

$$\mathbf{Diss} \left[ (v(t), \mathcal{O}(t))_{t \in [0,1]} \right] = \int_0^1 \int_{\mathcal{O}(t)} \mathbf{diss}[v] dx dt. \quad (2)$$

This is the same concept as employed by Dupuis et al. [54] and Miller et al. [55] in their pioneering diffeomorphism approach. They minimize a dissipation functional under the simplifying assumption that the material behaves equally viscous inside and outside the object. Also,  $\mathbf{diss}[v] = \frac{\lambda}{2}(\mathrm{tr}\epsilon[v])^2 + \mu\mathrm{tr}(\epsilon[v]^2)$  is replaced by a higher order quadratic form  $Lv \cdot v$  which plays the role of the local rate of dissipation in a multipolar fluid model [70]. Multipolar fluids are characterized by the fact that the stresses depend on higher spatial derivatives of the velocity. If the quadratic form associated with  $L$  acts only on  $\epsilon[v]$  and is symmetric, then rigid body motion invariance is incorporated in the multipolar fluid model (cf. Section 5). In contrast to this approach we here measure the rate of dissipation differently inside and outside the object and rely on classical (monopolar) material laws from fluid mechanics.

On this physical background we will now derive a Riemannian structure on the space of shapes  $\mathcal{S}$  in an admissible class of shapes  $\mathbf{S}$ . The associated metric  $\mathcal{G}_{\mathcal{S}}$  on the (infinite-dimensional) manifold  $\mathbf{S}$  is in abstract terms a bilinear mapping that assigns each element  $\mathcal{S} \in \mathbf{S}$  an inner product on variations  $\delta\mathcal{S}$  of  $\mathcal{S}$  (cf. Sec. 2 above). The associated length of a tangent vector  $\delta\mathcal{S}$  is given by  $\|\delta\mathcal{S}\| = \sqrt{\mathcal{G}_{\mathcal{S}}(\delta\mathcal{S}, \delta\mathcal{S})}$ . Furthermore, as we have already seen above the length of a differentiable curve  $\mathcal{S} : [0, 1] \rightarrow \mathbf{S}$  is then defined by  $\mathbf{L}[\mathcal{S}] = \int_0^1 \|\dot{\mathcal{S}}(t)\| dt = \int_0^1 \sqrt{\mathcal{G}_{\mathcal{S}(t)}(\dot{\mathcal{S}}(t), \dot{\mathcal{S}}(t))} dt$ , where  $\dot{\mathcal{S}}(t)$  is the temporal variation of  $\mathcal{S}$  at time  $t$ . The Riemannian distance between two shapes  $\mathcal{S}_A$  and  $\mathcal{S}_B$  on  $\mathbf{S}$  is given as the minimal length taken over all curves with  $\mathcal{S}(0) = \mathcal{S}_A$  and  $\mathcal{S}(1) = \mathcal{S}_B$  or equivalently (cf. Sec. 2 above) as the length of a minimizer of the functional  $\int_0^1 \mathcal{G}_{\mathcal{S}(t)}(\dot{\mathcal{S}}(t), \dot{\mathcal{S}}(t)) dt$ . For shapes  $\mathcal{S} \in \mathbf{S}$  an infinitesimal variation  $\delta\mathcal{S}$  of a shape  $\mathcal{S} = \partial\mathcal{O}$  is associated

with a transport field  $v : \overline{\mathcal{O}} \rightarrow \mathbb{R}^d$ . This transport field is obviously not unique. Indeed, given any vector field  $w$  on  $\overline{\mathcal{O}}$  with  $w(x) \in T_x \mathcal{S}$  for all  $x \in \mathcal{S} = \partial \mathcal{O}$  (where  $T_x \mathcal{S}$  denotes the  $(d-1)$ -dimensional tangent space to  $\mathcal{S}$  at  $x$ ), the transport field  $v + w$  is another possible representation of the shape variation  $\delta \mathcal{S}$ . Let us denote by  $\mathcal{V}(\delta \mathcal{S})$  the affine space of all these representations. As a geometric condition for  $v \in \mathcal{V}(\delta \mathcal{S})$  we obtain  $v(x) \cdot n[\mathcal{S}](x) = \delta \mathcal{S}(x) \cdot n[\mathcal{S}](x)$  for all  $x \in \mathcal{S}$ , where  $n[\mathcal{S}](x) \in \mathbb{R}^d$  denotes the outer normal to  $\mathcal{S} \subset \mathbb{R}^d$  in  $x \in \mathcal{S}$ . Given all possible representations we are interested in the optimal transport, i. e. the transport leading to the least dissipation. Thus, using the definition (1) of the local dissipation rate we finally define the metric  $\mathcal{G}_{\mathcal{S}}(\delta \mathcal{S}, \delta \mathcal{S})$  as the minimal dissipation rate on motion fields  $v$  which are consistent with the variation of the shape  $\delta \mathcal{S}$ ,

$$\mathcal{G}_{\mathcal{S}}(\delta \mathcal{S}, \delta \mathcal{S}) := \min_{v \in \mathcal{V}(\delta \mathcal{S})} \int_{\mathcal{O}} \mathbf{diss}[v] \, dx = \min_{v \in \mathcal{V}(\delta \mathcal{S})} \int_{\mathcal{O}} \frac{\lambda}{2} (\text{tr} \epsilon[v])^2 + \mu \text{tr}(\epsilon[v]^2) \, dx. \quad (3)$$

Let us remark that we distinguish explicitly between the metric  $\mathbf{g}(v, v) := \int_{\mathcal{O}} \mathbf{diss}[v] \, dx$  on motion fields and the metric  $\mathcal{G}_{\mathcal{S}}(\delta \mathcal{S}, \delta \mathcal{S})$  on shape variations. Finally, integration in time leads to the total dissipation (2) to be invested in the transport along a path  $(\mathcal{S}(t))_{t \in [0,1]}$  in the shape space  $\mathbf{S}$ . This implies the following definition of a time-continuous geodesic path in shape space:

**Definition 1 (Geodesic path)** *Given two shapes  $\mathcal{S}_A$  and  $\mathcal{S}_B$  in a shape space  $\mathbf{S}$ , a geodesic path between  $\mathcal{S}_A$  and  $\mathcal{S}_B$  is a curve  $(\mathcal{S}(t))_{t \in [0,1]} \subset \mathbf{S}$  with  $\mathcal{S}(0) = \mathcal{S}_A$  and  $\mathcal{S}(1) = \mathcal{S}_B$  which is a local solution of*

$$\min_{v(t) \in \mathcal{V}(\dot{\mathcal{S}}(t))} \mathbf{Diss} \left[ (v(t), \mathcal{O}(t))_{t \in [0,1]} \right]$$

among all differentiable paths in  $\mathbf{S}$ .

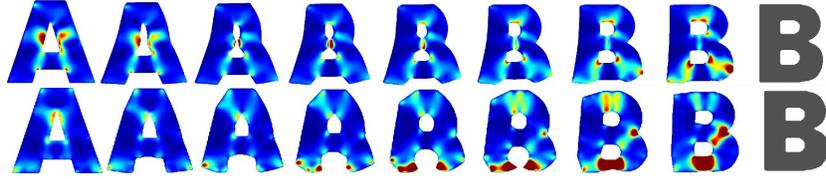
The Riemannian distance between two shapes  $\mathcal{S}_A$  and  $\mathcal{S}_B$  induced by this definition is given by the length of the shortest (geodesic) path  $\mathcal{S}(t)$  between the two shapes, i. e.

$$d_{\text{viscous}}(\mathcal{S}_A, \mathcal{S}_B) = \mathbf{L}[(\mathcal{S}(t))_{t \in [0,1]}].$$

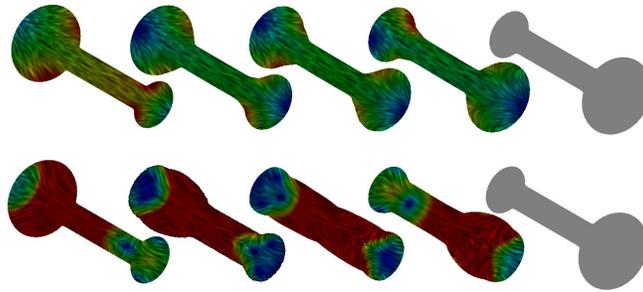
Fig. 3 shows two different paths between the same pair of shapes, one of them being a (numerically approximated) geodesic. Note that the chosen dissipation model combines the control of infinitesimal length changes via  $\text{tr}(\epsilon[v]^2)$ , and the control of compression via  $\text{tr}(\epsilon[v])^2$ . Fig. 4 evaluates the impact of these two terms on the shapes along a geodesic path.

### 3.2 State-based, path-independent elastic setup

Now, objects bounded by a shape contour  $\mathcal{S}$  are no longer composed of a viscous fluid but are considered to be elastic solids. To describe object deformations, we aim for an elastic energy which is not restricted to small displacements and which is consistent with first principles. Alongside the shape space modelling we will



**Fig. 3.** A geodesic (top, path length  $\mathbf{L} = 0.2225$  and total dissipation  $\mathbf{Diss} = 0.0497$ ) and a non-geodesic path (bottom,  $\mathbf{L} = 0.2886$ ,  $\mathbf{Diss} = 0.0880$ ) between an A and a B. The intermediate shapes of the bottom row are obtained via linear interpolation between the signed distance functions of the end shapes. The local dissipation rate is color-coded as .



**Fig. 4.** Two geodesic paths between dumbbell shapes varying in the size of the ends. In the top example the ratio  $\lambda/\mu$  between the dissipation parameters is 0.01 (leading to rather independent compression and expansion of the ends since the associated change of volume implies relatively low dissipation), and 100 in the bottom row (now mass is actually transported from one end to the other). The underlying texture on the objects is aligned to the transport direction, and the absolute value of the velocity  $v$  is color-coded as .

recall some background from elasticity. For details we refer to the comprehensive introductions in the books by Ciarlet [71] and Marsden and Hughes [72].

For two objects  $\mathcal{O}_A$  and  $\mathcal{O}_B$  with shapes  $\mathcal{S}_A = \partial\mathcal{O}_A$  and  $\mathcal{S}_B = \partial\mathcal{O}_B$  we assume a deformation  $\phi$  to be defined on  $\bar{\mathcal{O}}_A$  and constrained by the assumption  $\phi(\mathcal{S}_A) = \phi(\mathcal{S}_B)$ . For practical reasons one might consider  $\mathcal{O}_A$  to be embedded in a very soft elastic material occupying  $\Omega \setminus \mathcal{O}_A$  for some computational domain  $\Omega$ . There is an elastic energy  $\mathcal{W}_{\text{deform}}[\phi, \mathcal{O}_A]$  associated with the deformation  $\phi: \Omega \rightarrow \mathbb{R}^d$ . By definition, elastic means that this energy solely depends on the state and not on the path along which the deformation proceeds in time. More precisely, for so-called hyper-elastic materials,  $\mathcal{W}_{\text{deform}}[\phi, \mathcal{O}_A]$  is the integral of an energy density  $W$  depending solely on the Jacobian  $\mathcal{D}\phi$  of the deformation

$\phi$ , i. e.

$$\mathcal{W}_{\text{deform}}[\phi, \mathcal{O}_A] = \int_{\mathcal{O}_A} W(\mathcal{D}\phi) \, dx. \quad (4)$$

This elastic energy is considered as a dissimilarity measure between the shapes  $\mathcal{S}_A$  and  $\mathcal{S}_B$ . As a fundamental requirement one postulates the invariance of the deformation energy with respect to rigid body motions,  $\mathcal{W}_{\text{deform}}[Q \circ \phi + b, \mathcal{S}_A] = \mathcal{W}_{\text{deform}}[\phi, \mathcal{S}_A]$  for any orthogonal matrix  $Q \in SO(d)$  and translation vector  $b \in \mathbb{R}^d$  (the axiom of frame indifference in continuum mechanics). From this one deduces that the energy density only depends on the right Cauchy–Green deformation tensor  $\mathcal{D}\phi^T \mathcal{D}\phi$ . Hence, there is a function  $\tilde{W} : \mathbb{R}^{d,d} \rightarrow \mathbb{R}$  such that the energy density  $W$  satisfies  $W(F) = \tilde{W}(F^T F)$  for all  $F \in \mathbb{R}^{d,d}$ . The Cauchy–Green deformation tensor geometrically represents the metric measuring the deformed length in the undeformed reference configuration. For an isotropic material and for  $d = 3$  the energy density  $W$  can be further rewritten as a function  $\hat{W}(I_1, I_2, I_3)$  solely depending on the principal invariants of the Cauchy–Green tensor, namely  $I_1 = \text{tr}(\mathcal{D}\phi^T \mathcal{D}\phi)$ , controlling the local average change of length,  $I_2 = \text{tr}(\text{cof}(\mathcal{D}\phi^T \mathcal{D}\phi))$  ( $\text{cof} F := \det F F^{-T}$ ), reflecting the local average change of area, and  $I_3 = \det(\mathcal{D}\phi^T \mathcal{D}\phi)$ , which controls the local change of volume. For a detailed discussion we refer to [69,71]. We shall furthermore assume that the energy density is polyconvex [73], i. e. a convex function of  $\mathcal{D}\phi$ ,  $\text{cof} \mathcal{D}\phi$ , and  $\det \mathcal{D}\phi$ , and that isometries, i. e. deformations with  $\mathcal{D}\phi^T(x) \mathcal{D}\phi(x) = \mathbb{1}$ , are local minimizers with  $W(\mathcal{D}\phi) = \tilde{W}(\mathbb{1}) = 0$  [71]. Typical energy densities in this class are of the form

$$\hat{W}(I_1, I_2, I_3) = a_1 I_1^{\frac{p}{2}} + a_2 I_2^{\frac{q}{2}} + \Gamma(I_3) \quad (5)$$

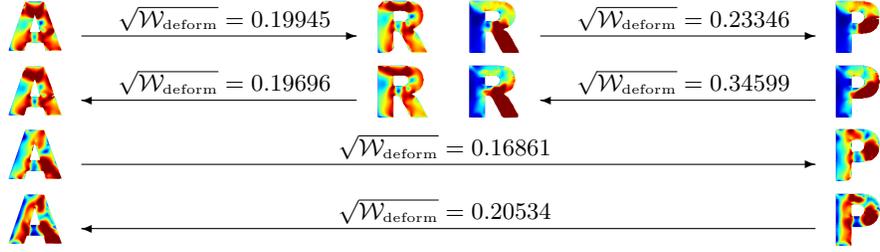
for  $a_1, a_2 > 0$  and a convex function  $\Gamma : [0, \infty) \rightarrow \mathbb{R}$  with  $\Gamma(I_3) \rightarrow \infty$  for  $I_3 \rightarrow 0$  and  $I_3 \rightarrow \infty$ . In nonlinear elasticity such material laws have been proposed by Ogden [74], and for  $p = q = 2$  (the case considered in our computations) we obtain the Mooney–Rivlin model [71]. The built-in penalization of volume shrinkage, i. e.  $\hat{W}(I_1, I_2, I_3) \xrightarrow{I_3 \rightarrow 0} \infty$ , enables us to control local injectivity (cf. [75]).

Incorporation of such a nonlinear elastic energy allows to describe large deformations with strong material and geometric nonlinearities, which cannot be treated by a linear elastic approach (cf. Hong et al. [38]). Furthermore, it balances in an intrinsic way expansion and collapse of the elastic objects and hence frees us from imposing artificial boundary conditions or constraints.

As in the previous section, the local force per area, induced by the deformation, is described at a point  $\phi(x) \in \phi(\mathcal{O})$  by the Cauchy stress tensor  $\sigma$ . It is related to the first Piola–Kirchhoff stress tensor  $\sigma^{\text{ref}} = W_{,F}(\mathcal{D}\phi) := \frac{\partial W(F)}{\partial F} \Big|_{F=\mathcal{D}\phi}$ , which measures the force density in the undeformed reference configuration, by  $\sigma^{\text{ref}} = \sigma \circ \phi \text{cof} \mathcal{D}\phi$ .

Based on these concepts from nonlinear elasticity we can now define a dissimilarity measure on shapes

$$d_{\text{elast}}(\mathcal{S}_A, \mathcal{S}_B) := \min_{\phi, \phi(\mathcal{S}_A) = \mathcal{S}_B} \sqrt{\mathcal{W}_{\text{deform}}[\phi, \mathcal{O}_A]}. \quad (6)$$



**Fig. 5.** Example of elastic distances between different shapes. The arrows indicate the direction of the deformation, the color-coding represents the local deformation energy density (in the reference as well as the deformed state).

Fig. 5 shows some applications of this measure. Obviously, the elastic energy is in general not symmetric so that  $d_{\text{elast}}(\mathcal{S}_A, \mathcal{S}_B) \neq d_{\text{elast}}(\mathcal{S}_B, \mathcal{S}_A)$ . Indeed, by construction  $d_{\text{elast}}(\cdot, \cdot)$  does not impose a metric structure on the space of shapes (we refer to Sec. 3.3 for a detailed discussion). Nevertheless, it can be applied to develop physically sound statistical tools for shapes such as shape averaging and a PCA on shapes, as outlined below in Sec. 4.

Let us make a brief remark on the mathematical relation between the two different concepts of elasticity and viscous fluids. If we assume the Hessian of the energy density  $W$  at the identity to be given by  $W_{,FF}(\mathbb{1})(G, G) = \lambda(\text{tr}G)^2 + \frac{\mu}{2}\text{tr}((G+G^T)^2)$  (which can be realized in (5) for a particular choice of  $a_1$ ,  $a_2$ , and  $\bar{\Gamma}$ , depending on the exponents  $p$  and  $q$ ), then by the ansatz  $\phi(x) = x + \tau v(x)$  and a second order Taylor expansion we obtain

$$\begin{aligned} W(\mathcal{D}\phi) &= W(\mathbb{1}) + \tau W_{,F}(\mathbb{1})(\mathcal{D}v) + \frac{\tau^2}{2} W_{,FF}(\mathbb{1})(\mathcal{D}v, \mathcal{D}v) + O(\tau^3) \\ &= 0 + 0 + \tau^2 \left( \frac{\lambda}{2} (\text{tr}\mathcal{D}v)^2 + \frac{\mu}{4} \text{tr} \left( (\mathcal{D}v + (\mathcal{D}v)^T)^2 \right) \right) + O(\tau^3). \end{aligned} \quad (7)$$

In effect, the Hessian of the nonlinear elastic energy leads to the energy density in linearized, isotropic elasticity

$$W^{\text{lin}}(\mathcal{D}u) = \frac{\lambda}{2} (\text{tr}\epsilon[u])^2 + \mu \text{tr}(\epsilon[u]^2) \quad (8)$$

for displacements  $u$  with  $\phi(x) = x + u(x)$ . This energy density, acting on displacements  $u$ , formally coincides with the local dissipation rate  $\mathbf{diss}[v]$ , acting on velocity fields  $v$ , in the viscous flow approach.

Finally, let us deal with the hard constraint  $\phi(\mathcal{S}_A) = \mathcal{S}_B$ , which is often inadequate in applications. Due to local shape fluctuations or noise in the shape acquisition, shape  $\mathcal{S}_A$  frequently contains details that are not present in  $\mathcal{S}_B$  and vice versa. These defects would imply high energies in a strict 1-1 matching approach. Hence, we have to relax the constraint and introduce some penalty

functional. Here, we either measure the symmetric difference of the input shapes  $\mathcal{S}_A$  and the pullback  $\phi^{-1}(\mathcal{S}_B)$  of the shape  $\mathcal{S}_B$  given by

$$\mathcal{F}[\mathcal{S}_A, \phi, \mathcal{S}_B] = \mathcal{H}^{d-1}(\mathcal{S}_A \Delta \phi^{-1}(\mathcal{S}_B)), \quad (9)$$

where  $A \Delta B = A \setminus B \cup B \setminus A$ , or alternatively the volume mismatch

$$\mathcal{F}[\mathcal{S}_A, \phi, \mathcal{S}_B] = \text{vol}(\mathcal{O}_A \Delta \phi^{-1}(\mathcal{O}_B)). \quad (10)$$

### 3.3 Conceptual differences between the path- and state-based dissimilarity measures

The state-based, elastic approach to dissimilarity measurement between shapes conceptually differs significantly from the path-based viscous flow approach. In the elastic setup, the axiom of elasticity implies that the energy at the deformed configuration  $\mathcal{S}_B = \phi(\mathcal{S}_A)$  is independent of the path from shape  $\mathcal{S}_A$  to shape  $\mathcal{S}_B$  along which the deformation is generated in time. Hence, there is no notion of shortest paths if we consider a purely elastic shape model, and different from a path-based approach there might not even exist an intermediate shape  $\mathcal{S}_C$  with  $d_{\text{elast}}(\mathcal{S}_A, \mathcal{S}_B) = d_{\text{elast}}(\mathcal{S}_A, \mathcal{S}_C) + d_{\text{elast}}(\mathcal{S}_C, \mathcal{S}_B)$ .

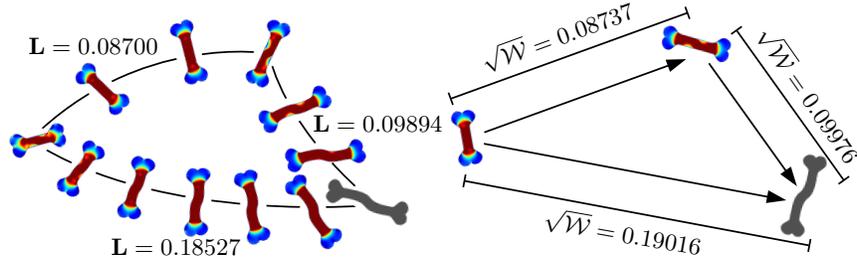
Unlike in the elasticity model, in the Newtonian model of viscous fluids the rate of dissipation and the induced stresses solely depend on the gradient of the motion field  $v$ . Even though the dissipation functional (2) looks like the deformation energy from linearized elasticity as outlined above, the underlying physics is only related in the sense that an infinitesimal displacement in the fluid leads to stresses caused by viscous friction, and these stresses are immediately absorbed via dissipation.

Surely, every (path-based) Riemannian space is metrizable (and in that sense state-based), and for many sufficiently regular (state-based) metric spaces we can devise a corresponding (path-based) Riemannian metric. However, from our mechanical perspective, the conceptual difference between the path-based, viscous and the state-based elastic approach is striking. In the *path-based* approach the structure of the space is too complicated for a closed formula of the geodesic distance so that the actual computation of a path is required. In the *state-based* approach there is no underlying path (i. e. no  $\mathcal{S}(t)_{t \in [0,1]}$  such that for any  $0 \leq t_1 \leq t_2 \leq t_3 \leq 1$  we have  $d(\mathcal{S}(t_1), \mathcal{S}(t_3)) = d(\mathcal{S}(t_1), \mathcal{S}(t_2)) + d(\mathcal{S}(t_2), \mathcal{S}(t_3))$ ), except the shape space structure is simple enough to allow for a closed formula of the geodesic distance as in Euclidean space.

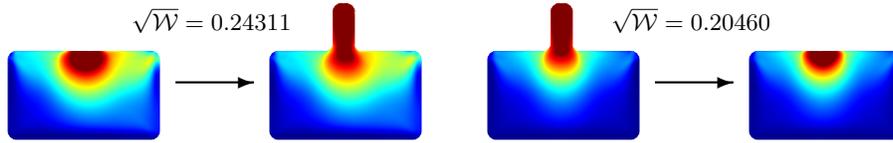
Mathematically, the path-based nature of the viscous flow approach and the fact that an inversion of the motion field  $v \rightarrow -v$  leads to a path from shape  $\mathcal{S}_B$  to  $\mathcal{S}_A$  in shape space with the same dissipation and length, i. e.

$$\mathbf{Diss} \left[ (v(t), \mathcal{O}(t))_{t \in [0,1]} \right] = \mathbf{Diss} \left[ (-v(1-t), \mathcal{O}(1-t))_{t \in [0,1]} \right],$$

ensures that the associated distance  $d_{\text{viscous}}$  is actually a metric. In particular, the symmetry condition  $d_{\text{viscous}}(\mathcal{S}_A, \mathcal{S}_B) = d_{\text{viscous}}(\mathcal{S}_B, \mathcal{S}_A)$  and the triangle inequality



**Fig. 6.** Left: Viscosity-based (time-discrete) geodesics between the shapes at the corners (the shapes are taken from [17]). The triangle inequality holds. Right: Elastic distances  $d_{\text{elast}}(\cdot, \cdot) = \sqrt{\mathcal{W}} \equiv \sqrt{\mathcal{W}_{\text{deform}}}$  between the same shapes, where the arrows point from the reference to the deformed configuration. The triangle inequality does not hold.



**Fig. 7.** The state-based elastic distance  $d_{\text{elast}}$  is not symmetric (as opposed to the path-based, viscous distance  $d_{\text{viscous}}$ ): In this example, it costs much more energy to drag out the protrusion than to push it in. The color-coding represents the local deformation energy density in the reference and the deformed configuration.

$d_{\text{viscous}}(\mathcal{S}_A, \mathcal{S}_C) \leq d_{\text{viscous}}(\mathcal{S}_A, \mathcal{S}_B) + d_{\text{viscous}}(\mathcal{S}_B, \mathcal{S}_C)$  hold. As we have already seen, the symmetry condition does not hold for the elastic dissimilarity measure. Also, the triangle inequality cannot be expected to hold. Indeed, if a deformation  $\phi_{A,B}$  maps  $\mathcal{O}_A$  onto  $\mathcal{O}_B$  and a deformation  $\phi_{B,C}$  maps  $\mathcal{O}_B$  onto  $\mathcal{O}_C$ , then  $\phi_{A,C} := \phi_{B,C} \circ \phi_{A,B}$  deforms  $\mathcal{O}_A$  onto  $\mathcal{O}_C$ . However, based on our elastic model,  $\mathcal{O}_B$  is considered to be stress-free when applying the deformation  $\phi_{B,C}$  (although it is actually obtained as the image of object  $\mathcal{O}_A$  under the deformation  $\phi_{A,B}$ ). Hence, the “history” of the deformation  $\phi_{A,B}$  is lost when measuring the energy of  $\phi_{B,C}$ . In addition, the energy density is highly nonlinear. As a consequence, in general we cannot expect  $d_{\text{elast}}(\mathcal{S}_A, \mathcal{S}_C) \leq d_{\text{elast}}(\mathcal{S}_A, \mathcal{S}_B) + d_{\text{elast}}(\mathcal{S}_B, \mathcal{S}_C)$ . Indeed, Fig. 6 gives an example where the triangle inequality holds in the viscous, path-based and fails in the elastic, state-based approach. Furthermore, Fig. 7 depicts another example for the lack of symmetry already apparent in Fig. 5 with a particularly pronounced mechanical difference of the two dissimilarity measures.

## 4 Elasticity-based shape space

In this section we will perform a statistical analysis on shapes up to the second moment, i. e. we will consider shape averaging and a principal component analysis on shapes as two exemplary applications of the state-based elastic shape space.

### 4.1 Elastic shape averaging

As usual, we consider objects  $\mathcal{O}$  as open sets in  $\mathbb{R}^d$  with the object shape given as  $\mathcal{S} := \partial\mathcal{O}$ . Given  $n$  sufficiently regular shapes  $\mathcal{S}_i = \partial\mathcal{O}_i$ ,  $i = 1, \dots, n$ , we are interested in an average shape which reflects the geometric characteristics of the input shapes in a physically intuitive manner. Suppose  $\mathcal{S} = \partial\mathcal{O} \subset \mathbb{R}^d$  denotes a candidate for this unknown shape. As it is characteristic for the elastic approach, the similarity of the input shapes  $\mathcal{S}_i$  to  $\mathcal{S}$  is measured by taking into account optimal elastic deformations  $\phi_i : \overline{\mathcal{O}_i} \rightarrow \mathbb{R}^d$  with  $\phi_i(\mathcal{S}_i) = \mathcal{S}$ . The elastic energy  $\mathcal{W}_{\text{deform}}[\phi_i, \mathcal{O}_i]$  of these deformations has the interpretation of a dissimilarity measure (cf. Sec. 3.2) so that we obtain a natural definition of an average shape as the minimizer of the sum of these terms (cf. Sec. 1).

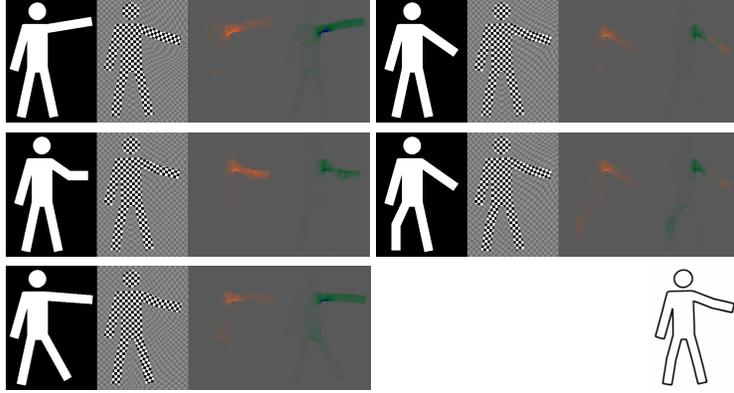
**Definition 2 (Elastic shape average)** *Given shapes  $\mathcal{S}_1, \dots, \mathcal{S}_n$  in some shape space  $\mathbf{S}$ , the elastic shape average  $\mathcal{S}$  is the minimizer of*

$$\sum_{i=1}^n d_{\text{elast}}(\mathcal{S}_i, \mathcal{S})^2 = \sum_{i=1}^n \inf_{\phi_i: \overline{\mathcal{O}_i} \rightarrow \mathbb{R}^d, \phi_i(\mathcal{S}_i) = \mathcal{S}} \mathcal{W}_{\text{deform}}[\phi_i, \mathcal{O}_i].$$

If the input objects  $\mathcal{O}_i$  have Lipschitz boundary and the integrand of the deformation energy  $\mathcal{W}_{\text{deform}}[\phi_i, \mathcal{O}_i] = \int_{\mathcal{O}_i} W(\mathcal{D}\phi_i) dx$  is polyconvex and bounded below by  $C_1 \|\mathcal{D}\phi_i\|^p - C_2$  for  $p > d$ ,  $C_1, C_2 > 0$ , the existence of a Hölder-continuous elastic shape average and deformations  $\phi_i \in W^{1,p}(\mathcal{O}_i)$  which realize the above infimum is guaranteed [76].

An example of a shape average is provided in Fig. 8. Obviously, the process of shape averaging is a constrained variational problem in which we simultaneously have to minimize over  $n$  deformations  $\phi_i$  and the unknown shape  $\mathcal{S}$  under the  $n$  constraints  $\phi_i(\mathcal{S}_i) = \mathcal{S}$ .

The necessary conditions for a set of minimizing deformations are the corresponding Euler–Lagrange equations. As usual, inner variations of one of the deformations lead to the classical system of PDEs  $\text{div } W_{,F}(\mathcal{D}\phi_i) = 0$  for every deformation  $\phi_i$  on  $\mathcal{O}_i \setminus \mathcal{S}_i$ , meaning a divergence-free, equilibrated stress field (cf. Sec. 3.2). Furthermore, the coupling between the deformations via the constraints  $(\phi_i(\mathcal{S}_i) = \mathcal{S})_{i=1, \dots, n}$  allows to derive a stress balance relation on  $\mathcal{S}$ : Consistent variation of all deformations  $\phi_i$  and the average  $\mathcal{S}$  by some displacement  $u : \overline{\mathcal{O}} \rightarrow \mathbb{R}^d$  via  $(\mathbb{1} + \delta u) \circ \phi_i$  and  $(\mathbb{1} + \delta u)(\mathcal{S})$  results in the optimality condition  $\frac{d}{d\delta} \sum_{i=1}^n \mathcal{W}_{\text{deform}}[(\mathbb{1} + \delta u) \circ \phi_i, \mathcal{O}_i] \Big|_{\delta=0} = 0$ , which after integration by parts leads to  $\sum_{i=1}^n \int_{\mathcal{S}_i} W_{,F}(\mathcal{D}\phi_i)(u \circ \phi_i) \cdot \nu[\mathcal{S}_i] da[\mathcal{S}_i] = 0$  for the outer normal



**Fig. 8.** Elastic shape average (bottom right) of five human silhouettes. For the computation, all shapes have actually been described as phase fields, and the elastic deformations are extended outside the input objects  $\mathcal{O}_i$  (cf. Sec. 7.2). The objects  $\mathcal{O}_i$  are depicted along with their deformations  $\phi_i$  (acting on a checkerboard) and the distribution of local length change  $\frac{1}{\sqrt{2}}\|\mathcal{D}\phi_i\|$  and volume change  $\det(\mathcal{D}\phi_i)$  (range  $[0.97, 1.03]$  color-coded as )

$\nu[\mathcal{S}_i]$  to  $\mathcal{S}_i$ . We have here exploited  $\operatorname{div}W_{,F}(\mathcal{D}\phi_i) = 0$  on  $\mathcal{O}_i \setminus \mathcal{S}_i$ . Now, we consider displacements  $u$  with local support and let this support collapse at some point  $x$  on  $\mathcal{S}$ . This yields the pointwise condition

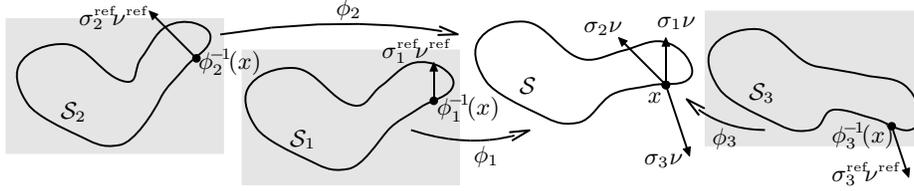
$$0 = \sum_{i=1}^n (\sigma_i^{\text{ref}} \nu[\mathcal{S}_i] da[\mathcal{S}_i])(\phi_i^{-1}(x)) \quad \text{and thus} \quad 0 = \sum_{i=1}^n (\sigma_i \nu[\mathcal{S}])(x) \quad (11)$$

for  $x \in \mathcal{S}$ , where we have used the relation

$$(\sigma_i^{\text{ref}} \nu[\mathcal{S}_i] da[\mathcal{S}_i])(\phi_i^{-1}(x)) = (\sigma_i \nu[\mathcal{S}] da[\mathcal{S}])(x)$$

between first Piola–Kirchhoff stress  $\sigma_i^{\text{ref}} = W_{,F}(\mathcal{D}\phi_i)$  and Cauchy stress  $\sigma_i = (\sigma_i^{\text{ref}}(\operatorname{cof}\mathcal{D}\phi_i)^{-1}) \circ \phi_i^{-1}$ . Hence, the shape average can be interpreted as that stable shape at which the boundary stresses of all deformed input shapes balance each other (Fig. 9). Obviously, there is a straightforward generalization involving jumps of normal stresses on interior interfaces in case of multi-component objects.

In order to ensure a certain regularity of the average shape  $\mathcal{S}$ , in addition to the sum of deformation energies in Def. 2 one can consider a further energy contribution which acts as a prior on  $\mathcal{S}$  in the variational approach. In the exemplary computations shown (Figs. 10 to 12), the  $(d-1)$ -dimensional Hausdorff measure  $\mathcal{L}[\mathcal{S}] = \mathcal{H}^{d-1}(\mathcal{S})$  has been employed as regularization.



**Fig. 9.** Sketch of the pointwise stress balance relation on the averaged shape.

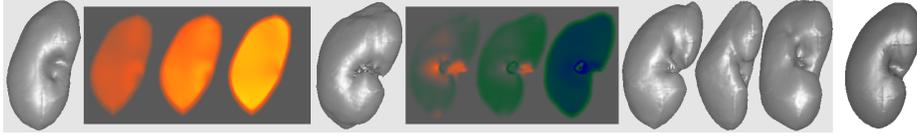


**Fig. 10.** Average of 18 hand silhouettes, taken from [1].

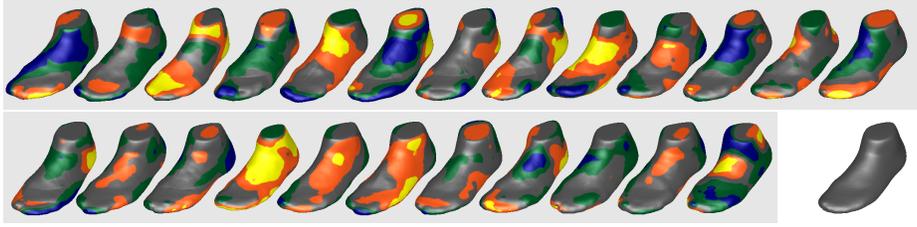
## 4.2 Elasticity-based PCA

As already explained in Sec. 2, a principal component analysis (PCA) is a linear statistical tool which decomposes a vector space into the direct sum of orthogonal subspaces. These subspaces are ordered according to the strength of variation which occurs along each subspace within a random set of sample vectors. We would like to interpret a given set of input shapes  $\mathcal{S}_1, \dots, \mathcal{S}_n$  as such a random sample and perform a corresponding PCA, however, due to the linearity of a PCA we first have to identify linear representatives for each shape on which a PCA can then be performed. For a Riemannian shape space, we have outlined in Sec. 2 that such linear representatives are given by the logarithmic map of the input shapes, but we have also learnt in Sec. 3.3 that a state-based elastic shape space is incompatible with a Riemannian structure.

To prepare the definition of appropriate linear representatives of shapes in an elastic shape space, let us briefly review the physical concept of boundary stresses. By the Cauchy stress principle, each deformation  $\phi_k : \mathcal{O}_k \rightarrow \mathcal{O}$  is characterized by pointwise boundary stresses on  $\mathcal{S} = \partial\mathcal{O}$  in the deformed configuration. The stress at some point  $x$  on  $\mathcal{S}$  is given by the application of the Cauchy stress tensor  $\sigma_k$  to the outer normal  $\nu$  on  $\mathcal{S}$ . The resulting stress  $\sigma_k \nu$  is a force density acting on a local surface element of  $\mathcal{S}$ . The shape  $\mathcal{S}$  is in an equilibrium configuration if the opposite force is applied as an external surface load (cf. Fig. 9). Otherwise, by the axiom of elasticity, releasing the object  $\mathcal{O}$ , the elastic body will snap back to the original reference configuration  $\mathcal{O}_k$ . Let us assume the relation between the energetically favorable deformation and its induced stresses to be one-to-one so that the average shape  $\mathcal{S}$  can be described in terms of the input shape  $\mathcal{S}_k$  and the boundary stress  $\sigma_k \nu$ , and we write  $\mathcal{S} = \mathcal{S}_k[\sigma_k \nu]$ . Upon scaling the stress with a weight  $t \in [0, 1]$  we obtain a one-parameter family of shapes  $\mathcal{S}(t) = \mathcal{S}_k[t\sigma_k \nu]$  connecting  $\mathcal{S}_k = \mathcal{S}(0)$  with  $\mathcal{S} = \mathcal{S}(1)$ . Thus, we can regard  $\sigma_k \nu$  as a representative of shape  $\mathcal{S}_k$  in the linear space of vector fields on  $\mathcal{S}$ .



**Fig. 11.** Five segmented kidneys and their average (right). For the first two input kidneys the distribution of  $\frac{1}{\sqrt{3}}\|\mathcal{D}\phi_i\|$ ,  $\frac{1}{\sqrt{3}}\|\text{cof}(\mathcal{D}\phi_i)\|$ , and  $\det(\mathcal{D}\phi_i)$  is shown on sagittal cross-sections (the range  $[0.85, 1.15]$  is color-coded as ). While the first kidney is dilated towards the average, the second is compressed.



**Fig. 12.** 24 given foot shapes (courtesy of adidas), textured with the distance to the surface of the average foot (bottom right). Values range from 6 mm inside the average foot to 6 mm outside, color-coded as .

Physically it is more intuitive to identify a displacement  $u_k$  instead of the normal stress  $\sigma_k\nu$  as the representative of an input shape  $\mathcal{S}_k$ . Hence, let us study how the average shape  $\mathcal{S}$  varies if we increase the impact of a particular input shape  $\mathcal{S}_k$  for some  $k \in \{1, \dots, n\}$ . For this purpose, we apply the Cauchy stress  $\sigma_k\nu$  to the average shape  $\mathcal{S}$ , scaled with a small constant  $\delta$ . This additional boundary stress  $\delta\sigma_k\nu$  may be seen as a first Piola–Kirchhoff stress acting on the (reference) configuration  $\mathcal{S}$ . The elastic response is given by a correspondingly scaled displacement  $u_k : \mathcal{O} \rightarrow \mathbb{R}^d$ . Here, to properly incorporate the nonlinear nature of the second moment analysis,  $\mathcal{O}$  should be interpreted as the compound object which is composed of all deformed and thus prestressed input objects  $\phi_i(\mathcal{O}_i)$ . This interpretation is reflected by the elastic material law employed to compute the displacements  $u_k$ . In detail,  $u_k$  is obtained as the minimizer of the free mechanical energy

$$\mathcal{E}_k[\delta, u] = \frac{1}{n} \sum_{i=1}^n \mathcal{W}_{\text{deform}}[(\mathbb{1} + \delta u) \circ \phi_i, \mathcal{O}_i] - \delta^2 \int_{\mathcal{S}} \sigma_k \nu \cdot u \, da \quad (12)$$

under the constraints  $\int_{\mathcal{O}} u_k \, dx = 0$  and  $\int_{\mathcal{O}} x \times u_k \, dx = 0$  of zero average translation and rotation. These displacements  $u_k$  are considered as representatives of the variation of the average shape  $\mathcal{S}$  with respect to the input shape  $\mathcal{S}_k$ , on which a PCA will be performed.

As long as  $F \mapsto W(F)$  is not quadratic in  $F$ ,  $u_k$  still solves a nonlinear elastic problem. The advantage of this nonlinear variational formulation is that

it is of the same type as the one for shape averaging, and it encodes in a natural way the compound elasticity configuration of the averaged shape domain  $\mathcal{O}$ . However, for the linearization of shape variations we are actually only interested in the displacements  $\delta u_k$  for small  $\delta$ . Therefore, we consider the limit of the Euler–Lagrange equations for  $\delta \rightarrow 0$  and after a little algebra obtain  $u_k$  as the solution of the linearized elasticity problem

$$\operatorname{div}(\mathbf{C}\epsilon[u]) = 0 \text{ in } \mathcal{O}, \quad \mathbf{C}\epsilon[u]\nu = \sigma_k\nu \text{ on } \mathcal{S} \quad (13)$$

for the symmetrized displacement gradient  $\epsilon[u] = \frac{1}{2}(\mathcal{D}u + \mathcal{D}u^T)$  under the constraints  $\int_{\mathcal{O}} u \, dx = 0$  and  $\int_{\mathcal{O}} x \times u \, dx = 0$ , where the in general inhomogeneous and anisotropic elasticity tensor  $\mathbf{C}$  reads

$$\mathbf{C} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{\det \mathcal{D}\phi_i} \mathcal{D}\phi_i W_{,FF}(\mathcal{D}\phi_i) \mathcal{D}\phi_i^T \right) \circ \phi_i^{-1}.$$

Next, for a PCA on the linearized shape variations  $u_k$  we select a suitable inner product (metric)  $g(u, \tilde{u})$  on displacements  $u, \tilde{u} : \mathcal{O} \rightarrow \mathbb{R}^d$ . Note that  $g$  induces a metric  $\tilde{g}(\sigma\nu, \tilde{\sigma}\nu) := g(u, \tilde{u})$  on the associated boundary stresses so that instead of analyzing the  $u_k$  the covariance analysis can equivalently be performed directly on the boundary stresses  $\sigma_1\nu, \dots, \sigma_n\nu$ , which we originally derived as linear shape representatives. Indeed, the solvability condition  $\int_{\mathcal{O}} \operatorname{div}(\mathbf{C}\nabla u) \, dx = \int_{\mathcal{S}} \mathbf{C}\nabla u\nu \, da[\mathcal{S}]$  is fulfilled, and thus the solution  $u_k$  for given boundary stress  $\sigma_k\nu = \mathbf{C}\nabla u\nu$  is uniquely determined up to a linearized rigid body motion (i. e. an affine displacement with skew-symmetric matrix representation), which is fixed by the conditions of zero mean displacement and angular momentum for  $u$ . Then, due to the linearity of the operator  $\sigma\nu \mapsto u$ , the metric  $\tilde{g}$  is bilinear and symmetric as well, and its positive definiteness follows from the positive definiteness of  $g$  and the injectivity of the map  $\sigma\nu \mapsto u$ .

We consider two different inner products on displacements  $u : \mathcal{O} \rightarrow \mathbb{R}^d$ :

- *The  $L^2$ -product.* Given two square integrable displacements  $u, \tilde{u}$  we define

$$g(u, \tilde{u}) := \int_{\mathcal{O}} u \cdot \tilde{u} \, dx$$

This product weights local displacements equally on the whole object  $\mathcal{O}$ .

- *The Hessian of the energy as inner product.* Different from the  $L^2$ -metric, we now measure displacement gradients in a non-homogeneous way. We define

$$g(u, \tilde{u}) := \int_{\mathcal{O}} \mathbf{C}\epsilon[u] : \epsilon[\tilde{u}] \, dx$$

for displacements  $u, \tilde{u}$  with square integrable gradients. Hence, the contribution to the inner product is larger in areas of the compound object which are in a significantly stressed configuration.

Given an inner product, we can define the covariance operator  $\mathbf{Cov}$  by

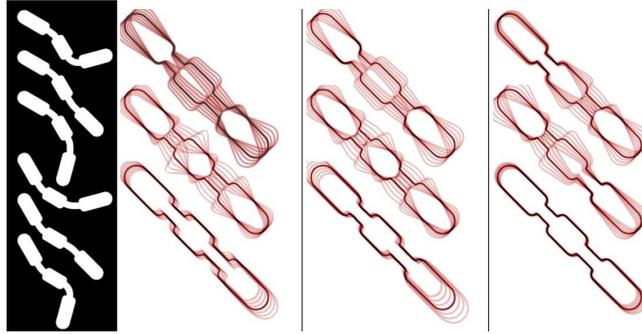
$$\mathbf{Cov} u := \frac{1}{n} \sum_{k=1}^n g(u, u_k) u_k$$

(note that the stresses  $\sigma_k \nu$  and thus also the displacements  $u_k$  have zero mean due to (11)). Obviously,  $\mathbf{Cov}$  is symmetric positive definite on  $\text{span}(u_1, \dots, u_n)$ . Hence, we can diagonalize  $\mathbf{Cov}$  on this finite-dimensional space and obtain a set of  $g$ -orthonormal eigenfunctions  $w_k : \mathcal{O} \rightarrow \mathbb{R}^d$  and eigenvalues  $\lambda_k > 0$  with  $\mathbf{Cov} w_k = \lambda_k w_k$ . These eigenfunctions can be considered as the principal modes of variation of the average object  $\mathcal{O}$  and hence of the average shape  $\mathcal{S}$ , given the  $n$  sample shapes  $\mathcal{S}_1, \dots, \mathcal{S}_n$ . Their eigenvalues encode the variation strength. The diagonalization of  $\mathbf{Cov}$  can be performed by diagonalizing the symmetric matrix  $\frac{1}{n} (g(u_i, u_j))_{ij} = O \Lambda O^T$ , where  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots)$  and  $O$  is orthogonal. The eigenfunctions are then obtained as  $w_k = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^n O_{jk} u_j$ .

Being displacements on  $\mathcal{O}$ , the modes of variation  $w_k$  can easily be visualized via a scalar modulation  $\delta w_k$  for varying  $\delta$  (cf. the visualization in Figs. 16 to 18 or the red lines in Figs. 13 and 15). If an amplified visualization of the modes is required, it is preferable to depict displacements  $w_\delta^k$  which are defined as minimizers of the nonlinear variational energy  $\frac{1}{n} \sum_{i=1}^n \mathcal{W}_{\text{deform}}[(\mathbb{1} + w) \circ \phi_i, \mathcal{O}_i] - \delta^2 \int_{\mathcal{S}} \mathbf{C} \nabla w_k \nu \cdot w \, da$  (cf. (12)).

Let us underline that this covariance analysis properly takes into account the usually strong geometric nonlinearity in shape analysis via the transfer of geometric shape variation to elastic stresses on the average shape, based on paradigms from nonlinear elasticity. Displacements or stresses are interpreted as the proper linearization of shapes. In abstract terms, either the space of displacements or stresses can be considered as the tangent space of shape space at the average shape, where the identification of displacements and stresses via (13) provides a suitable physical interpretation of stresses as shape variations.

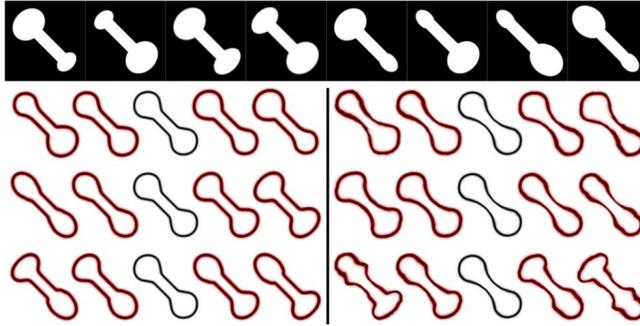
*The impact of the chosen metric.* Naturally, the modes of variation depend on the chosen inner product. We have already mentioned that in order to be physically meaningful, the inner product should act on displacements  $u_k$  of the compound object (which is composed of all deformed input shapes). If instead the  $u_k$  were obtained by applying the boundary stresses  $\sigma_k \nu$  to an object which just looks like the average shape but does not contain the information how strongly the input shapes had to be deformed to arrive at the average, we obtain a different result (Fig. 13, left): If the prestressed state of some object regions is neglected, it becomes easier to deform them which causes the prediction of stronger variations. Fig. 13 also hints at the differences between the employed metrics: The  $L^2$ -metric pronounces shape variations with large displacements even though they are energetically cheap (e.g. a rotation of some structure around a joint), while the Hessian of the elastic energy measures distances between displacements solely based on the associated change of elastic energy. Thus, displacements are weighted strongly in regions and directions which are significantly loaded.



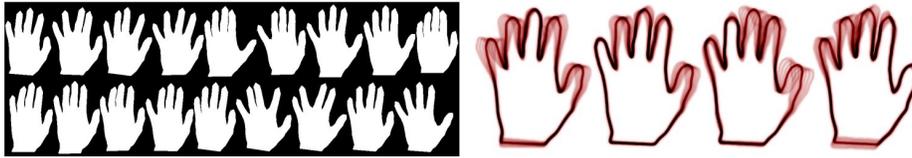
**Fig. 13.** First three dominant modes of variation for six input shapes (left), based on different metrics. Left:  $L^2$ -metric on displacements of a non-prestressed object (modes  $w_k$  with ratios  $\frac{\lambda_k}{\lambda_1}$  of 1, 0.23, 0.07). Middle:  $L^2$ -metric on displacements of the compound object ( $\frac{\lambda_k}{\lambda_1} = 1, 0.28, 0.03$ ). Right: Energy Hessian-based metric on displacements of the compound object ( $\frac{\lambda_k}{\lambda_1} = 1, 0.61, 0.24$ ).

*The impact of the nonlinear elasticity model.* Likewise, the particular choice of the nonlinear elastic energy density has a considerable effect on the average shape and its modes of variation. Fig. 14 has been obtained using  $W(\mathcal{D}\phi) = \frac{\mu}{2}\|\mathcal{D}\phi\|^2 + \frac{\lambda}{4}\det \mathcal{D}\phi^2 - (\mu + \frac{\lambda}{2})\log \det \mathcal{D}\phi - \mu - \frac{\lambda}{4}$ , where  $\mu$  and  $\lambda$  are the coefficients of length and volume change penalization, respectively. A low penalization of volume changes apparently leads to independent compression and inflation at the dumbbell ends (left), while for deformations with a strong volume change penalization (right), material is squeezed from one end to the other. Here, the underlying metric is the based on the Hessian of the energy.

Figs. 15 to 17 show the dominant modes of variation for the examples from the previous section. A statistical analysis of the hand shapes in Fig. 15 has also been performed in [1] and [25], where the shapes are represented as vectors of landmark positions. The average and the modes of variation are quite similar, representing different kinds of spreading the fingers. The dominant modes of variation for a set of 48 three-dimensional kidney shapes is depicted in Fig. 16, where for all modes  $w_k$  we show the average (middle) and its variation according to  $\delta w_k$  for varying  $\delta$ . Local structures seem to be quite well represented and preserved during the averaging process and the subsequent covariance analysis compared to e. g. the PCA on kidney shapes in [24] where a medial representation is used. The PCA of the 24 foot-shapes from Fig. 12 is shown in Fig. 17 and is much more intuitive than the color-coding in Fig. 12. The first mode apparently represents changing foot lengths, the second and third mode belong to different variants of combined width and length variation, and the fourth to sixth mode correspond to variations in relative heel position, ankle thickness, and instep height. Finally, Fig. 18 shows that the approach also works for image morphologies instead of shapes, using thorax CT scans as input. Here, the image edge set is considered as the corresponding shape, which is typically quite complex and characterized



**Fig. 14.** First three modes of variation for eight dumbbell shapes, left for a 100 times stronger penalization of length than of volume changes (with ratios  $\frac{\lambda_i}{\lambda_1}$  of 1, 0.22, 0.05), right for the reverse ( $\frac{\lambda_i}{\lambda_1} = 1, 0.41, 0.07$ ). Each row represents the variation of the average (middle shape) by  $\delta w_k$  for the mode  $w_k$  and varying  $\delta$ .



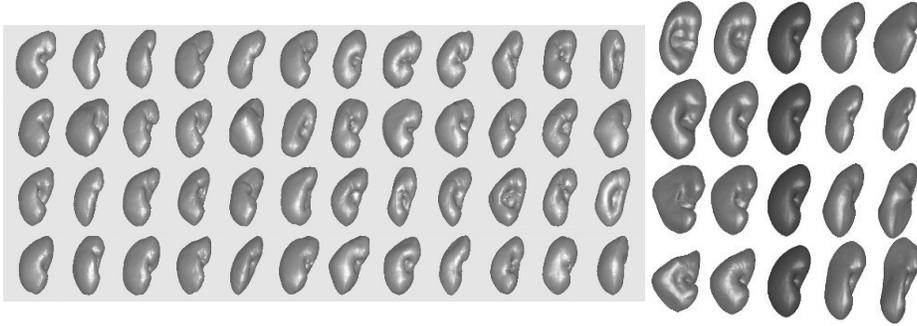
**Fig. 15.** First four modes of variation with ratios  $\frac{\lambda_i}{\lambda_1}$  of 1, 0.88, 0.42 and 0.25 for the 18 hand silhouettes from Fig. 10.

by nested contours. The first mode of variation represents a variation in chest size, the next mode corresponds to a change of heart and scapula shape, while the third mode mostly concerns the rib position.

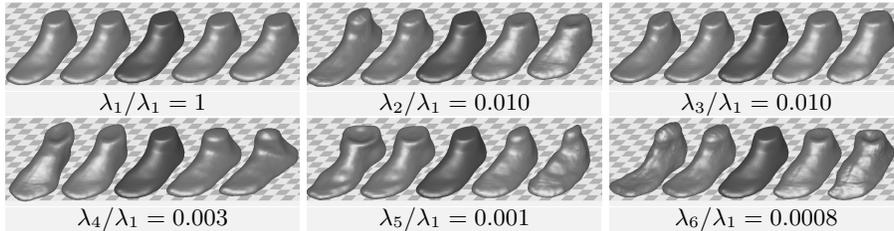
## 5 Viscous fluid-based shape space

As explained in Sec. 3.1, the viscous fluid shape space is by construction a (infinite-dimensional) Riemannian manifold and as such is based on the computation of shape paths as opposed to state-based approaches like the elastic shape space from the previous section. In the elastic, state-based approach, we have to find for each two shapes  $\mathcal{S}_A = \partial\mathcal{O}_A$  and  $\mathcal{S}_B = \partial\mathcal{O}_B$  one single optimal matching deformation  $\phi : \mathcal{O}_A \rightarrow \mathbb{R}^d$  via which the similarity between  $\mathcal{S}_A$  and  $\mathcal{S}_B$  is determined. In contrast, here we require more information to measure the distance between the two shapes, namely an optimal velocity field  $v(t) : \mathcal{O}(t) \rightarrow \mathbb{R}^d$  at each time  $t$  within the given time interval  $[0, 1]$ . In effect, this implies an increase of the dimension of the variational problem by the time component.

The two qualitatively different types of coordinates, the space coordinates (that span the space in which the shapes lie) and the time coordinate, are intuitively treated in different ways. One possibility is to regard the variational



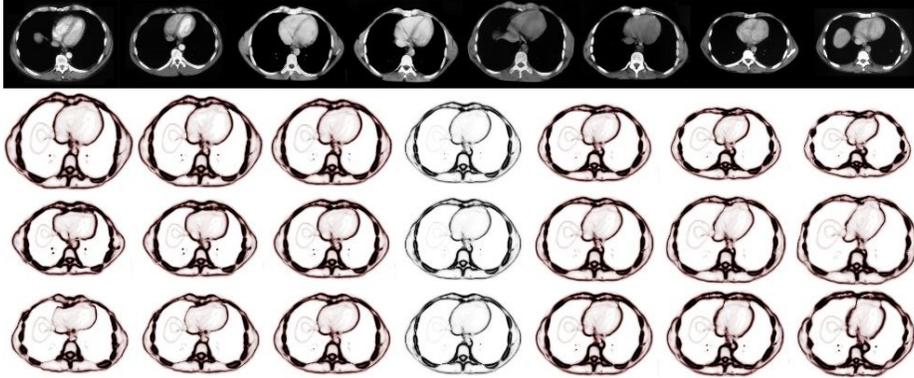
**Fig. 16.** 48 input kidneys (courtesy of Werner Bautz, radiology department at the University Hospital Erlangen, Germany) and their first four modes of variation with ratios  $\frac{\lambda_i}{\lambda_1}$  of 1, 0.72, 0.37, and 0.31.



**Fig. 17.** The first six dominant modes of variation for the feet from Fig. 12.

problem of computing a geodesic as a classical elliptic boundary value problem in time, in which each shape on a path seeks to be in equilibrium with its local neighborhood on the path. The equilibrizing force can be interpreted as an acceleration acting on the velocity field  $v$ . In this setting, it seems most natural to discretize first the time variable and approximate geodesics in shape space as discrete sequences  $\mathcal{S}_0, \dots, \mathcal{S}_K$  of shapes, where each shape is connected to and equilibrates with its neighbors and the path length along the discrete path  $\mathcal{S}_0, \dots, \mathcal{S}_K$  is approximated as a sum  $\sum_{k=1}^K \tilde{d}(\mathcal{S}_{k-1}, \mathcal{S}_k)$  of approximations  $\tilde{d}(\mathcal{S}_{k-1}, \mathcal{S}_k)$  of the geodesic distance between neighboring shapes. The distance  $\tilde{d}$  can be based on a matching deformation energy which will be elaborated on further down.

An alternative view starts from the underlying velocity field which generates the geodesic. Dupuis et al. [54] and Beg et al. [56] consider shapes (or rather images) embedded in a domain  $\Omega \subset \mathbb{R}^d$ . These shapes deform according to smooth, compactly supported velocity fields  $v \in L^2([0, 1]; W_0^{n,2}(\Omega; \mathbb{R}^d))$  with  $n > 2 + \frac{2}{d}$ . The regularity of the velocity fields is ensured by defining the path dissipation as  $\int_0^1 \int_{\Omega} Lv \cdot v \, dx \, dt$  and the path length as  $\int_0^1 \sqrt{\int_{\Omega} Lv \cdot v \, dx \, dt}$  for a differential operator  $L$  of sufficiently high order (cf. Section 3.1). The corresponding shape deformation  $\phi$  which is induced by the velocity field is obtained



**Fig. 18.** 8 thorax CT scans from different patients (courtesy of Bruno Wirth, urology department at the Hospital zum hl. Geist, Kempen, Germany) and their first three modes of variation with ratios  $\frac{\lambda_i}{\lambda_1}$  of 1, 0.12, and 0.07. Note that the thin lines which can be seen left of the heart correspond to contours of the liver, which are only visible in the first and last input image.

as the solution  $\phi = \phi_1$  of the pointwise, Lagrangian ordinary differential equation  $\frac{d}{dt}\phi_t(x) = v(\phi_t(x), t)$ .

In the first approach, the computation of a geodesic was seen as the concatenation of a number of local subproblems each of which represents the approximation of a geodesic segment between two intermediate shapes and each of which thus inherits the constraint that one shape is transferred exactly into the other. In contrast, in the second approach we have one single constraint, acting at the end of the geodesic and expressing that the accumulated flow  $\phi$  deforms the starting shape  $\mathcal{S}_A$  into the final shape  $\mathcal{S}_B$ ,  $\phi(\mathcal{S}_A) = \mathcal{S}_B$ .

Let us now focus on the first approach in which a geodesic path will be approximated via a finite sequence of shapes  $\mathcal{S}_0, \dots, \mathcal{S}_K$ , connected by deformations  $\phi_k : \mathcal{O}_{k-1} \rightarrow \mathbb{R}^d$  which are optimal in a variational sense and fulfill the constraint  $\phi_k(\mathcal{S}_{k-1}) = \mathcal{S}_k$ .

Given two shapes  $\mathcal{S}_A, \mathcal{S}_B$  in some given space of shapes  $\mathbf{S}$ , we define a discrete path of shapes as a sequence of shapes  $\mathcal{S}_0, \dots, \mathcal{S}_K \in \mathbf{S}$  with  $\mathcal{S}_0 = \mathcal{S}_A$  and  $\mathcal{S}_K = \mathcal{S}_B$ . For the time step  $\tau = \frac{1}{K}$  the shape  $\mathcal{S}_k$  is supposed to be an approximation of  $\mathcal{S}(t_k)$  with  $t_k = k\tau$ , where  $(\mathcal{S}(t))_{t \in [0,1]}$  is a continuous path connecting  $\mathcal{S}_A = \mathcal{S}(0)$  and  $\mathcal{S}_B = \mathcal{S}(1)$ . For each pair of consecutive shapes  $\mathcal{S}_{k-1}$  and  $\mathcal{S}_k$  we now consider a matching deformation  $\phi_k : \mathcal{O}_{k-1} \rightarrow \mathbb{R}^d$  which satisfies  $\phi_k(\mathcal{S}_{k-1}) = \mathcal{S}_k$ . With each deformation  $\phi_k$  we associate a deformation energy  $\mathcal{W}_{\text{deform}}[\phi_k, \mathcal{O}_{k-1}] = \int_{\mathcal{O}_{k-1}} W(\mathcal{D}\phi_k) dx$  of the same type as described in Sec. 3.2. If appropriately chosen, this energy will ensure sufficient regularity and a 1-1 matching property for deformations  $\phi_k$  with finite energy. As in elasticity, the energy is assumed to depend only on the local deformation, reflected by the Jacobian  $\mathcal{D}\phi$ . Yet, different from elasticity, we suppose the material to relax

instantaneously so that object  $\mathcal{O}_k$  is again in a stress-free configuration when applying  $\phi_{k+1}$  at the next time step. Let us also emphasize that the stored energy does not depend on the deformation history as in most plasticity models in engineering. This energy is now employed to define time-discrete counterparts to the dissipation and length of continuous paths from Sec. 3.1.

**Definition 3 (Discrete dissipation and discrete path length)** *Given a discrete path  $\mathcal{S}_0, \dots, \mathcal{S}_K \in \mathbf{S}$ , its dissipation is defined as*

$$\mathbf{Diss}_\tau(\mathcal{S}_0, \dots, \mathcal{S}_K) := \sum_{k=1}^K \frac{1}{\tau} \mathcal{W}_{\text{deform}}[\phi_k, \mathcal{O}_{k-1}],$$

where  $\phi_k : \overline{\mathcal{O}_{k-1}} \rightarrow \mathbb{R}^d$  is a minimizer of the deformation energy  $\mathcal{W}_{\text{deform}}[\phi_k, \cdot]$  under the constraint  $\phi_k(\mathcal{S}_{k-1}) = \mathcal{S}_k$ . Furthermore, the discrete path length is defined as

$$\mathbf{L}_\tau(\mathcal{S}_0, \dots, \mathcal{S}_K) := \sum_{k=1}^K \sqrt{\mathcal{W}_{\text{deform}}[\phi_k, \mathcal{O}_{k-1}]}.$$

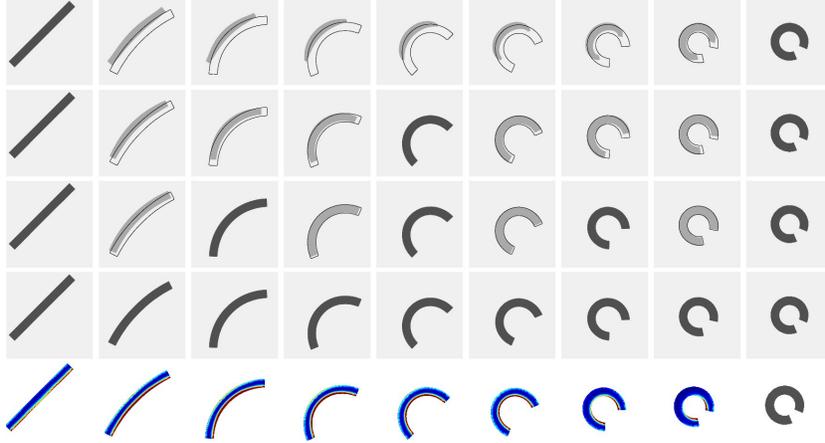
Let us make a brief remark on the proper scaling factors. The deformation energy  $\mathcal{W}_{\text{deform}}[\phi_k, \mathcal{O}_{k-1}]$  is expected to scale like  $\tau^2$  (cf. (7)). Hence, the factor  $\frac{1}{\tau}$  ensures the discrete dissipation measure to be conceptually independent of the time step size. The same holds for the discrete length measure  $\mathbf{L}_\tau(\mathcal{S}_0, \dots, \mathcal{S}_K)$ .

To ensure that the above-defined dissipation and length of discrete paths in shape space are well-defined, a minimizing deformation  $\phi_k$  of the elastic energy  $\mathcal{W}_{\text{deform}}[\cdot, \mathcal{O}_{k-1}]$  with  $\phi_k(\mathcal{S}_{k-1}) = \mathcal{S}_k$  has to exist. In fact, this holds for objects  $\mathcal{O}_{k-1}$  and  $\mathcal{O}_k$  with Lipschitz boundaries  $\mathcal{S}_{k-1}$  and  $\mathcal{S}_k$  for which there exists at least one bi-Lipschitz deformation  $\phi_k$  of  $\mathcal{O}_{k-1}$  into  $\mathcal{O}_k$  for  $k = 1, \dots, K$  [19].

With the notion of dissipation at hand we can define a discrete geodesic path following the standard paradigms in differential geometry.

**Definition 4 (Discrete geodesic path)** *A discrete path  $\mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_K$  in a set of admissible shapes  $\mathbf{S}$  connecting two shapes  $\mathcal{S}_A = \mathcal{S}_0$  and  $\mathcal{S}_B = \mathcal{S}_K$  is a discrete geodesic if there exists an associated family of deformations  $(\phi_k)_{k=1, \dots, K}$  such that  $(\phi_k, \mathcal{S}_k)_{k=1, \dots, K}$  minimize the total energy  $\sum_{k=1}^K \frac{1}{\tau} \mathcal{W}_{\text{deform}}[\tilde{\phi}_k, \tilde{\mathcal{O}}_{k-1}]$  over all intermediate shapes  $\tilde{\mathcal{S}}_1 = \partial \tilde{\mathcal{O}}_1, \dots, \tilde{\mathcal{S}}_{K-1} = \partial \tilde{\mathcal{O}}_{K-1} \in \mathbf{S}$  and all possible matching deformations  $\tilde{\phi}_1, \dots, \tilde{\phi}_K$  with  $\tilde{\phi}_k(\tilde{\mathcal{S}}_{k-1}) = \tilde{\mathcal{S}}_k$  for  $k = 1, \dots, K$ .*

Examples of discrete geodesics are provided in Figs. 19 to 20. Apparently, the frame indifference and the (local) injectivity property of the matching deformations, which are ensured by the nonlinear deformation energy  $\mathcal{W}_{\text{deform}}$ , allow the computation of reasonable discrete geodesics with only few intermediate shapes. Under sufficient growth conditions on the integrand of the deformation energy  $\mathcal{W}_{\text{deform}}$ , the existence of discrete geodesics is guaranteed at least for certain compact sets  $\mathbf{S}$  of admissible shapes, e. g. shapes  $\mathcal{S}$  which can be described by spline curves with a finite set of control points from some compact domain and which satisfy a uniform cone condition in the sense that each  $x \in \mathcal{S}$  is

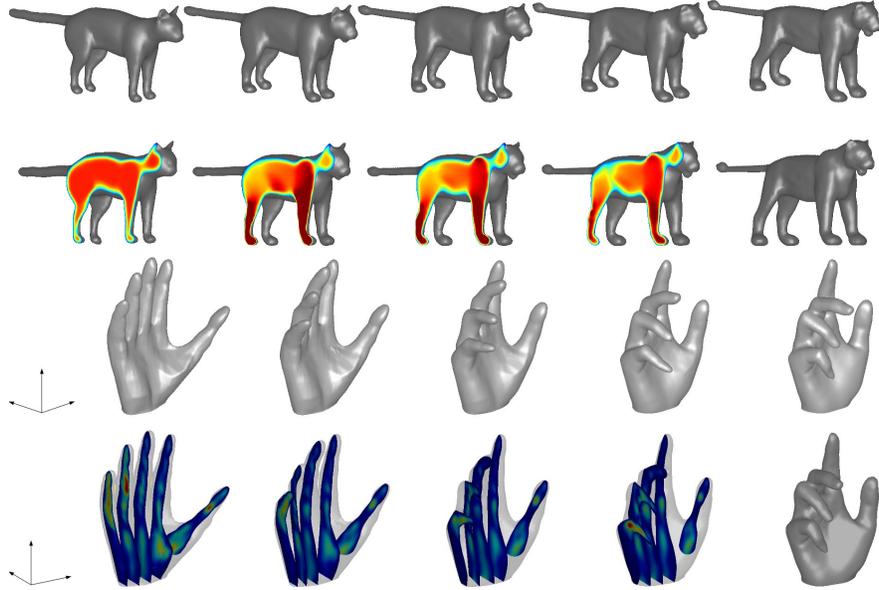


**Fig. 19.** Discrete geodesics between a straight and a rolled up bar, from first row to fourth row based on 1, 2, 4, and 8 time steps. The light gray shapes in the first, second, and third row show a linear interpolation of the deformations connecting the dark gray shapes. The shapes from the finest time discretization are overlaid over the others as thin black lines. In the last row the rate of viscous dissipation is rendered on the shape domains  $\mathcal{O}_1, \dots, \mathcal{O}_7$  from the previous row, color-coded as .

the tip of two cones with fixed height and opening angle which lie completely on either side of  $\mathcal{S}$  [19]. Such requirements on  $\mathbf{S}$  are necessary since the known regularity theory for deformation energies of the employed type does not allow to prove Lipschitz-regularity of optimal deformations so that the intermediate shapes might degenerate.

The discrete dissipation as the sum of matching deformation energies indeed represents an approximation to the time-continuous dissipation of a velocity field from Sec. 3.1. If a smooth path in shape space is considered which is interpolated at discrete times  $t_k = k\tau$ ,  $k = 0, \dots, K$ , and if for  $t \in [t_{k-1}, t_k)$ ,  $v_\tau(t) = \frac{(\phi_k - \mathbb{1})}{\tau} \circ \left( \frac{t_k - t}{\tau} \mathbb{1} + \frac{t - t_{k-1}}{\tau} \phi_k \right)^{-1}$  denotes the velocity field which generates the associated matching deformations  $\phi_k$ , then as the time step size  $\tau = \frac{1}{K}$  decreases and  $v_\tau$  converges against a smooth velocity field  $v$ , the discrete dissipation converges against the time-continuous dissipation (2) induced by  $v$  (cf. [19] for details).

Within this framework of geodesics in shape space, the strict constraints that one shape is deformed exactly into another one are often inadequate in applications as has already been discussed in Sec. 3.2 for the state-based, elastic setup. For the computation of an elastic distance, the single matching constraint could be relaxed as a mismatch penalty. In the Riemannian, viscous setting we pursue the same concept, however, the particular form of the employed constraints depends on the chosen view on shape geodesics. In the frame-

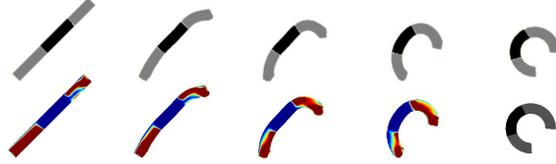


**Fig. 20.** Discrete geodesic between a cat and a lion and between the hand shapes m336 and m324 from the Princeton Shape Benchmark [77]. For both examples, the local dissipation is color-coded on slices through the shapes as .

work of geodesics as paths of diffeomorphisms, which we introduced at the beginning of this section, there is the single constraint  $\phi(\mathcal{S}_A) = \mathcal{S}_B$ , meaning that the induced diffeomorphism  $\phi$  maps the initial shape  $\mathcal{S}_A$  onto the final shape  $\mathcal{S}_B$ . This constraint can be relaxed in the same manner as in Sec. 3.2 via a penalty measuring the mismatch of the shapes or of the corresponding objects. For the time-discrete geodesic setting we have a sequence of matching constraints  $\phi_k(\mathcal{S}_{k-1}) = \mathcal{S}_k$ ,  $k = 1, \dots, K$ , each of which can again be relaxed by the same means. In fact, we add to the discrete dissipation of a set  $(\phi_k)_{k=1, \dots, K}$  of deformations a sum of mismatch penalties  $\sum_{k=1}^K \text{vol}(\mathcal{O}_{k-1} \Delta \phi_k^{-1}(\mathcal{O}_k))$ . In the limit for vanishing time step size  $\tau = \frac{1}{K}$  and under the same conditions as above, this sum can be shown to converge against the optical flow type functional  $\int_{\mathcal{T}} |(1, v(t)) \cdot n[t, \mathcal{S}(t)]| da$  for the unit outward normal  $n[t, \mathcal{S}(t)]$  to the space time shape tube  $\mathcal{T} = \bigcup_{t \in [0, 1]} \{t\} \times \mathcal{S}(t)$ . Furthermore,  $\sum_{k=1}^K \tau \mathcal{L}[\mathcal{S}_k]$  with  $\mathcal{L}[\mathcal{S}_k] = \mathcal{H}^{d-1}(\mathcal{S}_k)$  has been employed as regularization, which in the limit for  $\tau \rightarrow 0$  converges against the integral  $\int_0^1 \mathcal{H}^{d-1}(\mathcal{S}(t)) dt$ .

Real world objects are most often not only characterized by their outer contour, but contain also internal structures that have to be matched properly when computing the similarity between two objects. As an example, consider the straight and the folded rod in Fig. 21. The rods consist of three distinct components, which imposes a constraint on reasonable connecting paths: Each

component is to be mapped onto its correct counterpart. A shortest path under this constraint obviously differs significantly from the geodesic which just matches the outer contours (cf. Fig. 19).



**Fig. 21.** Discrete geodesic between the straight and the folded bar from Fig. 19, where the black region of the initial shape is constrained to be matched to the black region of the final shape. The bottom row shows a color-coding of the corresponding viscous dissipation. Due to the strong change in relative position of the black region, the intermediate shapes exhibit a strong asymmetry and high dissipation near the bar ends.

This observation calls for a generalization of shapes, an example of which we have already seen in the context of an elastic shape space in Fig. 18, where the edge set of an image was considered as a shape. Here, let us adopt a slightly different approach and regard shapes as being composed of a number of sub-components. In detail, instead of a geodesic between just two shapes  $\mathcal{S}_A = \partial\mathcal{O}_A$  and  $\mathcal{S}_B = \partial\mathcal{O}_B$ , we now seek a geodesic path  $(\mathcal{S}^i(t))_{i=1,\dots,m}$  with  $\mathcal{S}^i(t) = \partial\mathcal{O}^i(t)$  for  $t \in [0, 1]$ , between two collections of  $m$  separate shapes,  $(\mathcal{S}_A^i)_{i=1,\dots,m}$  with  $\mathcal{S}_A^i(t) = \partial\mathcal{O}_A^i(t)$  and  $(\mathcal{S}_B^i)_{i=1,\dots,m}$  with  $\mathcal{S}_B^i(t) = \partial\mathcal{O}_B^i(t)$ . The geodesic path is supposed to be generated by a joint motion field  $v(t) : \bigcup_{i=1}^m \mathcal{O}^i(t) \rightarrow \mathbb{R}^d$ . The single objects  $\mathcal{O}^i(t)$  can then be regarded as the subcomponents of an overall object  $\bigcup_{i=1}^m \mathcal{O}^i(t)$ . The total dissipation along the path is measured exactly as before by

$$\mathbf{Diss} \left[ (v(t), (\mathcal{O}^i(t))_{i=1,\dots,m})_{t \in [0,1]} \right] = \int_0^1 \int_{\bigcup_{i=1}^m \mathcal{O}^i(t)} \frac{\lambda}{2} (\operatorname{tr} \epsilon[v])^2 + \mu \operatorname{tr}(\epsilon[v]^2) \, dx \, dt.$$

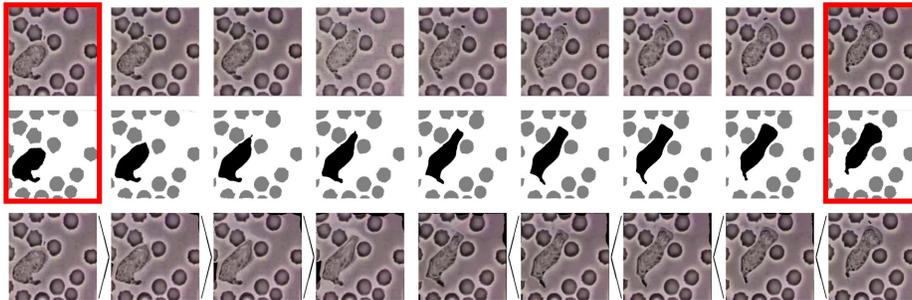
This naturally translates to the discrete dissipation of a path with  $K + 1$  intermediate shape collections  $(\mathcal{S}_k^i)_{i=1,\dots,m}$ ,  $k = 0, \dots, K$ ,

$$\sum_{k=1}^K \mathcal{W}_{\text{deform}}[\phi_k, (\mathcal{O}_{k-1}^i)_{i=1,\dots,m}] := \sum_{k=1}^K \int_{\bigcup_{i=1}^m \mathcal{O}_{k-1}^i} W(\mathcal{D}\phi_k) \, dx,$$

where the deformations  $\phi_k$  satisfy the constraints  $\phi_k(\mathcal{S}_{k-1}^i) = \mathcal{S}_k^i$  for  $k = 1, \dots, K$ ,  $i = 1, \dots, m$ , and  $\mathcal{S}_0^i = \mathcal{S}_A^i$ ,  $\mathcal{S}_K^i = \mathcal{S}_B^i$ ,  $i = 1, \dots, m$ .

The different object components can of course be assigned different material properties. Fig. 22 shows frames from a real video sequence of moving white and red blood cells (top) as well as a discrete geodesic between the first and

last frame (middle) for which the material parameters of the white blood cell were chosen twenty times weaker than for the red blood cells. The result is a nonlinear interpolation between distant frames which is in good agreement with the actually observed motion. Once geodesic distances between shapes are



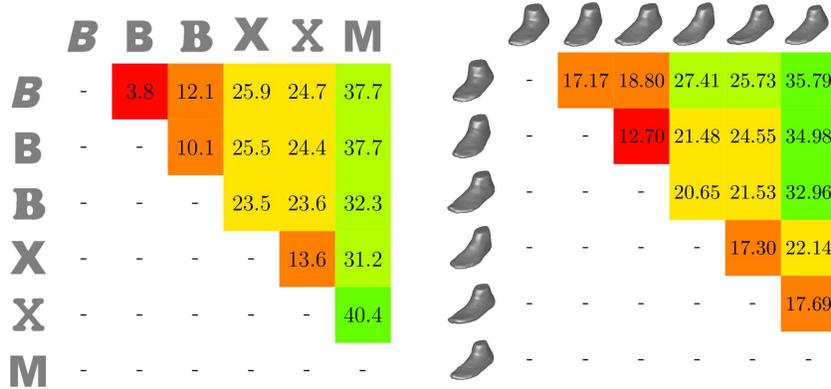
**Fig. 22.** Top: Frames from a real video sequence of a white blood cell among a number of red ones (courtesy Robert A. Freitas, Institute for Molecular Manufacturing, California, USA). Middle: Computed discrete geodesic between the segmented shapes in the first and the last frame. Bottom: Pushforward of the initial (first four shapes) and pullback of the final frame (last five shapes) according to the geodesic flow.

defined, one can statistically analyze ensembles of shapes and cluster them in groups based on the geodesic distance as a reliable measure for the similarity of shapes. Two exemplary examples are provided by the evaluation of geodesic distances between different 2D letters (Fig. 23, left) and between six different 3D foot shapes (Fig. 23, right). In the 2D example, we clearly identify three distinct clusters ( $B$ s,  $X$ s, and  $M$ ).

## 6 A comparison of path- and state-based shape space

Already in Sec. 3.3 we have studied the difference between the state-based dissimilarity measure  $d_{\text{elast}}$  and the path-based distance  $d_{\text{viscous}}$ . Based on the applications considered in the previous sections let us compare the underlying concepts now more on a conceptual level of the geometry of shape space:

- *Non-uniqueness of shape averages.* Due to the nonlinearity of the elastic variational problem, local minimizers of the elastic energy might be non-unique. There might even exist different minimizing deformations with the same elastic energy. Mechanically, this non-uniqueness is frequently associated with different buckling modes, which occur in case of large, geometrically nonlinear deformations. Hence, the shape average need not be uniquely defined, except in the small displacement case, where a linear elastic model (8) applies. In case of the path-based approach, (shortest) geodesics do not have



**Fig. 23.** Left: Pairwise geodesic distances between (also topologically) different letter shapes. Right: Pairwise geodesic distances between different scanned 3D feet. The feet have volumes  $499.5 \text{ cm}^3$ ,  $500.6 \text{ cm}^3$ ,  $497.6 \text{ cm}^3$ ,  $434.7 \text{ cm}^3$ ,  $432.0 \text{ cm}^3$ , and  $381.0 \text{ cm}^3$ , respectively.

to be unique either. Indeed, a geodesic is the unique shortest path only until the first conjugate point. Hence, the shape average is in a strict sense not well-defined if the distances are sufficiently large.

- *Different physical interpretation of the PCA.* In the Riemannian setup with the metric being the rate of viscous dissipation, the  $\log_{\mathcal{S}} \mathcal{S}_k$  corresponds to the initial velocity  $v_k : \mathcal{S} \rightarrow \mathbb{R}^d$  in the (optimal transport) flow of  $\mathcal{O}$  associated with shape  $\mathcal{S}$  into  $\mathcal{O}_k$  associated with the  $k$ th input shape  $\mathcal{S}_k$ . In the elastic model, the boundary stress  $\sigma_k \nu : \partial \mathcal{O} \rightarrow \mathbb{R}^d$  results from the deformation  $\phi_k$  of  $\mathcal{O}_k$  onto the average object  $\mathcal{O}$  and effectively is the restoring force acting on the average shape  $\mathcal{S}$ . Via the linearized elasticity problem in the pre-stressed compound configuration of the average object  $\mathcal{O}$ , these restoring forces are identified with displacements  $u_k$ . Depending on the model, either the flow velocities  $v_k$  or the linear elastic displacements  $u_k$  form the basis of a covariance analysis in the linear vector space of mappings  $\overline{\mathcal{O}} \rightarrow \mathbb{R}^d$ . The outcome are principal shape variations of the average shape, either generated by motion fields or displacements, respectively.
- *Quantitative shape analysis.* The Riemannian metric given by the rate of viscous dissipation in the path-based viscous fluid approach allows direct comparison of multiple ensembles of shapes via pairwise distance computations. Due to the lack of a triangle inequality this is possible only in a restricted sense in the state-based elastic approach, where dissimilarity measures for one fixed shape and a set of varying shapes can be computed.
- The method of choice depends on the *concrete application*. If shapes are considered as boundaries of objects with a viscous fluid inside then the path-based approach would be more appropriate. The state-based elastic approach is favorable for objects which behave more like deformable solids.

## 7 A collection of computational tools

So far, we have investigated some of the many aspects on mathematical models in shape space without any discussion of the corresponding computational tools and numerical algorithms. Hence, let us at least briefly mention some fundamental computational aspects to effectively deal with general classes of shapes as boundary contours of volumetric objects.

At first, we replace the strict separation between material inside the object and void outside by substituting the void with a material which is several orders of magnitude softer than inside the object. This relaxation is important with respect to the existence analysis and the stabilization of the computational method. In fact, we replace the deformation energy  $\mathcal{W}_{\text{deform}}[\phi, \mathcal{O}] = \int_{\mathcal{O}} W(\mathcal{D}\phi) dx$  by the energy  $\mathcal{W}_{\text{deform}}^{\eta}[\phi, \mathcal{O}] = \int_{\Omega} ((1 - \eta)\chi_{\mathcal{O}} + \eta) W(\mathcal{D}\phi) dx$  for a small constant  $\eta$ . In the implementation which underlies the above applications, for  $\eta = 10^{-4}$  one observes no significant qualitative impact of this regularization on the solution. Furthermore, as mentioned above, to ensure regularity of the shape contour  $\mathcal{S}$ , we take into account the area functional  $\mathcal{L}[\mathcal{S}] = \int_{\mathcal{S}} da$  as a prior, weighted with a small factor.

Compared to a parametric description of shapes, e. g. as a polygonal line or a triangular surface, an implicit description has several advantages. In particular, it does not require a remeshing even in case of large deformations, it allows for topological transitions without any extra handling of the associated singularities, and it can be combined with multi-scale relaxation schemes for an efficient minimization of the involved functionals.

In what follows, we consider a level set and a phase field description of shapes and outline the general framework of a multi-scale method based on finite element calculus. In fact, the phase field model has been used in the examples for the elastic shape averaging and the PCA, whereas the level set method has served as a numerical building block for the computation of time-discrete shape geodesics.

### 7.1 Shapes described by level set functions

The level set method first presented by Osher and Sethian [78] has been used for a wide range of applications [79,80]. Burger and Osher gave an overview in the context of shape optimization [81]. To numerically solve variational problems in shape space we assume a shape  $\mathcal{S}$  to be represented by the zero level set  $\{x \in \Omega : u(x) = 0\}$  of a scalar function  $u : \Omega \rightarrow \mathbb{R}$  on a computational domain  $\Omega \subset \mathbb{R}^d$ . Furthermore, the zero super level set  $\{x \in \Omega : u(x) > 0\}$  defines the corresponding object domain  $\mathcal{O}$ . This shape description can be incorporated in a variational approach following the approximation proposed by Chan and Vese [82]. In fact, the partition of the domain  $\Omega$  into object and background is encoded via a regularized Heaviside function  $H_{\varepsilon} \circ u$ . As in [82] we consider the function  $H_{\varepsilon}(x) := \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x}{\varepsilon}\right)$ , where  $\varepsilon$  is a scale parameter representing the width of the smeared-out shape contour. Then, a deformation energy  $\mathcal{W}_{\text{deform}}^{\eta}[\phi, \mathcal{O}] =$

$\int_{\Omega} ((1 - \eta)\chi_{\mathcal{O}} + \eta) W(\mathcal{D}\phi) dx$  is approximated by

$$\mathcal{W}_{\text{deform}}^{\varepsilon, \eta}[\phi, u] = \int_{\Omega} ((1 - \eta)H_{\varepsilon}(u) + \eta) W(\mathcal{D}\phi) dx.$$

Furthermore, the energy  $\mathcal{F}[\mathcal{S}_A, \phi, \mathcal{S}_B] = \text{vol}(\mathcal{O}_A \Delta \phi^{-1}(\mathcal{O}_B))$  measuring the volumetric mismatch between an object  $\mathcal{O}_A$  and the pullback of an object  $\mathcal{O}_B$  under a deformation  $\phi$  can be approximated by

$$\mathcal{F}^{\varepsilon}[u_A, \phi, u_B] = \int_{\Omega} (H_{\varepsilon}(u_B \circ \phi) - H_{\varepsilon}(u_A))^2 dx,$$

where  $u_A, u_B$  are level set representations of the shapes  $\mathcal{S}_A$  and  $\mathcal{S}_B$ , respectively. Finally, the surface area of a shape  $\mathcal{S}$ , which appears as a prior, is replaced by the total variation of  $H_{\varepsilon} \circ u$ , and we obtain

$$\mathcal{L}^{\varepsilon}[u] = \int_{\Omega} |\nabla H_{\varepsilon}(u)| dx.$$

Let us emphasize that in the actual energy minimization algorithm, the guidance of an initial zero level set towards the final shape relies on the nonlocal support of the derivative of the regularized Heaviside function (cf. [83]).

## 7.2 Shapes described via phase fields

An alternative to a level set description of shapes is a phase field representation. Physically, the phase field approach is inspired by the observation that interfaces are usually not sharp but characterized by a diffusive transition. Mathematically, there are two basic types of such phase field representations, a single phase approach as the one presented by Ambrosio and Tortorelli [84] for the approximation of the Mumford–Shah model [85] and the double phase approach by Modica and Mortola [86] used to approximate surface integrals. In the shape context studied here, let us focus on the single phase model. Thus, a shape  $\mathcal{S}$  is encoded by a continuous, piecewise smooth phase field function  $u : \Omega \rightarrow \mathbb{R}$  which is zero on  $\mathcal{S}$ , but close to one everywhere else. The specific profile of the phase field function  $u$  for a shape  $\mathcal{S}$  is determined via the phase field approximation

$$\mathcal{L}^{\varepsilon}[u] = \frac{1}{2} \int_{\Omega} \varepsilon |\nabla u|^2 + \frac{1}{\varepsilon} (u - 1)^2 dx$$

of the involved surface area  $\int_{\mathcal{S}} da$ . As in the above level set model the phase field parameter  $\varepsilon$  determines the width of the diffusive interface. Different from the level set model by Chan and Vese, the interface profile is not explicitly prescribed but implicitly encoded in the variational approach as the profile attained by minimizers of the functional. Based on this phase field model the penalty functional  $\mathcal{F}[\mathcal{S}_A, \phi, \mathcal{S}_B] = \mathcal{H}^{d-1}(\mathcal{S}_A \Delta \phi^{-1}(\mathcal{S}_B))$  measuring the area mismatch between a shape  $\mathcal{S}_A$  and the pullback of a shape  $\mathcal{S}_B$  under a deformation  $\phi$  can be approximated by

$$\mathcal{F}^{\varepsilon}[u_A, \phi, u_B] = \frac{1}{\varepsilon} \int_{\Omega} (u_B \circ \phi)^2 (1 - u_A)^2 + u_A^2 (1 - u_B \circ \phi)^2 dx,$$

where  $u_A, u_B$  are phase fields representing the shapes  $\mathcal{S}_A$  and  $\mathcal{S}_B$ , respectively. In this type of models the deformation energy  $\mathcal{W}_{\text{deform}}^\eta[\phi, \mathcal{O}]$  cannot be realized based on a phase field function  $u$  due to the fact that a single phase model allows to identify the shape itself but does not distinguish its inside and outside. Therefore, in the presented applications of elastic shape averaging and the elastic PCA the input objects and thus their characteristic functions  $\chi_{\mathcal{O}}$  were given a priori.

### 7.3 Multi-scale finite element approximation

For the spatial discretization of the functionals in the above variational approaches the finite element method can be applied. Hence, the level set function or the phase field  $u$ , representing a (unknown) shape  $\mathcal{S}$ , and the different components of the deformations  $\phi$  are represented by continuous, piecewise multilinear (trilinear in 3D and bilinear in 2D) finite element functions  $U$  and  $\Phi$  on a regular grid superimposed on the domain  $\Omega = [0, 1]^d$ . For the ease of implementation a dyadic grid resolution with  $2^L + 1$  vertices in each direction and a grid size  $h = 2^{-L}$  is chosen.

*Descent algorithm.* The functionals depend nonlinearly both on the discrete deformations  $\Phi$  (due to the concatenation  $U \circ \Phi$  and the nonlinear integrand  $W(\cdot)$  of the deformation energy) as well as on the discrete level set or phase field functions  $U$  (e. g. due to the concatenation of the level set function with the regularized Heaviside function  $H_\varepsilon(\cdot)$ ). In our energy relaxation algorithm for fixed grid size, we employ a gradient descent approach. We constantly alternate between performing a single gradient descent step for all deformations and the level set or phase field functions.

*Numerical quadrature.* Integral evaluations in the descent algorithm are performed by Gaussian quadrature of third order on each grid cell. For various terms we have to evaluate pullbacks  $U \circ \Phi$  of a discretized level set function  $U$  or a test function under a discretized deformation  $\Phi$ . Let us emphasize that quadrature based on nodal interpolation of  $U \circ \Phi$  would lead to artificial displacements near the shape edges accompanied by strong artificial tension. Hence, in our algorithm, if  $\Phi(x)$  lies inside  $\Omega$  for a quadrature point  $x$ , then the pullback is evaluated exactly at  $x$ . Otherwise, we project  $\Phi(x)$  back onto the boundary of  $\Omega$  and evaluate  $U$  at that projection point.

*Cascadic multi-scale algorithm.* The variational problem considered here is highly nonlinear, and for fixed time step size the proposed scheme is expected to have very slow convergence; also it might end up in some nearby local minimum. Here, a multi-level approach (initial optimization on a coarse scale and successive refinement) turns out to be indispensable in order to accelerate convergence and not to be trapped in undesirable local minima. Due to our assumption of a dyadic resolution  $2^L + 1$  in each grid direction, we are able to build a hierarchy of grids with  $2^l + 1$  nodes in each direction for  $l = L, \dots, 0$ . Via a simple restriction operation we project every finite element function to any of these coarse grid spaces. Starting the optimization on a coarse grid, the results from coarse scales

are successively prolonged onto the next grid level for a refinement of the solution [87]. Hence, the construction of a grid hierarchy allows to solve coarse scale problems in our multi-scale approach on coarse grids. Since the width  $\varepsilon$  of the diffusive shape representation should naturally scale with the grid width  $h$ , we choose  $\varepsilon = h$ .

## Acknowledgements

The model proposed in Sec. 5 has been developed in cooperation with Leah Bar and Guillermo Sapiro from the University of Minnesota. Benedikt Wirth has been funded by the Bonn International Graduate School in Mathematics. Furthermore, the work was supported by the Deutsche Forschungsgemeinschaft, SPP 1253 "Optimization with Partial Differential Equations".

## References

1. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. *Computer Vision and Image Understanding* **61**(1) (1995) 38–59
2. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical shape influence in geodesic active contours. In: 5th IEEE EMBS International Summer School on Biomedical Imaging, 2002. (2002)
3. Cremers, D., Kohlberger, T., Schnörr, C.: Shape statistics in kernel space for variational image segmentation. *Pattern Recognition* **36** (2003) 1929–1943
4. Elad (Elbaz), A., Kimmel, R.: On bending invariant signatures for surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(10) (2003) 1285–1295
5. Mémoli, F., Sapiro, G.: A theoretical and computational framework for isometry invariant recognition of point cloud data. *Foundations of Computational Mathematics* **5** (2005) 313–347
6. McNeill, G., Vijayakumar, S.: 2d shape classification and retrieval. In: Proceedings of the 19th international joint conference on artificial intelligence. (2005) 1483–1488
7. Ling, H., Jacobs, D.W.: Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(2) (2007) 286–299
8. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* **23** (2004) 151–160 Supplement 1.
9. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging* **22**(8) (2003) 1014–1025
10. Christensen, G.E., Rabbitt, R.D., Miller, M.I.: 3D brain mapping using a deformable neuroanatomy. *Phys. Med. Biol.* **39**(3) (1994) 609–618
11. Kendall, D.G.: Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* **16** (1984) 81–121
12. Klassen, E., Srivastava, A., Mio, W., Joshi, S.H.: Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(3) (2004) 372–383

13. Yezzi, A.J., Mennucci, A.: Conformal metrics and true “gradient flows” for curves. In: ICCV 2005: Proceedings of the 10th IEEE International Conference on Computer Vision. (2005) 913–919
14. Michor, P.W., Mumford, D., Shah, J., Younes, L.: A metric on shape space with explicit geodesics. *Rend. Lincei Mat. Appl.* (9) (2008) 25–57
15. Kilian, M., Mitra, N.J., Pottmann, H.: Geometric modeling in shape space. In: ACM Transactions on Graphics. Volume 26. (2007) #64, 1–8
16. Eckstein, I., Pons, J.P., Tong, Y., Kuo, C.C., Desbrun, M.: Generalized surface flows for mesh processing. In: Eurographics Symposium on Geometry Processing. (2007)
17. Fuchs, M., Jüttler, B., Scherzer, O., Yang, H.: Shape metrics based on elastic deformations. *J. Math. Imaging Vis.* **35**(1) (2009) 86–102
18. Rumpf, M., Wirth, B.: A nonlinear elastic shape averaging approach. *SIAM Journal on Imaging Sciences* **2**(3) (2009) 800–833
19. Wirth, B., Bar, L., Rumpf, M., Sapiro, G.: A continuum mechanical approach to geodesics in shape space. submitted to IJCV (2010)
20. Droske, M., Rumpf, M.: Multi scale joint segmentation and registration of image morphology. *IEEE Transaction on Pattern Recognition and Machine Intelligence* **29**(12) (2007) 2181–2194
21. Charpiat, G., Faugeras, O., Keriven, R.: Approximations of shape metrics and application to shape warping and empirical shape statistics. *Foundations of Computational Mathematics* **5**(1) (2005) 1–58
22. Miller, M.I., Younes, L.: Group actions, homeomorphisms, and matching: a general framework. *International Journal of Computer Vision* **41**(1–2) (2001) 61–84
23. Michor, P.W., Mumford, D.: Riemannian geometries on spaces of plane curves. *J. Eur. Math. Soc.* **8** (2006) 1–48
24. Fletcher, P.T., Lu, C., Joshi, S.: Statistics of shape via principal geodesic analysis on Lie groups. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR. Volume 1. (2003) 95–101
25. Fletcher, T., Venkatasubramanian, S., Joshi, S.: Robust statistics on Riemannian manifolds via the geometric median. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2008)
26. Berkels, B., Linkmann, G., Rumpf, M.: An  $SL(2)$  invariant shape median. (2009) submitted.
27. Charpiat, G., Faugeras, O., Keriven, R., Maurel, P.: Distance-based shape statistics. In: Acoustics, Speech and Signal Processing, 2006 (ICASSP 2006). Volume 5. (2006)
28. Rumpf, M., Wirth, B.: An elasticity approach to principal modes of shape variation. In: Proceedings of the Second International Conference on Scale Space Methods and Variational Methods in Computer Vision (SSVM 2009). Volume 5567 of Lecture Notes in Computer Science. (2009) 709–720
29. Chen, S.E., Parent, R.E.: Shape averaging and its applications to industrial design. *IEEE Computer Graphics and Applications* **9**(1) (1989) 47–54
30. Hafner, B., Zachariah, S., Sanders, J.: Characterisation of three-dimensional anatomic shapes using principal components: application to the proximal tibia. *Med. Biol. Eng. Comput.* **38** (2000) 9–16
31. Perperidis, D., Mohiaddin, R., Rueckert, D.: Construction of a 4d statistical atlas of the cardiac anatomy and its use in classification. In Duncan, J., Gerig, G., eds.: Medical Image Computing and Computer Assisted Intervention. Volume 3750 of LNCS. (2005) 402–410

32. Söhn, M., Birkner, M., Yan, D., Alber, M.: Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation. *Phys. Med. Biol.* **50** (2005) 5893–5908
33. Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W.E., Willsky, A.: A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging* **22**(2) (2003) 137–154
34. Dambreville, S., Rathi, Y., Tannenbaum, A.: A shape-based approach to robust image segmentation. In Campilho, A., Kamel, M., eds.: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Volume 4141 of LNCS. (2006) 173–183
35. Bronstein, A., Bronstein, M., Kimmel, R.: *Numerical Geometry of Non-Rigid Shapes*. Monographs in Computer Science. Springer (2008)
36. Mémoli, F.: Gromov-Hausdorff distances in euclidean spaces. In: *Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (CVPR workshop, NORDIA'08)*. (2008)
37. Manay, S., Cremers, D., Hong, B.W., Yezzi, A.J., Soatto, S.: Integral invariants for shape matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(10) (2006) 1602–1618
38. Hong, B.W., Soatto, S., Vese, L.: Enforcing local context into shape statistics. *Advances in Computational Mathematics* **online first** (2008)
39. Pennec, X., Stefanescu, R., Arsigny, V., Fillard, P., Ayache, N.: Riemannian elasticity: A statistical regularization framework for non-linear registration. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2005*. LNCS (2005) 943–950
40. Pennec, X.: Left-invariant Riemannian elasticity: a distance on shape diffeomorphisms? In: *Mathematical Foundations of Computational Anatomy - MFCA 2006*. (2006) 1–14
41. Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré* **10** (1948) 215–310
42. Karcher, H.: Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics* **30**(5) (1977) 509–541
43. Yushkevich, P., Fletcher, P.T., Joshi, S., Thalla, A., Pizer, S.M.: Continuous medial representations for geometric object modeling in 2d and 3d. *Image and Vision Computing* **21**(1) (2003) 17–27
44. Fletcher, P., Lu, C., Pizer, S., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. *Medical Imaging, IEEE Transactions on* **23**(8) (2004) 995–1005
45. Fuchs, M., Scherzer, O.: Segmentation of biologic image data with a-priori knowledge. FSP Report, Forschungsschwerpunkt S92 52, Universität Innsbruck (May 2007)
46. Fuchs, M., Scherzer, O.: Regularized reconstruction of shapes with statistical a priori knowledge. *International Journal of Computer Vision* **79**(2) (2008) 119–135
47. Liu, X., Shi, Y., Dinov, I., Mio, W.: A computational model of multidimensional shape. *International Journal of Computer Vision* **Online First** (2010)
48. Schmidt, F.R., Clausen, M., Cremers, D.: Shape matching by variational computation of geodesics on a manifold. In: *Pattern Recognition*. Volume 4174 of LNCS., Springer (2006) 142–151
49. Srivastava, A., Jain, A., Joshi, S., Kaziska, D.: Statistical shape models using elastic-string representations. In Narayanan, P., ed.: *Asian Conference on Computer Vision*. Volume 3851 of LNCS. (2006) 612–621

50. Sundaramoorthi, G., Yezzi, A., Mennucci, A.: Sobolev active contours. *International Journal of Computer Vision*. **73**(3) (2007) 345–366
51. Younes, L.: Computable elastic distances between shapes. *SIAM J. Appl. Math* **58**(2) (April 1998) 565–586
52. Delfour, M.C., Zolésio, J.: *Geometries and Shapes: Analysis, Differential Calculus and Optimization*. Adv. Des. Control 4. SIAM, Philadelphia (2001)
53. Zolésio, J.P.: Shape topology by tube geodesic. In: *IFIP Conference on System Modeling and Optimization No 21.*, (2004) 185–204
54. Dupuis, D., Grenander, U., Miller, M.: Variational problems on flows of diffeomorphisms for image matching. *Quarterly of Applied Mathematics* **56** (1998) 587–600
55. Miller, M., Trounev, A., Younes, L.: On the metrics and Euler-Lagrange equations of computational anatomy. *Annual Review of Biomedical Engineering* **4** (2002) 375–405
56. Beg, M.F., Miller, M.I., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision* **61**(2) (February 2005) 139–157
57. Vaillant, M., Glaunès, J.: Surface matching via currents. In: *IPMI 2005: Information Processing in Medical Imaging*. Volume 3565 of *Lecture Notes in Computer Science*. (2005) 381–392
58. Glaunès, J., Qiu, A., Miller, M.I., Younes, L.: Large deformation diffeomorphic metric curve mapping. *International Journal of Computer Vision* **80**(3) (2008) 317–336
59. Younes, L., Qiu, A., Winslow, R.L., Miller, M.I.: Transport of relational structures in groups of diffeomorphisms. *Journal of Mathematical Imaging and Vision* **32**(1) (2008) 41–56
60. Fletcher, P., Whitaker, R.: Riemannian metrics on the space of solid shapes. In: *MICCAI 2006: Med Image Comput Comput Assist Interv.* (2006)
61. Wirth, B., Bar, L., Rumpf, M., Sapiro, G.: Geodesics in shape space via variational time discretization. In: *Proceedings of the 7th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'09)*. Volume 5681 of *LNCS*. (2009) 288–302
62. Rathi, Y., Dambreville, S., Tannenbaum, A.: Statistical shape analysis using kernel PCA. In: *Proceedings of SPIE*. Volume 6064. (2006)
63. Rathi, Y., Dambreville, S., Tannenbaum, A.: Comparative analysis of kernel methods for statistical shape learning. In Beichel, R., Sonka, M., eds.: *computer vision approaches to medical image analysis*. Volume 4241 of *LNCS*. (2006) 96–107
64. Thorstensen, N., Segonne, F., Keriven, R.: Pre-image as karcher mean using diffusion maps: Application to shape and image denoising. In: *Proceedings of the Second International Conference on Scale Space Methods and Variational Methods in Computer Vision (SSVM 2009)*. Volume 5567 of *Lecture Notes in Computer Science*. (2009) 721–732
65. do Carmo, M.P.: *Riemannian Geometry*. Birkhäuser, Boston (1992)
66. Klingenberg, W.P.A.: *Riemannian Geometry*. Walter de Gruyter (1995)
67. Spivak, M.: *A comprehensive introduction to differential geometry*. Vol. I. Publish or Perish Inc., Boston, Mass. (1970)
68. Chorin, A.J., Marsden, J.E.: *A Mathematical Introduction to Fluid Mechanics*. Volume 4 of *Texts in Applied Mathematics*. Springer (1990)
69. Truesdell, C., Noll, W.: *The Non-Linear Field Theories of Mechanics*. Springer, Berlin (2004)
70. Nečas, J., Šilhavý, M.: Multipolar viscous fluids. *Quarterly of Applied Mathematics* **49**(2) (1991) 247–265

71. Ciarlet, P.G.: Three-dimensional elasticity. Elsevier Science Publishers B. V. (1988)
72. Marsden, J.E., Hughes, T.J.R.: Mathematical foundations of Elasticity. Prentice-Hall, Englewood Cliffs (1983)
73. Dacorogna, B.: Direct methods in the calculus of variations. Springer-Verlag, New York (1989)
74. Ogden, R.W.: Non-Linear Elastic Deformations. John Wiley (1984)
75. Ball, J.: Global invertibility of Sobolev functions and the interpenetration of matter. Proceedings of the royal Society of Edinburgh **88A** (1981) 315–328
76. Wirth, B.: Variational methods in shape space. Dissertation, University Bonn (2009)
77. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The princeton shape benchmark. In: Proceedings of the Shape Modeling International 2004, Genova. (2004) 167–178
78. Osher, S., Sethian, J.A.: Fronts propagating with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. Journal of Computational Physics **79**(1) (1988) 12–49
79. Osher, S., Fedkiw, R.: Level set methods and dynamic implicit surfaces. Volume 153 of Applied Mathematical Sciences. Springer-Verlag, New York (2003)
80. Sethian, J.A.: Level Set Methods and Fast Marching Methods. Cambridge University Press (1999)
81. Burger, M., Osher, S.J.: A survey on level set methods for inverse problems and optimal design. European Journal of Applied Mathematics **16**(2) (2005) 263–301
82. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on Image Processing **10**(2) (2001) 266–277
83. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision **22**(1) (1997) 61–79
84. Ambrosio, L., Tortorelli, V.M.: On the approximation of free discontinuity problems. Bollettino dell’Unione Matematica Italiana, Sezione B **6**(7) (1992) 105–123
85. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. Communications on Pure Applied Mathematics **42** (1989) 577–685
86. Modica, L., Mortola, S.: Un esempio di  $\Gamma$ -convergenza. Boll. Un. Mat. Ital. B (5) **14**(1) (1977) 285–299
87. Bornemann, F., Deuffhard, P.: The cascadic multigrid method for elliptic problems. Num. Math. **75**(2) (1996) 135–152