Basis restricted elastic shape analysis on the space of unregistered surfaces

Emmanuel Hartman, Emery Pierson, Martin Bauer, Mohamed Daoudi, Nicolas Charon

Basis restricted elastic shape analysis on the space of unregistered surfaces

Emmanuel Hartman[1†], Emery Pierson[2†], Martin Bauer[1*], Mohamed Daoudi[3,4], Nicolas Charon[5]

[1*]Department of Mathematics, Florida State University, Tallahassee, USA.
[2]LIX , Ecole Polytechnique, Palaiseau, 91120, France.
[3]Univ. Lille, CNRS, Centrale Lille, Institut Mines-Télécom, UMR 9189 CRIStAL, F-59000 Lille, France.
[4]IMT Nord Europe, Institut Mines-Télécom, Univ. Lille, Centre for Digital Systems, F-59000 Lille, France.
[5]Department of Mathematics, University of Houston, Houston, USA.

*Corresponding author(s). E-mail(s): bauer@math.fsu.edu;
Contributing authors: ehartman@math.fsu.edu; emery.pierson@courrier.dev;
mohamed.daoudi@imt-nord-europe.fr; ncharon@central.uh.edu;
†These authors contributed equally to this work.

Abstract

This paper introduces a new framework for surface analysis derived from the general setting of elastic Riemannian metrics on shape spaces. Traditionally, those metrics are defined over the infinite dimensional manifold of immersed surfaces and satisfy specific invariance properties enabling the comparison of surfaces modulo shape preserving transformations such as reparametrizations. The specificity of our approach is to restrict the space of allowable transformations to predefined finite dimensional bases of deformation fields. These are estimated in a data-driven way so as to emulate specific types of surface transformations. This allows us to simplify the representation of the corresponding shape space to a finite dimensional latent space. However, in sharp contrast with methods involving e.g. mesh autoencoders, the latent space is equipped with a non-Euclidean Riemannian metric inherited from the family of elastic metrics. We demonstrate how this model can be effectively implemented to perform a variety of tasks on surface meshes which, importantly, does not assume these to be pre-registered or to even have a consistent mesh structure. We specifically validate our approach on human body shape and pose data as well as human face and hand scans for problems such as shape registration, interpolation, motion transfer or random pose generation.

Keywords: Elastic Shape Analysis, Human Body Analysis, Geometric Deep Learning, Riemannian Shape Metrics

1 Introduction

Overview and Main Contributions: In this article, we introduce a novel pipeline designed to quantify the geometric difference between the shapes of surfaces. Furthermore, we are not only interested in quantifying shape differences between two individual data points, but we aim to estimate in addition plausible deformation processes between different shapes and allow for

statistical shape analysis tasks such as extrapolation of deformations, transposition to new data, and the generation of random shapes. Finally, the proposed model does not assume a consistent mesh structure across the data, making it applicable to a variety of tasks on surface meshes with real data.

Our approach is grounded in the field of Elastic Shape Analysis (ESA) (Srivastava and Klassen, 2016) and further leverages the varifold representation of surfaces (Charon and Trouvé, 2013), thereby bypassing the common requirement of having consistent mesh structures and available point correspondences across the dataset. In contrast to standard ESA, our method relies on enforcing specific structure on the deformation model via the introduction of a data driven basis for the space of admissible shape changes. In the terminology of machine learning, this can be interpreted as a latent space representation but, unlike typical approaches involving autoencoders, in our framework, this latent space is equipped with a non-Euclidean Riemannian metric inherited from the class of second-order invariant Sobolev metrics on the space of surfaces. In comparison to existing geometric deep learning frameworks, our approach's training process is notably straightforward and does not demand a substantial amount of training data. Moreover, as our results suggest, it leads to strong out-of-sample generalization properties when dealing with unseen data. We demonstrate the usability of our framework in a variety of different experiments (registration, interpolation, extrapolation, random shape generation and motion transfer) on two distinct types of data – human body meshes from the FAUST, DFAUST and SHREC repositories as well as face scans from the COMA dataset.

This work is based on the authors' previous conference publication (Hartman et al., 2023a), but introduces several new important additions to that initial approach. This includes in particular:

- a more general and in-depth presentation of the mathematical framework of basis restricted elastic shape analysis. Indeed, Hartman et al. (2023a) is entirely focused on the setting of human body shapes and assumed a specific splitting of the latent space into two subspaces associated to changes in body identity and pose changes respectively; here, we lift such restrictions and introduce our framework with any generic space of admissible deformations of a reference template, see Sec. 2 for more details. While it leads to optimization problems formally equivalent to those presented in Hartman et al. (2023a), this extended framework allows us to go beyond the sole case of human body shapes and investigate applications to other types of data.
- a more comprehensive description and justification of the computational model and proposed methodology, including several ablation studies to validate our choice of number of basis vector fields, shape matching functions and Riemannian metric on the latent space, see Sec. 8.
- an extended comparison with state-of-the-art latent space methods for body shape analysis (including FARM and 3D-coded), c.f. Sec. 6;
- experiments on an extra dataset of human bodies from SHREC, c.f. Sec. 6.
- an experiment on the shape and pose disentanglement properties of our framework both with and without the presence of labels in the training data, cf. Sec. 6.4;
- a new application of the method on different data highlighting the effectiveness of this approach for scans of human faces and human hands; c.f. Sec. 6.5 and Sec. 7.2.
- a new section on constructing the deformation bases in the absence of 4D training data, c.f. Sec. 7;
- an open source coding package for basis restricted ESA with precomputed bases for human body and face analysis, available at https://github.com/emmanuel-hartman/BaRe-ESA.

1.1 Related Work and Motivation

Analyzing three-dimensional (3D) *surfaces* has become an increasingly important topic, where the need for such algorithms is motivated by the emergence of high-accuracy 3D scanning devices, that have resulted in a significant increase in the availability of such data. The resulting application range from human health analysis (Desrosiers et al., 2017), facial animation (Qin et al., 2023; Otberdout et al., 2022a), computer graphics (Deng et al., 2022) or synthetic human data generation (Zhang et al., 2020) to computational anatomy (Grenander and Miller, 1998).

Although the framework developed in this article is fairly general and, we believe, could be relevant for a variety of real data applications, our simulations will primarily focus on datasets of 3D human bodies and faces. These involve particularly challenging problems given the high degree of variation in shape and pose, and the lack of point correspondences and consistent mesh structure across such datasets.

Geometric shape analysis: The general field of Riemannian shape analysis has produced several mathematical frameworks and numerical pipelines to tackle some of the key problems in the comparison and statistical analysis of 3D surfaces. These models are built from a Riemannian metric on a "shape space", in which the "shape" of a surface is usually regarded as what information remains after factoring out shape-preserving transformation groups such as reparametrizations or rigid motions. Two main frameworks have in particular stood out in constructing Riemannian metrics on such shape spaces: on one hand, the diffeomorphic approach of Beg et al. (2005); Younes (2019) and, on the other, the elastic metric setting introduced in (Younes, 1998; Srivastava and Klassen, 2016). An important aspect in both models is that the formulation of basic shape analysis tasks such as the estimation of the geodesic distance between two given surfaces for instance, is typically framed as the minimization of a *reparametrization invariant matching energy* in which computation of the distance and of the optimal registration (i.e. of the unknown point correspondences) must be tackled *jointly*. This should be viewed in sharp contrast with the majority of traditional approaches in shape analysis (Audette et al., 2000) in which registration is performed as a pre-processing step using methods such as functional maps (Ovsjanikov et al., 2012) and where the subsequent analysis is then done independently of this registration. This practice has been, however, increasingly questioned as it can, in some cases, lead to a severe loss of data structure/information or generate bias in the analysis, see e.g. Srivastava and Klassen (2016) and the references therein. On the other hand, the joint estimation of distance and registration often induces several practical challenges in particular when working with simplicial meshes such as triangulated surfaces. Some approaches (Kurtek et al., 2012; Jermyn et al., 2017; Su et al., 2020a; Tumpach et al., 2016; Laga et al., 2022) rely on analytical representations or approximations for the surfaces and the reparametrization group (using e.g. spherical harmonics) but are therefore often limited to a predefined topology. As an alternative, it was proposed, first for the diffeomorphic model in (Vaillant and Glaunes, 2005; Charon and Trouvé, 2013), and later adapted to the ESA framework (Bauer et al., 2021; Hartman et al., 2023b), to instead introduce discrepancy loss functions built from measure representations of surfaces. Those discrepancy functions, in particular the metrics derived from the framework of *varifolds*, have been shown to provide robustness to scan inconsistencies, such as varying mesh samplings and topological noise.

Despite those successes, one of the key remaining limitation of Riemannian shape analysis is the fact that pure geodesic trajectories are often not inherently representative of realistic longitudinal changes in real data. For instance, in one of the data application of this paper, it has been observed that simple geodesic interpolation between two human body poses does not generally reproduce the natural body motion that would be expected, c.f. Section VI of the supplementary material. An important current research challenge is thus to develop ways to enforce various types of physical, biological or data-specific constraints within Riemannian shape frameworks. In the diffeomorphic setting, some progress has been made towards that goal either through the introduction of sub-Riemannian (Arguillère et al., 2015; Gris et al., 2018) or other types of constrained models (Charlier et al., 2018; Hsieh et al., 2022; Charon and Younes, 2023). Yet these approaches are typically built around user specified constraints or principles rather than being entirely data-driven and are also known to be numerically costly when working with high resolution data. The basis restricted approach of the present work in part overcomes those difficulties by leveraging, on the one hand, the advantages of the elastic metric framework when it comes to numerical complexity and, on the other hand, by extracting from the dataset itself the adequate constrained subspace of deformations. In the registered setting a similar approach has been used in the conference paper (Pierson et al., 2022) by some of the authors and Tumpach. Such basis models are highly related to latent space models, popular in geometric deep learning (Bronstein et al., 2017, 2021), which we will describe next.

Low dimensional deep deformation models: Recently, deep deformation models have become increasingly popular for shape representation and deformation. These approaches propose to build a deformation model for different types of deformable shapes, such as the human body (Bouritsas et al., 2019; Lemeunier et al., 2022; Huang et al., 2021; Groueix et al., 2018a), the human face (Bouritsas et al., 2019; Otberdout et al., 2022b; Besnier et al., 2023), or animals (Huang et al., 2021) based on a limited training set of parameterized shapes.

However, those methods need to deal with parameterization invariance at inference. This is often done using a PointNet encoder (Qi et al., 2017; Besnier

et al., 2023), which maps a shape to its latent vector independently of its discretization. Other approaches have been proposed, but they come with an high training cost (Trappolini et al., 2021) or use intrinsic quantities such as the Laplacian (Sharp et al., 2022; Wiersma et al., 2022), that can be sensitive to topological changes. We note, however, that in practice, the invariance of those methods remains limited, because of their reliance on large datasets of parameterized surfaces for training purposes. They often need additional post-processing registration steps in inference to reproduce plausible geometric reconstruction of shapes (Huang et al., 2021; Groueix et al., 2018b; Trappolini et al., 2021).

Moreover, the performance of these methods is often limited in the context of large deformations: they regularly fail to sufficiently learn the non-linear map from the flat latent space to the shape space. Consequently, they are lacking generalizability when confronted with unseen data. To address these issues multiple deformation energy losses have been introduced in the training phase, such as geodesic distances (Cosmo et al., 2020), ARAP (Huang et al., 2021; Muralikrishnan et al., 2022), or volumetric constraints (Atzmon et al., 2021). Manifold regularization of learned pose spaces (Tiwari et al., 2022; Freifeld and Black, 2012) has also been proposed. Those geometric quantities however increase the total training costs of those approaches.

In contrast, our approach does not rely on a non-linear map but imposes an affine map, called the affine decoder, from a given low dimensional latent space to a corresponding space of shapes. This space is defined using pre-estimated basis. We impose non-linearity on the deformation space via the pullback of a second-order, parametrization-invariant, Sobolev (Riemannian) metric. The registration of a scan becomes an interpolation problem between the template and the scan representation in the low dimensional space, cf. Eq. (5), proposing plausible registrations of the shape. Moreover, interpolation and extrapolation problems are formulated as geodesic boundary and initial value problems and are easily implemented using modern scientific computation libraries.

2 Mathematical background

2.1 The Riemannian shape space of immersed surfaces

In this article, the "shapes" of interest are surfaces immersed in the Euclidean space $\mathbb{R}^3$. Mathematically, and from the continuous viewpoint, we define a *parametrized shape* as an immersion from a generic 2D parameter domain $\mathcal{T}$ (a compact 2-dimensional manifold) into $\mathbb{R}^3$, i.e. a smooth mapping $q : \mathcal{T} \to \mathbb{R}^3$ such that the Jacobian map $dq(u)$ is injective for all $u \in \mathcal{T}$. For instance, $\mathcal{T}$ can be taken as a compact domain of $\mathbb{R}^2$ if one considers open surfaces (such as human faces) or the sphere S^2 in the case of closed surfaces (such as whole human body surfaces). We shall denote by Imm the space of all immersions from $\mathcal{T}$ to $\mathbb{R}^3$.

In order to provide a quantitative way to compare such shapes, one needs to introduce a similarity measure on Imm. As pointed out in the introduction, we are here interested in similarity measures that originate from a Riemannian setting, in other words we wish to view Imm as an infinite dimensional manifold and equip it with a Riemannian metric. In this setup the corresponding geodesic distance function provides the similarity measure for shape comparison. In addition this will allow us to reduce tasks such as shape interpolation and extrapolation to geometric operations – the geodesic initial and boundary value problem. First, note that, as an open subset of $C^\infty(\mathcal{T}, \mathbb{R}^3)$, Imm can be directly viewed as a Fréchet manifold for which the tangent space $T_q \mathrm{Imm}$ at each immersion q can be naturally identified with $C^\infty(\mathcal{T}, \mathbb{R}^3)$ or equivalently as the space of smooth deformation fields along the parametrized surface q, cf. Figure 1 for an explanatory illustration of the shape space of parametrized immersed surfaces.

In this setting, a Riemannian metric G is a family of inner products $G_q : C^\infty(\mathcal{T}, \mathbb{R}^3) \times C^\infty(\mathcal{T}, \mathbb{R}^3) \to \mathbb{R}_+$ that depends smoothly on the foot point $q \in \mathrm{Imm}$. We further recall that, from G, one obtains a "distance" on Imm which, for any $q_0, q_1 \in \mathrm{Imm}$, is obtained as:

$$d_G(q_0, q_1)^2 = \inf \left\{ \int_0^1 G_{q(t)}(h(t)), h(t)) \right\} \quad (1)$$

where $h(t) = \partial_t q(t) \in T_{q(t)} \mathrm{Imm}$ and where the infimum is taken over all paths $q : [0, 1] \to \mathrm{Imm}$ with $q(0) = q_0$ and $q(1) = q_1$. We call (1) the

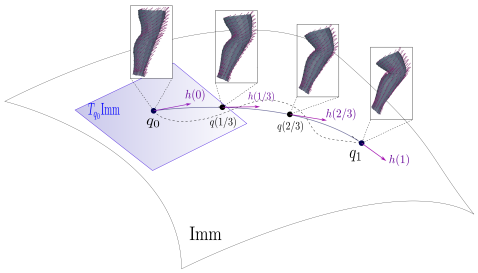

Fig. 1: Illustration of the Riemannian shape space of parametrized immersed surfaces.

parametrized matching problem. Any minimizing path, if it exists, is then a *geodesic* between q_0 and q_1.

To define the Riemannian metric G we will rely on the setting of elastic shape analysis (ESA) which has derived various families of metrics that further satisfy the key property of reparametrization-invariance. To explain this property we introduce the notion of a *reparametrization* ϕ as an element of the diffeomorphism group $\mathrm{Diff}(\mathcal{T})$, i.e., the set of smooth and bijective maps on the parameter domain . This group acts on any given immersion q by right composition, i.e., $(q, \phi) \mapsto q \circ \phi$. The metric G is called reparametrization invariant if for any $\phi \in \mathrm{Diff}(\mathcal{T})$ we have

$$G_{q\circ\phi}(h \circ \phi, k \circ \phi) = G_q(h, k) \tag{2}$$

and the importance of this property will become clear below in Section 2.2, where we will quotient out the action of this group.

Perhaps the simplest of those metrics is the so called invariant L^2 metric defined for all $q \in \mathrm{Imm}$ and $h, k \in C^\infty(\mathcal{T}, \mathbb{R}^3)$ via:

$$G_q(h, k) = \int_{\mathcal{T}} \langle h, k\rangle \,\mathrm{vol}_q$$

where vol_q is the volume measure on $\mathcal{T}$ induced by q, that is, denoting (u_1, u_2) some coordinates on $\mathcal{T}$, $\mathrm{vol}_q = |\partial_{u_1} q \wedge \partial_{u_2} q|$. The integration with respect to this induced volume measure is precisely what leads to the invariance of the metric (and by extension of the geodesic distance). One crucial shortcoming of the above metric however, which was first shown by Michor and Mumford in (Michor and Mumford, 2005; Bauer et al., 2012), is that the associated d_G turns out to be fully degenerate and a fortiori not a true distance, i.e., with respect to this metric all shapes are considered to be equal.

One way to address this issue is by introducing higher-order metrics on Imm. In this article, we shall focus on the class of *second order invariant Sobolev metrics* which has shown several desirable properties in past works (Bauer et al., 2011; Hartman et al., 2023b). More specifically we will consider the 6-parameters family of metrics obtained by the following combination of 0-th, 1-st and 2-nd order terms weighted by the nonnegative constants a_0, a_1, b_1, c_1, d_1 and a_2:

$$\begin{aligned} G_q(h, k) = \int_{\mathcal{T}} \Big(& a_0\langle h, k\rangle + a_1 g_q^{-1}(dh_m, dk_m) \\ & + b_1 g_q^{-1}(dh_+, dk_+) + c_1 g_q^{-1}(dh_\perp, dk_\perp) \\ & + d_1 g_q^{-1}(dh_0, dk_0) + a_2\langle \Delta_q h, \Delta_q k\rangle \Big) \,\mathrm{vol}_q\,. \end{aligned} \tag{3}$$

In the above, dh denotes the vector-valued 1-form on $\mathcal{T}$ given by the differential of h which we can alternatively view, in a given coordinate system, as a 3×2 matrix field on $\mathcal{T}$, g_q is the pullback of the Euclidean metric on $\mathbb{R}^3$ which we may view as a 2×2 symmetric positive definite matrix field on $\mathcal{T}$, in which case $g_q^{-1}(dh, dh) = \mathrm{tr}(dh g_q^{-1} dh^T)$. The second-order term involves the vector Laplacian Δ_q induced by the parametrization which in coordinates can be written as $\Delta_q h = \frac{1}{\sqrt{\det(g_q)}} \partial_{u_1} \left(\sqrt{\det(g_q)} g_q^{u_1 u_2} \partial_{u_2} h \right)$. Lastly, let us briefly comment on the particular splitting of the first-order part of the metric in the four different terms appearing in (3). For the sake concision, we shall refer the interested reader to the appendix (or to (Su et al., 2020b; Charon and Younes, 2023)) for the technical definition of the orthogonal decomposition of dh into the sum of the tensors $dh_m, dh_+, dh_\perp, dh_0$. We will only say at this point that such a splitting is motivated by its interpretation from linear elasticity theory, with the terms weighted by a_1, b_1, c_1 in (3) corresponding to thin shell shearing, stretching and bending energies induced by the deformation field h respectively: more specifically: the term involving dh_m measures the change in the local surface metric (this corresponds to shearing of the surface), whereas the term involving dh_+ measures the change of the area volume density which corresponds to local stretching of the surface. Finally, the term involving $dh_\perp$ measures the change of the surface normal vector, which can be associated to surface bending.

Consequently, the class of invariant H^2 metrics (3) provides a flexible family allowing, through the

selection of the weighting coefficients, to emphasize or penalize different types of deformations. Each of those metrics is reparametrization-invariant and unlike the L^2 case it induces a true distance on Imm:

Theorem 1. *Let $a_0 > 0$ and let either $a_1, b_1, c_1, d_1 > 0$ or $a_2 > 0$ then the induced geodesic distance of the metric G on the space Imm is non-degenerate, i.e., for any two surfaces $q_0, q_1 \in \mathrm{Imm}$ with $q_0 \neq q_1$ we have $d_G(q_0, q_1) > 0$.*

For a proof of this result we refer to the supplementary material. Furthermore, as we shall explain later, there are natural discretization schemes to compute such second-order metrics on e.g. triangulated meshes.

2.2 Quotienting out reparametrizations

Note that the model described so far leads to distances and geodesics between parametrized shapes. From a practical standpoint, this intrinsically assumes known point to point correspondences, namely point $q_0(u)$ on the source surface is matched to $q_1(u)$ on the target. Apart from pre-registered datasets (such as the DFAUST one described below), it is common in most applications that raw or segmented surface meshes do not come with such given correspondences, and even display inconsistent number of vertices and/or mesh structures. Thus one is typically interested in comparing surfaces **independently** of how they are parametrized/sampled.

Mathematically, this can be done by looking at the quotient shape space $\mathcal{S} = \mathrm{Imm} \,/\, \mathrm{Diff}(\mathcal{T})$ of the equivalence classes $[q] = \{q \circ \phi \ : \ \phi \in \mathrm{Diff}(\mathcal{T})\}$ of all possible reparametrizations of q. A key advantage of the invariant metric framework introduced in the previous section, and in particular of the invariant Sobolev metrics (3), is that one can recover a Riemannian distance on $\mathcal{S}$ as follows. Given unparametrized surfaces $[q_0]$ and $[q_1]$, the quotient distance is obtained by fixing a parametrization q_0 and solving the following **unparametrized matching problem**:

$$\overline{d}_G([q_0],[q_1])^2 = \inf_{(q(\cdot),\phi)} \left\{ \int_0^1 G_{q(t)}(\partial_t q(t), \partial_t q(t)) \right\} \quad (4)$$

where the minimization is now over paths $q : [0,1] \to \mathrm{Imm}$ **and** reparametrization $\phi \in \mathrm{Diff}(\mathcal{T})$, with the constraint that $q(0) = q_0$ and $q(1) = q_1 \circ \phi$ i.e. $[q(1)] = [q_1]$. In other words, the quotient distance is obtained by jointly finding an optimal path from q_0 to an optimal reparametrization of the target.

However, the variational problem (4) is generally challenging to tackle and to implement on discrete surface meshes. It involves estimating parametrizations of the two surfaces over a predefined domain (such as the sphere) and then requires discretizing and optimizing over the group $\mathrm{Diff}(\mathcal{T})$ (Jermyn et al., 2017; Su et al., 2020b). An alternative approach in registration problems is rather to enforce the matching constraint $[q(1)] = [q_1]$ indirectly via a discrepancy function $\Gamma([q(1)],[q_1])$ that only depends on the unparametrized shapes and therefore consider the **relaxed matching problem**:

$$\inf_{q(\cdot)} \left\{ \int_0^1 G_{q(t)}(\partial_t q(t), \partial_t q(t)) + \lambda \Gamma([q(1)],[q_1]) \right\} \quad (5)$$

in which $\lambda > 0$ acts as a Lagrange multiplier for the terminal constraint, and the minimization is now only over parametrized surface paths $t \mapsto q(t)$; in other words, we bypass the need for directly optimizing reparametrizations.

To define Γ, one typically introduces a measure of similarity between the geometric point sets $q_1(\mathcal{T})$ and $q(1)(\mathcal{T})$; we discuss a few possible options in the Supplementary Material, including the Hausdorff and Chamfer distances often used for that purpose in computer vision. In this paper, following many other works and our previous publication (Hartman et al., 2023b), we instead rely on similarity terms derived from geometric measure theory, specifically the family of kernel metrics on the space of *varifolds* (Kaltenmark et al., 2017). A notable advantage of this framework is that it leads to actual distances that can be differentiated with respect to the point positions of either shape. Although we will abstain from presenting this construction in the main text for concision, we refer the reader to the Appendix for details and qualitative comparison of varifold metrics with other classical point set discrepancies.

3 Restricted latent space model

As highlighted in the introduction, one limitation of the general H^2 metric framework is that it does not impose any restriction on deformation fields beyond the energy penalties in the metric (3). When it comes to modelling human body motion for example, it has been observed that geodesics between two poses most often do not emulate a "natural" interpolation of the

pose, despite the flexibility in the choice of metric coefficients. A second practical downside is the numerical complexity of having to solve a very high dimensional optimization problem over paths of surfaces for any estimation of a distance and geodesic, which can become quite significant when generalizing that approach for more complex statistical tasks such as Fréchet mean estimation or parallel transport.

3.1 Latent space representation

As a way to address the above challenges, we propose a simplified and **linearized finite dimensional** shape space model by restricting ourselves to parametrized surfaces q that result from a fixed template surface $\bar{q} \in \mathrm{Imm}$ and a predefined admissible set of P linearly independent deformation fields $\{h_i\}_{i=1}^P$ of the template. More precisely, with the affine mapping $F : \mathbb{R}^P \to C^\infty(\mathcal{T}, \mathbb{R}^3)$ defined by:

$$F : (\alpha_i)_{i=1,\dots,P} \mapsto \bar{q} + \sum_{i=1}^{P} \alpha_i h_i, \tag{6}$$

we introduce the space $\mathcal{L}_{\bar{q}} = F(\mathbb{R}^P) \bigcap \mathrm{Imm}$.

Remark 1 (Relation to Linear Blend Skinning formulations)**.** *The formulation resembles the Linear Blend Skinning (LBS) formulation present in common models of the human body (Loper et al., 2015; Anguelov et al., 2005): the shape is represented as a template deformation, which is a sum of body pose and identity deformations. The main difference is that standard LBS formulations the pose deformation is done using a precomputed skeleton which is linearly rigged to the template mesh, inducing non-linear pose deformation as a combination of skeletal articulation rotations. In our model, the non-linearity directly comes from the Riemannian metric, and no skeleton is needed.*

Consequently any surface $q \in \mathcal{L}_{\bar{q}}$ can be then represented uniquely by a finite-dimensional vector $\alpha = (\alpha_i) \in \mathbb{R}^P$ we will call the *latent code* of q, thus allowing us to work on a potentially much lower-dimensional space. Yet $\mathcal{L}_{\bar{q}}$ should still remain rich enough so as to express the predominant geometric variations in the dataset of interest. As we shall address in 5.2, this suggests using a basis $\{h_i\}$ that is built in a *data-driven way*. Furthermore, this latent space model allows the use of composite bases, where different subsets of vector fields are associated to distinct types of morphological variations. This will prove particularly relevant to the applications of this paper when we are interested in e.g. disentangling body pose from body type changes or facial expression from facial morphology changes.

Remark 2. *Note that, in general, $\mathcal{L}_{\bar{q}}$ is an open subset of the affine space $F(\mathbb{R}^P)$ and contains $\bar{q}$. However, not all elements of $F(\mathbb{R}^P)$ are immersions unless certain specific conditions on the vector fields are satisfied. This holds in particular if for all $i = 1, \dots, N$ and $u \in \mathcal{T}$, $dh_i(u)^T d\bar{q}(u) = 0$. However, we will not assume this condition in the rest of the paper.*

3.2 Induced Riemannian metric

The next logical question to address is which metric to take on the above latent space. In sharp contrast to most encoder models in geometric deep learning which often implicitly consider the standard Euclidean structure, our approach is rather to take advantage of the properties of invariant metrics on shape spaces and pull the metric back to the latent space. Namely, for any $\alpha \in F^{-1}(\mathcal{L}_{\bar{q}})$ and $\beta, \eta \in \mathbb{R}^P$, let us define: $\overline{G}_\alpha(\beta, \eta) := G_{F(\alpha)}(d_\alpha F(\beta), d_\alpha F(\eta))$ in which G is a Riemannian metric on Imm which we shall take from the invariant family (3). As the mapping F is affine, this pull back metric on $\mathbb{R}^P$ can be expressed more explicitly as:

$$\overline{G}_\alpha(\beta, \eta) = G_{F(\alpha)}\left(\sum_{i=1}^{P} \beta_i h_i, \sum_{j=1}^{P} \eta_j h_j\right) = \beta^T \overline{G}_\alpha \eta$$

where, in the last equation, $\overline{G}_\alpha = [G_{F(\alpha)}(h_i, h_j)]$ is the symmetric positive definite $P \times P$ matrix giving the metric at the latent code α. Estimation of the distance between any two surfaces $q_0 = F(\alpha_0)$ and $q_1 = F(\alpha_1)$ then reduces to standard finite-dimensional Riemannian geometry and is obtained by finding a path of coefficients $t \mapsto \alpha(t) \in \mathbb{R}^P$ minimizing $E(\alpha) = \int_0^1 (\partial_t \alpha)^T \overline{G}_\alpha (\partial_t \alpha) dt$ with $\alpha(0) = \alpha_0$ and $\alpha(1) = \alpha_1$.

4 Shape analysis in latent space

Relying on the latent space representation and its Riemmanian metric introduced in the previous section, one can perform efficiently a variety of shape analysis related tasks, which we describe in the following paragraphs.

4.1 Calculating latent space representations

We start by describing how we can calculate a latent space representation that is (up to numerical accuracy) independent of the parametrization of the surface, i.e., given a surface $q \in \mathrm{Imm}$ we aim to find a latent code representation α such that $F(\alpha) = q \circ \phi$ for some (unknown) reparametrization function $\phi \in \mathrm{Diff}(\mathcal{T})$. To tackle this problem we rely again on the varifold similarity term, i.e., we reformulate the latent representation problem as the task of finding a latent code representation α such that

$$\Gamma(F(\alpha), q) = 0. \tag{7}$$

One remaining difficulty is that, for most datasets such as those of Section 5.1, raw surface scans are not given with consistent mesh structures and a fortiori cannot be assumed to all belong to $\mathcal{L}_{\bar{q}}$ for a given fixed template $\bar{q}$. To circumvent this difficulty we consider a relaxed formulation of the latent code representation problem; instead of searching for a latent code α satisfying equation (7) we simply aim to minimize the varifold distance $\Gamma(F(\alpha), q)$ over all latent codes $\alpha \in \mathbb{R}^P$. In our experiments it turned out to be beneficial to add an extra regularizing term to this minimization problem, which we choose to be the geodesic distance of $F(\alpha)$ to the template $\bar{q}$, i.e., we minimize the energy

$$\left(\Gamma(F(\alpha), q) + \frac{1}{\lambda} d_G^{\mathcal{L}_{\bar{q}}}(\bar{q}, F(\alpha))^2\right) \tag{8}$$

over all $\alpha \in \mathbb{R}^P$, where $\lambda > 0$ is a weight parameter. Using the definition of the geodesic distance $d_G^{\mathcal{L}_{\bar{q}}}$ on the latent space $\mathcal{L}_{\bar{q}}$ this requires us to minimize the path energy

$$\Gamma(F(\alpha(1)), q) + \frac{1}{\lambda} \int_0^1 \overline{G}_\alpha(\partial_t \alpha(t), \partial_t \alpha(t)) dt \tag{9}$$

over all paths $\alpha : [0,1] \to \mathbb{R}^P$. Numerically, we consider time-discrete paths of coefficients $\alpha = (\alpha(0), \alpha(1/T), \alpha(2/T), \ldots, \alpha(1))$ for a selected number of time steps T, with $\partial_t \alpha$ being approximated by forward finite difference. Furthermore, q and $\bar{q}$ are in practice given as sets of vertices and triangular meshes while each h_i is of a collection of vectors sampled on the vertices of $\bar{q}$. This turns the problem into an unconstrained minimization over $\mathbb{R}^{P(T-1)}$ for which we use the L-BFGS algorithm of the *scipy* library; here the free variables are only in $\mathbb{R}^{P(T-1)}$ as the path starts at $\bar{q}$ and thus $\alpha(0) = 0$. The precise discretization of the different terms in (11), based on the principles of discrete differential geometry, is detailed in the Supplementary Material. Our implementation, that builds on some of the authors' previous package for surface matching[1], is done in Python and relies on libraries such as *PyTorch* and *PyKeops* which allow to automatically differentiate those terms on the GPU. Our implementation is also publicly available on Github[2] and relies on the same libraries.

4.2 Shape comparison and interpolation

Quantifying the global difference between surfaces is generally essential when attempting for example to cluster data in a population. The Riemannian metric setting gives a direct way to measure such differences via the distance itself and, what is more, lead to geodesic paths that interpolate between the objects. The availability of such geodesic paths has the double advantage of allowing to interpret the properties and behaviour of the distance while also providing a way to reconstruct a dynamical evolution from one data point to another.

Within the framework of Section 3, we have seen that the estimation of distance and geodesics between two surfaces $q_0 = F(\alpha_0)$ and $q_1 = F(\alpha_1)$ in $\mathcal{L}$ can be done by finding a path of least Riemannian energy in the latent space, i.e., by minimizing the path energy

$$\int_0^1 \overline{G}_\alpha(\partial_t \alpha(t), \partial_t \alpha(t)) dt \tag{10}$$

over all paths $\alpha : [0,1] \to \mathbb{R}^P$ such that $\alpha(0) = \alpha_0$ and $\alpha(1) = \alpha_1$. Discretizing the path in time t this leads to an unconstrained minimization problem over $\mathbb{R}^{P(T-2)}$ with the free variables being $\alpha(1/T), \alpha(2/T), \ldots, \alpha((T-1)/T))$ as $\alpha(0) = \alpha_0$ and $\alpha(1) = \alpha_1$ are fixed.

Given new data, for which we have not yet calculated a latent space representation, we could proceed as follows: calculate first a latent space representation using the method of the previous section and then solve the geodesic problem using the above algorithm. In practice it is, however, more effective to solve both of these tasks in one step. This can be done using again the varifold distance and by considering the path

[1] https://github.com/emmanuel-hartman/H2_SurfaceMatch
[2] https://github.com/emmanuel-hartman/BaRe-ESA

minimization problem

$$\int_0^1 \overline{G}_\alpha(\partial_t\alpha, \partial_t\alpha)dt + \lambda\Gamma(F(\alpha(0)), q_0) + \lambda\Gamma(F(\alpha(1)), q_1). \quad (11)$$

where $\alpha : [0,1] \to \mathbb{R}^P$ is again a path in the latent space $\mathcal{L}_{\bar{q}}$. The presence of the two discrepancy terms in (11) is necessary to make the above problem well-defined for **any** q_0 and q_1 in Imm and not just in $\mathcal{L}_{\bar{q}}$. The solution (11) can be thus interpreted as the distance and geodesic between the closest approximations of q_0 and q_1 by elements of the latent space.

4.3 Shape extrapolation

The shape extrapolation problem consists in predicting the future evolution of a surface given an initial deformation direction. In our Riemannian framework this reduces to solving the geodesic equation with given initial condition $q(0) = q_0$ (the initial pose) and $\partial_t q(0) = h$ (the deformation direction), cf. Figure 1. The geodesic equation is the first order optimality condition of the energy functional; it is a non-linear PDE, that is second order in time t and forth order in space (twice the order of the metric). For the exact formula of this equation, which is rather lengthy and not particularly insightful, we refer the interested reader to the literature, see eg. Bauer et al. (2011). To solve such initial value problems in our latent space, we modify methods of discrete geodesic calculus (Rumpf and Wirth, 2013) to our setting. We approximate the geodesic starting at α^0 in the direction of β with a PL path with $N+1$ evenly spaced breakpoints. At the first step, we set $\alpha^1 = \alpha^0 + \frac{1}{N}\beta$ and find α^2 such that $F(\alpha^1)$ is the geodesic midpoint of $F(\alpha^0)$ and $F(\alpha^2)$, i.e., we solve for α^2 such that

$$\alpha^1 = \underset{\tilde{\alpha}}{\mathrm{argmin}}[\overline{G}_{\alpha^0}(\beta_0, \beta_0) + \overline{G}_{\tilde{\alpha}}(\tilde{\beta}, \tilde{\beta})]$$

where $\beta_0 = \tilde{\alpha} - \alpha^0$ and $\tilde{\beta} = \alpha^2 - \tilde{\alpha}$. Differentiating with respect to $\tilde{\alpha}$ and evaluating the resulting expression at α^1, we obtain the system of equations

$$\begin{aligned} 2\overline{G}_{\alpha^0}(\beta_0, h_i) - 2\overline{G}_{\alpha^1}(\tilde{\beta}, h_i) + D_{\alpha^1}\overline{G}.(\tilde{\beta}, \tilde{\beta})_i = 0, \\ 2\overline{G}_{\alpha^0}(\beta_0, k_i) - 2\overline{G}_{\alpha^1}(\tilde{\beta}, k_i) + D_{\alpha^1}\overline{G}.(\tilde{\beta}, \tilde{\beta})_{i+m} = 0 \end{aligned} \quad (12)$$

where $\{h_i, k_i\}$ is our basis of deformations as introduced above. We denote the system of equations in (12) by $\Phi(\alpha^2; \alpha^1, \alpha^0) = 0$, where we stress again that α^0 and α^1 are here fixed and known. We solve this system of equations for α^2 using a nonlinear least squares approach, i.e., by computing

$$\alpha^2 = \underset{\tilde{\alpha}}{\mathrm{argmin}} \|\Phi(\tilde{\alpha}; \alpha^1, \alpha^0)\|_2^2.$$

We repeat this process $N-1$ times, thereby constructing the discrete solution up to time $t = 1$.

4.4 Motion transfer in latent space

As previously discussed, composite bases offer a means to independently depict various modes of shape deformation. Specifically, when applied to human body and facial morphology, these bases allow us to separate identity and pose alterations, enabling motion transfer. In practical terms, when presented with a series of unregistered scans depicting a single identity engaged in an action, we can obtain latent code representations for each frame of the action. We then substitute the coefficients of the shape basis with the shape coefficients of the desired identity. This process yields a sequence of shapes that faithfully embodies the desired motion transferred onto the desired identity. Note that this is significantly simpler (albeit different) than performing parallel transport in the Riemannian manifold of surfaces as done in e.g. Hartman et al. (2023b).

4.5 Random shape generation

Additionally, we can utilize the Riemannian structure on our latent space representation to offer a data-driven method for generating random shapes from unregistered data. We may do this by learning an empirical distribution on the tangent space of the template shape. Given a data set of unregistered shapes, we solve the latent code retrieval problem and compute the initial vectors of the resulting geodesics in the latent space. We can then fit Gaussian mixture model on the resulting collection of tangent vectors and solve the initial value problem from the template in the direction of the vector generated from this model. In the case where we compute multiple bases to describe different modalities of shape change, the model may be fit to independently generate different types of shape changes.

5 Experimental Methodology

In this section we will describe the different datasets, which we will use in the experimental section, as well as the corresponding basis construction and the choice of parameters. In addition we will present different ablation studies, that further motivate the chosen energy functional.

5.1 Used Datasets

Human Body dataset: The main type of data considered in this article consists of human body scans. To construct our basis we will make use of the publicly available Dynamic FAUST (DFAUST) (Bogo et al., 2017) dataset. This dataset contains high quality scans, along with corresponding registered meshes that will be used as training data. More specifically DFAUST (Bogo et al., 2017) is comprised of 4D scans captured at 60 Hz of 10 individuals performing 14 in-place motions. Due to the high speed of the recording, DFAUST scans contains several singularities in the surface, such as holes or even artificial objects (eg. parts of walls). The corresponding registered surfaces to each scan are created using image texture information and a novel body motion model. A set of 7 long range sequence are left for testing. The remaining 133 sequences, which we denote DFaustT, make up the training set from which we compute the deformation and motion basis.

For the quantitative experiments, we consider three testing datasets on which we validate our model trained on DFaustT; first, we consider a subset of the static FAUST dataset (Bogo et al., 2014) for testing our models performance for registration and point correspondences. The static FAUST database is a 3D static scan dataset designed for human mesh registration tasks, that contains scans of minimally clothed humans and corresponding registered meshes. We selected scans of 10 individuals in 9 different poses from the training set that show no rotations along with the corresponding ground truth registrations and use them as our first testing set, denoted FaustE. In addition, we consider a subset of the SHREC dataset (Marin et al., 2020) to demonstrate the generalizability of our model in shape reconstruction tasks as it contains human shapes from significantly different modalities than that of our training set including scans of clothed humans and synthetic shapes of human bodies. For our third and final testing set, we divide the 7 sequences from DFAUST left aside for testing into 10 representative mini-sequences which we use to evaluate our framework's ability to reconstruct human motions. We denote this DFaustE.

Face scan dataset: As a second validation dataset, we consider human face scans from the COMA (Ranjan et al., 2018) database. It contains high-quality scans of human faces, along with corresponding registered meshes in the FLAME topology (Li et al., 2017) that will be used as training data. More specifically COMA is comprised of 4D scans of human faces captured at 60 Hz of 12 individuals performing 12 extreme facial expressions. The scans are available as raw scans of the whole face and often contain significant parts of the chest that are not present in the final registrations. Moreover, some detailed parts can be cropped or disappear in the scans, e.g. ears of the individual. The corresponding registered surfaces to each scan are created using image texture information, face landmarks and the FLAME model. A set of 12 sequences are left for testing and the remaining 132 sequences were used to compute the deformation and motion basis.

Hand dataset: As a third type of data we also consider human hand scans from the MANO database (Romero et al., 2017); a dataset comprising more than 800 registered hands with various poses from 50 individuals. These individuals were asked to reproduce daily life poses that were then scanned. The training set comprises approximately 800 registered hand poses and the MANO database provides a separate testing set consisting of 50 scans with available ground truth registrations.

5.2 Constructing the space $\mathcal{L}_{\bar{q}}$

To construct the deformation bases for motion and identity changes, we interpret registered mesh sequences of motions (expressions, resp.) as paths in shape space whose tangent vectors are implicitly restricted to the space of valid motions. We first collect meshes of the same pose (expression) from each identity and compute the (unrestricted) pairwise geodesics between these meshes with respect to our second-order Sobolev metric, where we use the Pytorch implementation of Hartman et al. (2023b). Note that these meshes show only moderate deformations and thus there are no difficulties with applying the unrestricted matching algorithm. We then collect the tangent vectors to these paths and perform PCA to define our basis for shape/identity deformations.

It would be possible to adopt a similar strategy for generating the body pose (or face expression)

deformation basis, i.e., collect shapes with the same body type (face identity, resp.) and calculate the unrestricted pairwise geodesics between these meshes. However, we had previously noticed Hartman et al. (2023a) that this may sometimes lead to unnatural motions for large movements. Instead, we shall first take advantage of the available 4D data in our targeted application datasets, allowing us to perform principle component analysis directly on the tangent vectors of those real motion sequences to obtain a valid pose (expression) data basis. This will be the approach used in the experiments of Section 6. Yet, in order to also validate our approach in the absence of 4D data, we present, in section 7.1, results obtained by following the same procedure as for the identity basis i.e. based only on 3D data for the basis construction. In the final experiments involving the MANO dataset (Section 7.2), we use an even simpler strategy to construct the bases: namely we simply consider linear deformations between all shapes in the training data as set of vectors for our PCA construction. As we will demonstrate, for this application, this cheaper procedure already produces satisfactory results outperforming the benchmark methods.

We should also note that we here pre-construct all bases from a fixed predefined training set. Another possible approach, used for instance in Muralikrishnan et al. (2023) (albeit only for shape deformations), is to progressively enrich some initial estimation of a basis via a bootstrapping scheme, providing a possible alternative way to build shape/pose deformation bases from only a small training set of registered meshes.

5.3 Parameter selection

Next we describe the choice of parameters in our experiments. For the human bodies the coefficients for the H^2-metric were chosen to enforce close to isometric deformations that allow for some stretching and shearing to allow change in body type. In the case of human body faces, we reduce the strecthing and shearing penalization, and enforce normal consistency. We added a small coefficient to the remaining terms to further regularize the deformations. The final six parameters for the H^2-metric are set to $(1, 1000, 100, 1, 1, 1)$ for human bodies and and $(1, 10, 10, 10, 1, 1)$ for human faces and hands. The basis size for all three applications is as follows (the number of basis vectors was chosen experimentally, cf. Section 8): the motion basis has $n = 130$ elements (70 elements for hands), whereas the basis for the body type variation has only $m = 40$ elements. Furthermore, we perform sequential minimizations where the parameter σ of the varifold term is decreased from 0.4 to 0.025 and the balancing term λ is increased from 10^2 to 10^8. In the applications to human faces and to human hands, we needed only two minimizations with the parameter σ of the varifold term at 0.01 and 0.005 and the balancing term λ at 10^6 and 10^{10}.

5.4 Evaluation methods

In our experiments, we will evaluate results quality using different similarity measures (distances) between the outputs of the different methods and the original scan. The "shape" matching is evaluated by comparing each method against the original scans using three different remeshing invariant similarity measures. First, we evaluate the methods using the varifold metric introduced before. As our method minimizes this distance during the registration process, we include two additional metrics to avoid bias: the widely used Hausdorff distance, which provides a good insight for the quality of a mesh reconstruction, but can be sensitive to single outliers present in low-quality scans and the Chamfer distance (Fan et al., 2017; Groueix et al., 2018b), which is more robust to such outliers.

In the first set of experiments – latent code retrieval, Section 6.1 – we will in addition evaluate the quality of the obtained point correspondences – in this section, we use data with given ground truth point correspondences. Therefore we will compute the mean squared error of each method to the ground truth registrations of the testing set. Unfortunately, one method (LIMP) does not return the same mesh structure as the ground truth registrations and thus we could not compare it this way. We thus add the geodesic error metric, to evaluate the matching quality. From the registered mesh, we extract point-to-point correspondences between the template and the given scan. Then for each point of the scan, we compute the geodesic distance (on the template mesh) between the proposed corresponding point and the ground truth correspondence. The final computed metric is the mean of these errors. For a detailed description of all these evaluation metrics, we refer to the supplementary material.

5.5 Comparison methods

Finally, we will briefly describe the other state-of-the-art methods that we considered for comparison. A more detailed description of these methods can be

found in the supplementary material. We primarily compare to methods that rely on latent space learning for registration, interpolation, and extrapolation tasks and do not consider other methods that can potentially tackle the same tasks but without a low dimensional latent space (Eisenberger et al., 2021), or that are specifically designed for other tasks (Muralikrishnan et al., 2022). We compare our approach to LIMP (Cosmo et al., 2020), which models shape deformations using a variational auto-encoder with geodesic constraints; ARAPReg (Huang et al., 2021), which models deformations using an auto-decoder with regularization through the as rigid as possible energy; and 3D-Coded (Groueix et al., 2018a), which is similar to LIMP but with lighter training and without geometric loss regularization. LIMP and 3D-Coded both utilize a PointNet architecture as an encoder, which enables invariance to parameterization. On the other hand, ARAPReg recovers latent vectors within a registered setting utilizing the L^2 metric, which assumes that the target meshes possess an identical mesh structure as the model's output. To make this framework viable for our application we replace the L^2-metric by the varifold distance thereby extending ARAPReg to unregistered point clouds. We trained all three networks on the DFAUST dataset using reported training details from the respective papers. As a final comparison, we consider the FARM approach (Melzi et al., 2019) from the class of functional maps-based methods. As FARM does not compute any interpolation or extrapolation of shape changes, we will however exclusively compare to this method for shape registration tasks.

6 Experimental Results using $4D$-training data

In this section, we will demonstrate the capabilities of our framework in several different experiments. For human body scans, which will be our main targeted application, we will present a thorough comparison to several other state-of-the-art algorithms. Therefore we will provide quantitative and qualitative analysis of the registration and point correspondence accuracy, the shape reconstruction quality, and the accuracy of interpolations and extrapolations to recreate real sequences of human motions. Furthermore, we give qualitative examples of our framework applied to random shape generation and motion transfer tasks. At the end of the section, we will present similar experiments for the COMA dataset, which consists of human face scans. The computational cost of our method is discussed in the supplementary material.

	LIMP	ARAPReg	3D-Coded	FARM	BaRe-ESA
MSE	NA	0.035	0.053	0.043	**0.014**
Geodesic Error	0.15	0.031	0.038	0.038	**0.013**

Table 1: Human body shape registration results. We compute the registration error on the FaustE data set. Where applicable, we compute the mean squared error (MSE) and geodesic error between each method's outputs and the ground truth registrations of FaustE.

	Hausdorff		Chamfer		Varifold	
	FAUST	SHREC	FAUST	SHREC	FAUST	SHREC
LIMP	0.23	0.17	0.098	0.070	0.073	0.057
ARAPReg	0.11	0.11	0.117	0.028	0.021	0.036
3D-Coded	**0.07**	**0.07**	0.020	**0.022**	0.023	**0.034**
BaRe-ESA	0.08	0.13	**0.019**	0.029	**0.014**	**0.034**

Table 2: Human body shape reconstruction results. We compute the Hausdorff, Chamfer, and Varifold reconstruction errors between the outputs of the methods and the original scans. We evaluate these methods on the FaustE and ShrecE testing sets.

6.1 Mesh invariant latent code retrieval

To demonstrate the effectiveness of our latent code retrieval algorithm, cf. Section 4.1, we tested its performance on the three human body testing data sets described in Section 5.1. In this experiment, we construct latent code representations with BaRe-ESA, LIMP, 3D-Coded, ARAPreg, and FARM and measure the distance from the reconstructed meshes to the original scans using the evaluation methods outlined in Section 5.4. In Fig. 2 we present a qualitative comparison of the obtained results. A quantitative comparison of the performance of the different methods is presented, with shape registration evaluation in Tab. 1 and geometric reconstruction of the human shape in Tab. 2. Both evaluations demonstrate that BaRe-ESA significantly outperforms the mesh autoencoder methods with respect to the registration and reconstruction evaluation metrics, with a performance quite similar to 3D-Coded in terms of reconstruction quality.

6.2 Interpolation and Extrapolation Results

We now turn our attention to the interpolation problem for human bodies, i.e., the task of constructing a deformation between two different human body poses, that

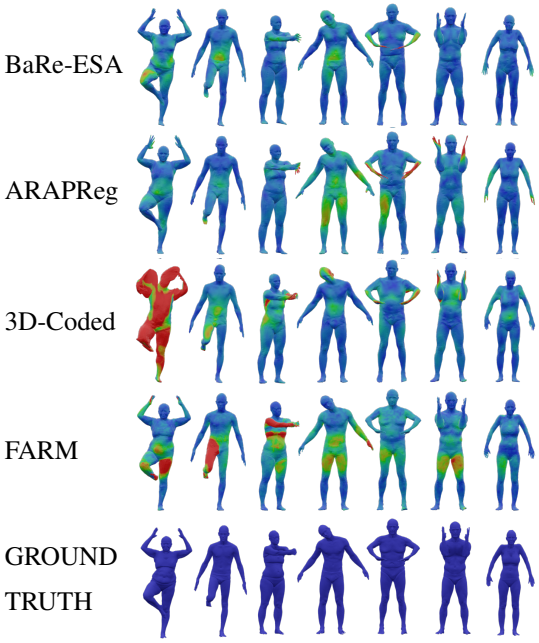

Fig. 2: Registration of seven elements of FAUST using four methods trained on DFAUST. The registrations produced by 3D-Coded, FARM, and ARAPReg have regions with large deformation errors. BaRe-ESA consistently produces a decent representation in all examples. The coloring of each mesh encodes the pointwise registration error from the ground truth with blue encoding 0mm error and red encoding $\geq$ 15mm error.

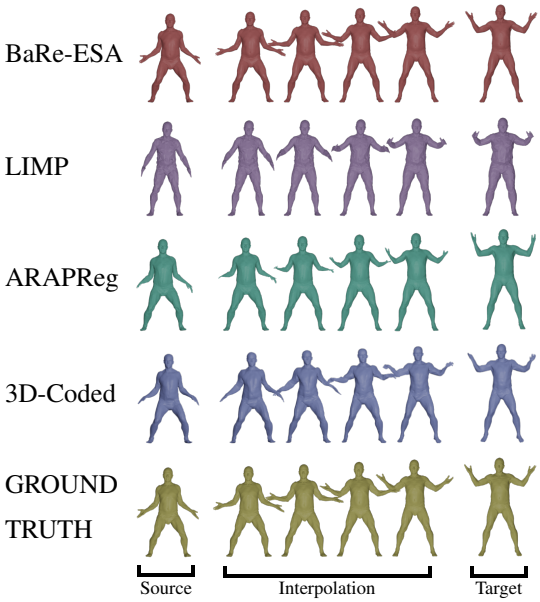

Fig. 3: Interpolation results comparison between our method, LIMP, ARAPReg and the Ground Truth from DFAUST. While the path produced by LIMP does not properly register the endpoints and the path produced by ARAPreg does not stay in the space of human bodies, BaRe-ESA successfully produces a path of human shapes whose endpoints match the source and target shapes.

follows a "realistic" motion pattern. We use the start and end points of our 10 test mini-sequences from the DFAUST data set as the input for these experiments. This allows us to compare the obtained results to the full mini-sequences, seen as a ground truth motion (see the supplementary material for their corresponding animations). In Fig. 3, we show a qualitative comparison of our method with ARAPReg, 3D-Coded, and LIMP. Our method is successful at recovering the latent codes that represent the endpoints and producing interpolations that remain in the space of human shapes. We further perform a quantitative comparison of the methods by measuring the distance to the ground-truth sequences at each break point with respect to the evaluation metrics introduced in Section 5.4; these results are displayed in Tab. 3. One can clearly observe that our method again outperforms the others both qualitatively and quantitatively.

Next, we consider the related problem of human body shape extrapolation, i.e., the task of predicting the future movement given a body shape and an initial movement (deformation). We consider again the 10 mini-sequences from the DFAUST dataset. We then recover the latent codes of the first two meshes in the sequence and use the first latent code and the difference of the codes as input to the method described in Section 4.3. In Fig. 4, we present again a qualitative comparison of our results to the extrapolations computed using LIMP, 3D-Coded, and ARAPreg (see the supplementary material for their corresponding animations). One can see that our method is successful at producing extrapolations that capture the correct motion of the mesh without any extraneous movements and without leaving the space of human bodies. As with the interpolation comparison, we measure the distance to the ground-truth sequences at each breakpoint and display the results of the quantitative

	Interpolation											
	Hausdorff				Chamfer				Varifold			
	LIMP	ARAPReg	3D-Coded	BaRe-ESA	LIMP	ARAPReg	3D-Coded	BaRe-ESA	LIMP	ARAPReg	3D-Coded	BaRe-ESA
punching	4.650	4.786	4.882	**1.009**	1.488	1.553	1.694	**0.350**	1.373	0.869	1.182	**0.252**
running on spot	2.045	0.977	1.357	**0.820**	1.026	**0.334**	0.454	0.475	0.786	**0.359**	0.441	0.372
running on spot b	2.367	1.726	1.931	**1.134**	1.039	0.653	0.706	**0.548**	0.767	0.488	0.545	**0.366**
shake arms	1.698	1.145	1.456	**0.847**	0.764	0.327	0.496	**0.326**	0.672	0.206	0.391	**0.180**
chicken wings	4.774	4.926	4.951	**1.289**	2.058	2.356	2.535	**0.636**	1.276	0.666	0.807	**0.296**
knees	12.898	2.797	19.593	**0.718**	8.803	0.496	18.153	**0.461**	2.067	0.627	1.925	**0.338**
knees b	5.516	1.055	**0.738**	1.995	1.862	0.262	**0.249**	0.693	0.748	0.298	**0.279**	0.347
jumping jacks	1.397	1.320	1.164	**0.811**	0.762	0.350	0.380	**0.333**	0.769	0.253	0.329	**0.229**
jumping jacks b	3.518	2.140	2.607	**1.482**	1.635	**0.672**	1.005	0.692	0.882	0.369	0.523	**0.254**
one leg jump	1.931	0.748	0.853	**0.616**	0.806	0.274	0.281	**0.221**	0.739	0.329	0.367	**0.264**
mean	4.079	2.162	3.953	**1.072**	2.024	0.728	2.595	**0.474**	1.008	0.447	0.679	**0.290**

Table 3: Full interpolation comparison on 10 DFAUST sequences. The Hausdorff, Chamfer and varifold distance are computed against ground truth sequences.

	Extrapolation											
	Hausdorff				Chamfer				Varifold			
	LIMP	ARAPReg	3D-Coded	BaRe-ESA	LIMP	ARAPReg	3D-Coded	BaRe-ESA	LIMP	ARAPReg	3D-Coded	BaRe-ESA
punching	4.232	8.142	5.792	**4.952**	1.494	2.436	2.685	**1.424**	1.506	1.551	1.441	**0.901**
running on spot	2.846	3.437	2.340	**1.973**	1.184	1.617	1.095	**1.071**	0.805	1.135	**0.607**	0.788
running on spot b	2.404	2.435	1.699	**1.392**	1.122	0.828	**0.759**	1.073	0.787	0.749	**0.515**	0.839
shake arms	2.090	2.737	1.734	**1.109**	1.017	0.892	0.630	**0.421**	0.771	0.528	0.520	**0.330**
chicken wings	4.778	12.790	5.224	**4.952**	2.230	5.127	2.536	**2.373**	1.475	1.673	**1.117**	1.121
knees	42.529	6.713	49.820	**3.632**	32.943	**1.144**	39.805	2.074	6.794	1.470	2.699	**1.428**
knees b	9.993	2.418	**1.942**	3.455	3.343	**0.554**	1.050	1.323	1.380	0.633	**0.506**	0.722
jumping jacks	4.116	5.873	8.696	**2.149**	1.767	2.345	6.449	**0.917**	1.099	1.038	0.699	**0.476**
jumping jacks b	2.219	3.519	1.759	**1.436**	0.992	0.984	0.702	**0.411**	0.765	0.623	0.498	**0.270**
one leg jump	2.195	1.970	1.989	**0.867**	0.906	0.757	0.915	**0.427**	0.758	0.858	1.800	**0.540**
mean	7.740	5.004	8.100	**2.592**	4.700	1.668	5.663	**1.151**	1.614	1.026	1.040	**0.742**

Table 4: Full extrapolation comparison on 10 DFAUST sequences. The Hausdorff, Chamfer and varifold distance are computed against ground truth sequences.

comparison in Tab. 4. Similar to the previous experiments, our method significantly outperforms LIMP, ARAPReg and 3D-Coded.

6.3 Motion Transfer and Random Shape Generation

As two further examples of the capabilities of the proposed framework, we present applications to motion transfer and random shape generation. To perform motion transfer, we first represent a motion as a sequence of latent codes and then simply replace the shape coefficients of each element of the sequence with the shape coefficients of the target shape. An example of this method in action is displayed in Fig. 5. Another possible application of our framework is random shape generation. The idea is to use a data-driven distribution on the human shape tangent space. Therefore we first perform latent code retrieval on a subset of DFAUST. We then compute the initial tangent vector of each of these paths in the latent space, separated in pose and shape components. For each of these collections of tangent vectors, we fit a Gaussian mixture model, which is popular for generating human shapes (Bogo et al., 2016; Omran et al., 2018). We used 10 and 6 components respectively, which proved

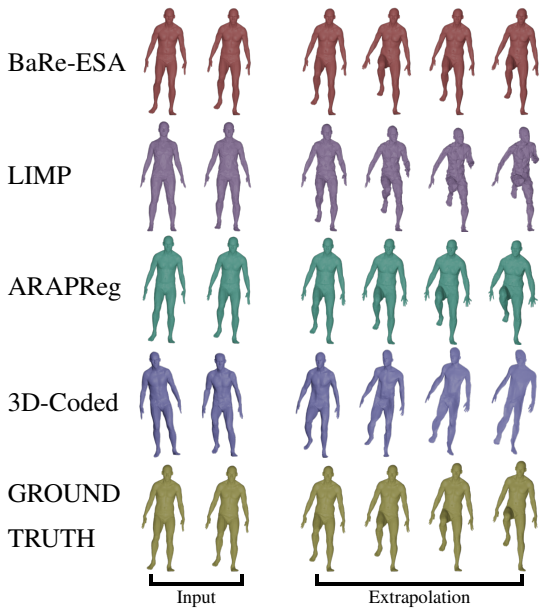

Fig. 4: Extrapolation results comparison between our method, LIMP, ARAPReg and DFAUST Ground Truth. While all methods capture the primary motion of lifting a leg, the extrapolations of LIMP and ARAPreg include extraneous motions of arms and slight changes in body type.

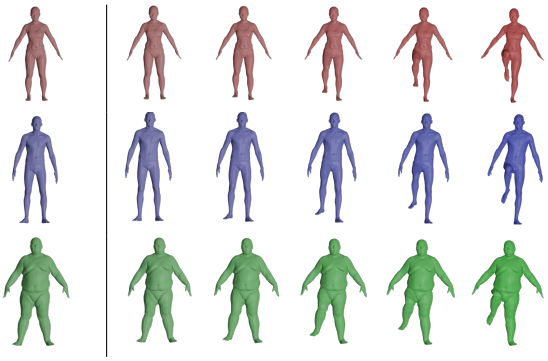

Fig. 5: Motion Transfer: We display the original motion in the top row and the transfer of the motion to the target shapes in the second and third row.

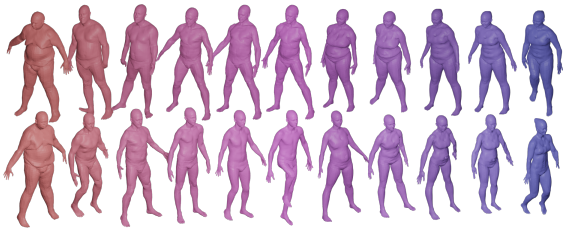

Fig. 6: Random Shapes: 22 random shapes generated using a Gaussian mixture model on the space of initial velocities.

to be sufficient to get visually satisfying random body surfaces. The generation process consists of sampling a pose and shape vector in the tangent space and solving the corresponding geodesic initial value problem from the template in the direction of the generated vector. We display a selection of 22 generated shapes in Fig. 6.

6.4 Supervised and Unsupervised Disentanglement

All the above experiments made use of data with labels that distinguish identities and poses in the training phase. In this section, we will show that the obtained latent space representation inherits this property from the training phase, i.e., that the obtained latent variables do split into a set of pose deformations and a set of identity deformations. Besides, recent works (Yang et al., 2023) have started to explore approaches that alleviate this label dependency by constructing latent spaces that automatically disentangle these different types of deformations. In the second part of this section, we will thus show that

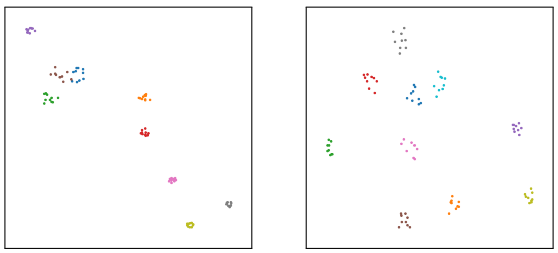

Fig. 7: On the left, we display a TSNE plot of the coefficients of a testing set of surfaces corresponding to pose deformations. The colors in this plot correspond to the ground truth poses of the testing set. On the right, we repeat this process with the coefficients of the deformation basis vectors with the colors corresponding to the ground truth identity labels.

the BaRe-ESA framework shares this capability, i.e., that the presence of labels is not a necessity and that our method can be adapted so as to automatically disentangle identity and pose information (and more generally between multiple different deformation subspaces, cf. Remark 3).

We start by calculating the latent codes of a set of testing data from DFAUST. We then calculate t-distributed Stochastic Neighbor Embeddings (TSNE) of the coefficients corresponding to both the pose and the identity coefficients. As expected, the clustering shown in Figure 7 matches the ground truth exactly, proving that the developed deformation basis effectively separates changes in pose from those in identity.

Next we demonstrate our method's ability to disentangle shape deformation modalities without relying on any prior labels. Therefore we consider again a training set of human body scans from DFAUST. In contrast to the previously described method, we will, however, construct our PCA basis without using any label information, i.e., we will treat paths (tangent vectors, resp.) stemming from motions exactly the same as paths (tangent vectors, resp.) stemming from changes in identity. In a first step, we then perform PCA on the tangent vectors of all these paths with respect to the H^2-inner product at the template shape.

After the construction of this unified deformation basis (containing both pose and identity changes) we calculate the H^2-norm of each vector in the PCA basis. Guided by the empirical observation that deformations due to changes in pose are on a different H^2 scale as compared to those caused by changes in

body shape, we separate these vectors in two groups using an automatically chosen threshold on the H^2 norm using Otsu's method (Otsu, 1979). We treat the first group as the pose deformation basis and the second group as the basis for identity deformations. Both the H^2-norms of these basis vectors and the chosen threshold are shown in Fig. 8. To compare the results with the original (label-based) basis, we take the same total number of vectors by taking the 130 vectors above the threshold as the first group and the top 40 below the threshold as the second group. We compute the chordal Grassmann distance between the subspace spanned by the pose vectors of the original basis and the subspace spanned by first group of vectors from the basis computed without prior label information as 0.00398. For comparison we (experimentally) computed the mean distance between random subspaces of the same dimension based on 1000 simulated pairs of random subspaces of the same dimension, which turned out to be 17.085. A similar computation for the shape vectors of the original bases and the second group of vectors from the unsupervised basis returns a chordal Grassmann distance of 0.00166. For comparison we computed again the mean distance between random subspaces of this dimension, which turned out to be 9.581. This experiment demonstrates that the unsupervised basis construction leads to essentially the same bases for pose and shape deformations as the original label based basis construction thereby showing the label-independence of our framework.

To further demonstrate this result qualitatively, we compute the latent coefficients with respect to the unsupervised basis for the same testing set of human bodies as used in Fig. 7. To illustrate the disentanglement of this basis, we again present TSNE plots of the coefficients corresponding to the basis vectors above and below the chosen threshold, with points colored according to their ground truth poses and identities, respectively. The resulting clustering, displayed in Fig. 8, again aligns perfectly with the ground truth as in the experiment using labeled data. This demonstrates that also the unsupervised disentanglement accurately separates the deformation basis into changes in pose and shape.

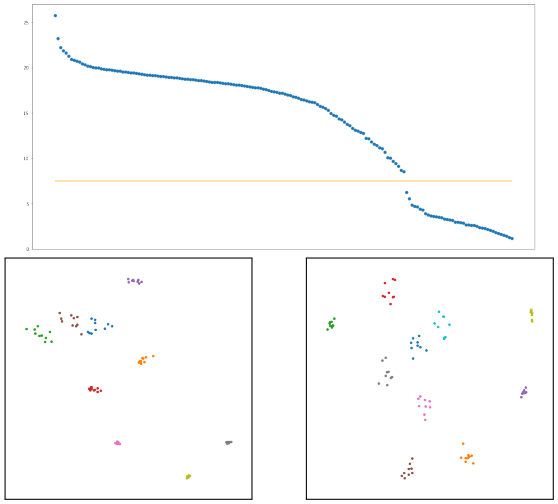

Fig. 8: Unsupervised disentanglement capabilities of BaRe-ESA: in the top plot, we present the H^2 norms of the elements of a deformation basis constructed without prior pose and shape information and the chosen threshold by which we separate the basis. Below on the left, we display a TSNE plot of the coefficients of a testing set of surfaces corresponding to the basis vectors above the threshold. The colors in this plot correspond to the ground truth poses of the testing set. On the right, we repeat this process with the coefficients of the deformation basis vectors (those below the threshold) and with the colors corresponding here to the ground truth identity labels.

Remark 3 (Automatic Disentanglement in the presence of multiple deformation modules)**.** *We should emphasize that the approach described in this section to automatically separate pose and identity deformation basis vectors extends to situations in which one is interested in splitting a given basis into multiple distinct deformation modules (subspaces), at least in situations where those different types of deformations are expected to result in H^2 energies of different orders. Following a similar process, one can simply compute the H^2 norms of each vector in the PCA basis estimated from the training set, and, using e.g. a multilevel Otsu thresholding method (Liao et al., 2001), separate those basis vectors into a finite number of groups. The final deformation modules are then constructed by choosing the dominant deformation directions from each of these clusters. More generally, when the different type of deformations are not expected to operate on different H^2-scales, other approaches to separate the latent space can be explored; e.g. one might construct non-local metrics specifically penalizing deformations of certain parts of the shape or adapt methods similar to those in the recently proposed Geolatent framework (Yang et al., 2023).*

	Hausdorff	Chamfer	Varifold
LIMP	0.15	0.087	0.034
ARAPReg	**0.12**	**0.015**	**0.0089**
3D-Coded	0.16	0.059	0.020
BaRe-ESA	**0.12**	0.016	0.0091

Table 5: Face reconstruction results. We compute the Hausdorff, Chamfer, and Varifold reconstruction errors between the outputs of the methods and the corresponding scans.

6.5 Application to Human Faces

In this part we showcase the capabilities of our framework in the context of human face scan analysis from the COMA dataset, where the chosen parameters and the testing and training data are described in Sections 5.1-5.3. As a first meansurent we claculated again the shape reconstruction error for BaRe-ESA, ARAPReg and 3D-Coded, cf. Table Tab. 5. In this task ARAPReg performed the best, with BaRe-ESA's performance being only marginally lower. The performance of 3D-Coded is an order of magnitude worse. One reason for the better performance of ARAPReg as compared to LIMP and 3D-Coded is probably the use of the varifold distance in our adapted implementation of this approach, the original implementation of ARAPReg not being capable of dealing with unregistered data. The other learning-based methods (LIMP and 3D-coded) use instead the Chamfer distance. We believe that this might be one source of the significantly worse performance of these methods on the COMA dataset.

In Figure 9, we show two latent code reconstructions of two different noisy scans of human faces, an example of an interpolation between two different expressions and an expression transfer to a different identity. Additional examples of registration, interpolations, and a qualitative comparison to the other deep learning methods are shown in the supplementary material. One can see again that our method leads to more natural interpolation and extrapolation results as compared to the other methods.

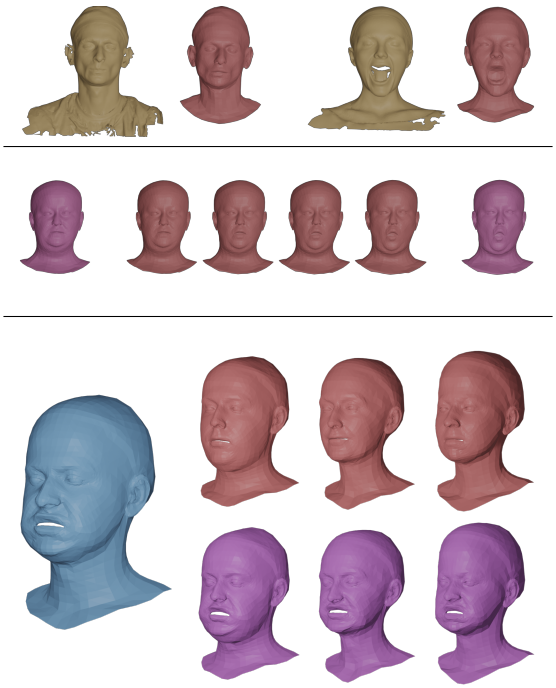

Fig. 9: Experimental results for COMA faces. Here we present several qualitative results from this framework applied to the COMA dataset. In the first row, we show our latent code reconstruction (red) of two different noisy scans of human faces (yellow). In the second row, we display an example of a solution to a geodesic boundary value problem to interpolate between two shapes (purple). In the third row, we display an example of expression transfer in our framework. The blue mesh on the left represents a target face registered with our framework, each red mesh on the right represents three additional identities and the purple meshes below represent the transfer of the expression onto these identities.

7 Experimental Results from $3D$-training data

In the previous section we took advantage of the existence of full $4D$-data during the training phase. In this section we will demonstrate the capabilities of our framework starting solely from $3D$-data. Here we will consider again the DFAUST dataset, and in addition present experiments on the MANO dataset, a (static) database of human hands.

7.1 A comparison between 3D and 4D training data using the DFaust dataset

Here we compare the results of our framework where we generate our motion basis using two different methods:

1. first, we use the same method as in the previous section, namely using all of DFaustT with real 4D data sequences containing 39159 meshes;

2. second, we start from extremely limited 3D-data: we consider only 270 scans from the FAUST dataset and we generate the necessary motion and deformation paths with an elastic matching algorithm during the training phase.

In Tab. 7 we present the mean error for the interpolation problem for the same ten DFAUST sequences considered in Section 6.2: comparing these results with those of Tab. 3 one can observe that the interpolation error is indeed higher as compared to the error obtained by training BaRe-ESA with 4D-data. Nevertheless we still outperform the three other baselines (LIMP, ARAPReg, 3D-Coded) in all three measures of performance (Hausdorff, Chamfer, Varifold). As one can see in Tab. 6 and Tab. 8 the same holds true for the extrapolation task and for the registration tasks. For the shape reconstruction task the performance drops to the level of ARAPReg and LIMP. This certainly demonstrate the advantage of having access to $4D$-training data (or at least signifincantly large training data), but at the same time shows the capability of our framework to lead to a superior performance without this additional information. We want to emphasize that all the comparison methods are trained with the full 39159 meshes, – i.e., with more than hundred times the amount of meshes – making the results of BaRe-ESA from this very limited training data all the more remarkable.

	MSE	Hausdorff	Chamfer	Varifold
mean	0.028	0.21	0.046	0.025

Table 6: Registration results using only data from Faust (3D) for training. The mean errors are calculated for a testing set from DFAUST.

	Hausdorff	Chamfer	Varifold
mean	2.004	0.683	0.405

Table 7: Interpolation results using only data from Faust (3D) for training. The mean errors are calculated for the same ten sequences from DFAUST as presented in 6.2.

7.2 Application to Human Hands

Finally we showcase the capability of our framework for human hands analysis. We apply our method to the MANO dataset. The chosen parameter and the training and testing data are described in Sections 5.1-5.3. Different from the two other modalities, MANO contains only static shapes. To create the PCA basis we use simple linear deformations between the registered hands in the training data to build our shape and pose basis. We started again with calculating the reconstruction errors, which are shown in Tab. 9. As one can see, our approach significantly outperformed the other methods by an order of magnitude. We believe that the reason lies in the limited size of the training data compared, which is not sufficient for these other methods; in the previous Sec. 7.1 we only trained our method from a small set of training data, but the comparison methods still saw the full set of training data. Qualitative results of reconstruction and interpolation are shown in Fig. 10; while there is no ground truth available the obtain interpolations follow a visually natural hand movement.

	Hausdorff	Chamfer	Varifold
mean	4.853	1.256	0.899

Table 8: Extrapolation results using only data from Faust (3D) for training. The mean errors are calculated for the same ten sequences from DFAUST as presented in 6.2.

	Hausdorff	Chamfer	Varifold
LIMP	0.063	0.035	0.0067
ARAPReg	0.039	0.016	0.0022
3D-Coded	0.068	0.031	0.0051
BaRe-ESA	**0.0053**	**0.0048**	**0.0003**

Table 9: Hands reconstruction results. We compute the Hausdorff, Chamfer, and Varifold reconstruction errors between the outputs of the methods and the corresponding scans.

8 Ablation Studies

Within this section, we conduct a sequence of ablation studies to validate our selections regarding the number of shapes and pose basis elements, the shape-matching function, and the Riemannian metric employed in the calculation of path energy.

8.1 Choice of basis size

Our first ablation study considers the choice of basis size: in Table 10, we present the registration error

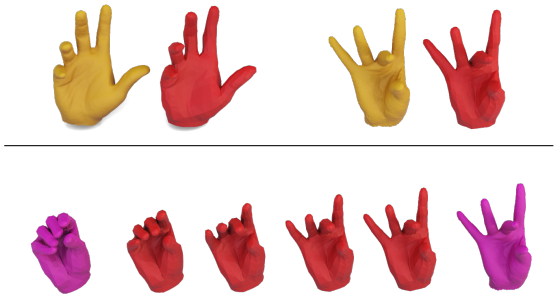

Fig. 10: Experimental results on MANO hands. Here we present several qualitative results from this framework applied to the MANO dataset. In the first row, we show our latent code reconstruction (red) of two different scans of human hands (yellow). In the second row, we display an example of a solution to a geodesic boundary value problem to interpolate between two shapes (purple).

	10	70	130	190
10	0.090	0.072	0.016	0.015
40	0.092	0.068	**0.014**	0.015
70	0.089	0.063	0.015	0.014
100	0.087	0.062	0.016	0.016

Table 10: Ablation study on the number of basis elements. We report the registration errors where we vary the number of pose and shape basis vectors of used in the matching process.The basis vectors are derived from training with DFAUST and the errors of the methods are calculated using data from FAUST.

corresponding to various numbers of shape and pose basis vectors. Each value in the table is determined by optimizing the latent vector reconstruction energy, as detailed in Equation (9), for an identical number of optimization iterations. The obtained results suggest that our choice of basis size provides the ideal balance of minimizing the latent space dimension while maximizing the expressivity of the obtained shape model.

8.2 Choice of matching functional and latent space metric

To justify our choice of shape matching functional and path energy, we compute the mean registration errors and interpolation errors for different combination of shape matching and path energy functionals. In particular, we experiment with replacing the varifold distance with the Chamfer distance and the elastic energy with the Euclidean distance on the latent space. The results of these experiments are reported in Table 11. First, we compare the performance of the Chamfer and varifold distances and demonstrate that the choice of the varifold metric leads to significantly lower registration errors than the Chamfer distance. In a second experiment, we demonstrate that the elastic matching energy produces significantly lower interpolation errors than that of an Euclidean path energy on the latent space.

Shape Matching Term	Path Energy	Registration Error	Interpolation Errors		
			Haus.	Cham.	Var.
Chamfer	H^2	0.032	1.273	0.510	0.348
Varifold	Euclidean	0.015	1.787	0.632	0.335
Varifold	H^2	**0.014**	**1.072**	**0.474**	**0.290**

Table 11: Ablation study on the shape matching and path energy functions. We report the registration and interpolation errors for each combination of shape matching and path energy functions. For these experiments, we train the method using DFAUST and tested with FAUST.

9 Conclusions

In this paper, we proposed a general framework for basis restricted elastic shape analysis on the space of unregistered surfaces. We demonstrated superior performance compared to state-of-the-art methods in various tasks such as shape registration, interpolation, motion transfer, and random pose generation. Our framework utilizes a finite-dimensional latent space representation, which we equip with a non-Euclidean Riemannian metric inherited from the family of elastic metrics. This allows for a simplified representation of shape space while preserving the ability to compare surfaces modulo shape preserving transformations, i.e., our approach does not assume pre-registered surfaces or consistent mesh structures, making it applicable to a wide range of surface meshes with real data. Furthermore, the framework shows good generalization properties and does not require a substantial amount of training data. The paper presents qualitative examples and quantitative analysis to support the effectiveness of the proposed framework in various experiments, including human body shape and pose data as well as human face and hand scans.

Lastly, we want to mention limitations and corresponding open directions for future work. Therefore we first point out that, as compared to some of the other latent space methods, the non-Euclidean nature of the latent space comes at the price of solving optimization problems to estimate interpolated or extrapolated geodesic paths, which can encumber to significant computational cost for large data applications. A possible way around this limitation would be to train neural networks in a supervised setting to learn the geometry of the latent space, i.e., to approximate the solutions of the interpolation and extrapolation problems.

Finally, we want to mention a simple yet potentially relevant extension of our model, namely to introduce distinct Sobolev Riemannian metrics on the different shape modalities, e.g. for the shape change and the pose change deformation field in the human body motions. This comes with the idea of adapting the metric to the different nature of those deformations, and thus even better disentangling these quantities.

Acknowledgments. This work was supported by ANR project ANR-19-CE23-0020 (Human4D); by NSF grants DMS-1912037, DMS-1953244, DMS-1945224 and DMS-1953267, and by FWF grant FWF-P 35813-N. Most of this work was done when Emery Pierson was at the University of Lille.

Data Availability Statement. The authors confirm that all data generated or analysed during this study are included in this published article.

Appendix A

A.1 Geodesic distance bounds

In this section we will study the induced geodesic distance of the second order Sobolev metric used in this article. For a finite dimensional Riemannian manifold the induced geodesic distance is always a true distance function, i.e., it is symmetric, satisfies the triangle inequality and is non-degenerate. This last property can, however, fail in infnite dimensions: there exists Riemannian geometries such that the geodesic distance between distinct points is zero or it might even vanish on the whole manifold. This startling phenomenon was first observed by Eliashberg and Polterovich (1993) for the H^{-1} metric on the symplectomorphism group and later by Michor and Mumford for the L^2 metric on spaces of immersions and diffeomorphisms (Michor and Mumford, 2005; Bauer et al., 2012). In the following theorem, we will prove that, under certain conditions on the parameters, the geodesic distance of the family of elastic Riemannian metric used in this article is non-degenerate:

Theorem 1. *Let $a_0 > 0$ and let either $a_1, b_1, c_1, d_1 > 0$ or $a_2 > 0$ then the induced geodesic distance of the metric G on the space Imm is non-degenerate, i.e., for any two surfaces $q_0, q_1 \in \mathrm{Imm}$ with $q_0 \neq q_1$ we have $d_G(q_0, q_1) > 0$.*

Proof. We start with the case that $a_1, b_1, c_1, d_1 > 0$. For this case we will make use of a generalization of the SRNF (Jermyn et al., 2012, 2017) as introduced in Su et al. (2020a). To be more specific in Su et al. (2020a) they considered the mapping

$$\begin{aligned} \mathcal{Q} : \mathrm{Imm} &\to \mathrm{Met}(\mathcal{T}) \times C^\infty(\mathcal{T}, \mathbb{R}^3) \\ q &\mapsto (q^* \langle ., . \rangle, \psi_q), \end{aligned} \tag{A1}$$

where $\mathrm{Met}(\mathcal{T})$ denotes the space of all Riemannian metrics on $\mathcal{T}$ and where ψ_q denotes the SRNF of q. On the space of all Riemannian metrics there exists a one parameter family of Riemannian metrics G^E, called the Ebin or DeWitt metric (DeWitt, 1967; Ebin, 1970). Among other beneficial properties this Riemannian metric admits an explicit formula for its corresponding geodesic distance as derived by Clarke (2010) and Gil-Medrano and Michor (1991). For the precise formula we refer to (Su et al., 2020a, Theorem 2). For the purpose of this proof it is only important that this distance is non-degenerate, i.e., $d_{G^E}(g_0, g_1) > 0$ if $g_0 \neq g_1$. On the second factor of the image of Q, i.e., on $C^\infty(\mathcal{T}, \mathbb{R}^3)$ we consider the standard non-invariant L^2 inner product as a Riemannina metric. This has again an explicit expression for the geodesic distance given by $d_{L^2}(\psi_1, \psi_2) = \|\psi_1 - \psi_2\|^2_{L^2}$. The relevance of these results for our family of metrics can be found in the fact, that the pull-back of this product Riemannian metric via the mapping Q yields exactly the Riemannian metric G with parameters $a_0 = d_1 = a_2 = 0$ and $a_1, b_1, c_1 \neq 0$ (depending on the parameter choice in the DeWitt metric and of the weighting of the two Riemannian metrics on the product space, see (Su et al., 2020a, Theorem 3) for the precise statement of this result).

Unfortunately the image of the map Q in the product space $\mathrm{Met}(\mathcal{T}) \times C^\infty(\mathcal{T}, \mathbb{R}^3)$ is far from being totally geodesic and thus we cannot directly calculate the geodesic distance of the metric G via this transform. Nevertheless, this construction still provides a

lower bound for the geodesic distance of G on Imm, i.e., we have:

$$d_G(q_0, q_1) \geq d_{G^E}(g_0, g_1) + \|\psi_0 - \psi_1\|^2_{L^2}, \quad \text{(A2)}$$

where $(g_i, \psi_i) = Q(q)$. Next we note, that $Q(q_0) = Q(q_1)$ if and only if q_0 and q_1 differs only by a translation and we have shown that the geodesic distance of the elastic metric G is non-degenerate on the quotient space $\mathrm{Imm}\,/\mathrm{translation}$. It remains to deal with the case that $q_0 = q_1 + v$ for some $v \in \mathbb{R}^3$. In this case the immersions q_0 and q_1 are also different elements in the quotient shape space $\mathcal{S}$ of unparametrized immersions, where the non-degeneracy has been shown using an area-swept-outbound, see Bauer et al. (2011). This concludes the proof assuming $a_1, b_1, c_1, d_1 > 0$. It remains to prove the result under the assumption that $a_2 > 0$, but in this situation the result follows directly from the above and the Sobolev embedding theorem. □

A.2 Discretization of invariant H^2 metrics

In this section, we detail the computation of the Riemannian metric term $G_q(h, h)$ for discrete meshes and vector fields. We shall however refer to Crane (2018) for a more comprehensive presentation and justification of the discrete differential approximations being used here. Let us assume that q is a triangulated oriented surface mesh given by the ordered list of vertices $V = (v_1, v_2, \ldots, v_N)$ with each $v_i \in \mathbb{R}^3$ and set of triangle faces F where each $f \in F$ corresponds to an ordered triplet of distinct indices $f = (f_0, f_1, f_2)$ of $\{1, \ldots, n\}$. We then view the vector field h as a list of vectors $(h_i)_{i=1,\ldots,N}$ attached to each vertex of q. Note that, equivalently, one can interpret the discrete q and h as piecewise affine linear maps on each face of the mesh, by interpolation of the values at the vertices.

We start with the L^2 term of the metric: $\int_{\mathcal{T}} \langle h, h\rangle \,\mathrm{vol}_q$. The discrete volume form can be first expressed over the mesh triangular faces. Specifically, for each face $f \in F$, we can calculate its area as $\mathrm{vol}_f = \|(v_{f_1} - v_{f_0}) \times (v_{f_2} - v_{f_0})\|$. The volume form on the vertices is then obtained by distributing the areas of the adjacent faces, namely for each vertex v_i, we take $\mathrm{vol}_{v_i} = \frac{1}{3}\sum_{f \ni i} \mathrm{vol}_f$. This leads to the following discrete version of the L^2 term:

$$\int_{\mathcal{T}} \langle h, h\rangle \,\mathrm{vol}_q \approx \sum_{i=1}^{N} \|h_i\|^2 \,\mathrm{vol}_{x_i}\,.$$

Next, we consider the first order terms of the metric. For any face $f \in F$, we can view both q and h as affine maps on f, by interpolation of their values at the three vertices of f. Then their differentials are constant on f and given by the following (3×2) matrices:

$$\begin{aligned} dq_f &= \left[h_{f_1} - h_{f_0}, h_{f_2} - h_{f_0}\right], \\ dh_f &= \left[h_{f_1} - h_{f_0}, h_{f_2} - h_{f_0}\right] \end{aligned}$$

We further have the following discrete versions of the metric g_q and unit normal n_q on the face f:

$$\begin{aligned} g_f &= \begin{bmatrix} \|e_{01}\|^2 & e_{01} \cdot e_{02} \\ e_{01} \cdot e_{02} & \|e_{02}\|^2 \end{bmatrix}, \\ n_f &= \frac{e_{01} \times e_{02}}{\|e_{01} \times e_{02}\|}. \end{aligned}$$

where $e_{01} = v_{f_1} - v_{f_0}$, $e_{02} = v_{f_2} - v_{f_0}$ are the two edges of the face f passing through the vertex v_{f_0}. We then rely on the interpretation and the discretization of the different first order terms introduced in Su et al. (2020b). Namely,

$$\begin{aligned} &\textstyle\int_{\mathcal{T}} g_q^{-1}(dh_m, dh_m)\,\mathrm{vol}_q \approx \\ &\quad \textstyle\sum_{f \in F} \mathrm{tr}(g_f^{-1}\delta g_f g_f^{-1}\delta g_f)\,\mathrm{vol}_f \end{aligned}$$

in which δg_f represents the variation of the metric tensor g_f resulting from the variation of the vertices of the mesh in the direction of the vector field h. In practice, in the computation of geodesics, δg_f is calculated from one discrete time point to the next by taking the difference of the respective metric tensors of face f. Similarly,

$$\int_{\mathcal{T}} g_q^{-1}(dh_+, dh_+)\,\mathrm{vol}_q \approx \sum_{f \in F} \mathrm{tr}(g_f^{-1}\delta g_f)^2\,\mathrm{vol}_f$$

where each term inside the sum can be interpreted as the change of in the area of the face f and

$$\int_{\mathcal{T}} g_q^{-1}(dh_\perp, dh_\perp)\,\mathrm{vol}_q \approx \sum_{f \in F} \langle \delta n_f, \delta n_f\rangle\,\mathrm{vol}_f$$

in which δn_f stands for the variation of the normal vector n_f resulting from the variation of the vertices

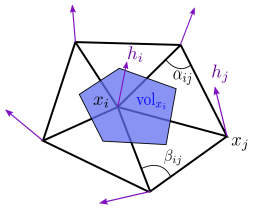

Fig. A1: Discrete volume form and Laplacian on a mesh.

of the mesh in the direction of the vector field h. The last first order term is discretized as follows:

$$\int_{\mathcal{T}} g_q^{-1}(dh_0, dh_0)\,\mathrm{vol}_q \approx \sum_{f\in F} \mathrm{tr}(g_f^{-1}\xi_f g_f^{-1}\xi_f^T)\,\mathrm{vol}_f$$

where $\xi_f = dq_f^T dh_f - dh_f^T dq_f$.

Finally, for the discretization of the Laplacian in the second order term of the metric, we use the standard approximation on triangular meshes based on the cotangent formula. Letting E be the set of oriented edges in the mesh viewed as ordered pairs of distinct vertex indices, we take for any $i \in \{1, \dots, N\}$:

$$(\Delta_q h)_{v_i} = \sum_{\substack{j|(i,j)\in E \\ \text{or } (j,i)\in E}} (\cot(\alpha_{ij}) + \cot(\beta_{ij}))(h_i - h_j).$$

where α_{ij} and β_{ij} are the angles defined as in Figure A1. Then the full discrete second order order term is obtained as:

$$\int_{\mathcal{T}} \langle \Delta_q h, \Delta_q k\rangle\,\mathrm{vol}_q \approx \sum_{i=1}^{N} \|(\Delta_q h)_{v_i}\|^2\,\mathrm{vol}_{x_i}\,.$$

A.3 Mesh invariant similarity measures

In this section, we add some details regarding the similarity metrics being used in the registration procedure as well as for the evaluation and comparison of the different methods. With similar notations to the previous section, we consider two discrete surfaces q and q' with possibly different number of vertices and mesh structure. We denote by $(v_1, \dots, v_N)$ the vertices of q and F its set of faces, and similarly $(v'_1, \dots, v'_{N'})$ and F' the vertices and faces of q'.

First, we remind that the Hausdorff distance between the two shapes is given by the formula:

$$d_{\mathrm{H}}(q, q') = \max\left\{ \sup_{i=1,\dots,N} \inf_{j=1,\dots,N'} \|v_i - v'_j\|, \sup_{j=1,\dots,N'} \inf_{i=1,\dots,N} \|v'_j - v_i\| \right\}$$

In our numerical experiments, we use the approximate implementation provided by libigl (Jacobson et al., 2018). Note that this metric is typically very sensitive to outliers.

In contrast, the Chamfer distance (Fan et al., 2017; Groueix et al., 2018b) provides a more regular similarity cost which is defined as:

$$d_{\mathrm{Ch}}(q, q') = \frac{1}{N}\sum_{i=1}^{N} \inf_{j=1,\dots,N'} \|v_i - v'_j\| + \frac{1}{N'}\sum_{j=1}^{N'} \inf_{i=1,\dots,N} \|v'_j - v_i\|.$$

We use the Pytorch implementation of Thibault Groueix[3]. One of the downsides of this metric for comparing discrete surfaces, however, is that it is not necessarily robust to local changes of point density since it is designed as a distance between point clouds (without taking the triangle mesh into account) and it remains somewhat sensitive to outliers and noise (Wu et al., 2021).

As similarity terms for the algorithms of this paper and final measure of reconstruction quality, we instead favor distances that are based on measure representations of shapes, as introduced in (Charon and Trouvé, 2013; Kaltenmark et al., 2017). Specifically, we rely on the representation of surfaces as *varifolds* equipped with kernel Hilbert metrics. The resulting family of metrics is equally defined for continuous and discrete surfaces and the properties of those metrics have been well studied, c.f. the aforementioned papers. In practice, they are computed via the following formula:

$$d_{\mathrm{Var}}(q, q')^2 = \sum_{f,\tilde{f}\in F} k(c_f, n_f, c_{\tilde{f}}, n_{\tilde{f}})\,\mathrm{vol}_f\,\mathrm{vol}_{\tilde{f}} - 2\sum_{\substack{f\in F \\ f'\in F'}} k(c_f, n_f, c_{f'}, n_{f'})\,\mathrm{vol}_f\,\mathrm{vol}_{f'}$$

[3] https://github.com/ThibaultGROUEIX/ChamferDistancePytorch

Method	Training	Retrieval	Interpolation	
LIMP	1.5w	<1s	<1s	
3D-Coded	12h	160s	<1s	160s
ARAPReg	2w	160s	<1s	160s
BaRe-ESA	<1h	160s	91s	160s

Table A1: This table presents the computation costs for the different methods and the three tasks of model training, latent code retrieval (once trained) and interpolation between two shapes. In the case of interpolation, we display on the left the running times when latent codes are available vs, on the right, when they are not.

$$+ \sum_{f', \tilde{f}' \in F'} k(c'_{f'}, n'_{f'}, c'_{\tilde{f}'}, n_{\tilde{f}'}) \operatorname{vol}_f \operatorname{vol}_{\tilde{f}'}$$

where c_f (resp. $c'_{f'}$) denote the barycenter of the triangle face f (resp. f') in q (resp. q'). Here k is a positive definite kernel function on $\mathbb{R}^3 \times S^2$. While several different families of kernels are possible (see the discussion in Kaltenmark et al. (2017)), in all the experiments of this paper, we specifically take $k(x, n, x', n') = e^{-\frac{\|x-x'\|^2}{\sigma^2}}(n \cdot n')^2$ where σ can be interpreted as a spatial scale of sensitivity of the metric which is chosen to be quite small ($\sigma = 0.025$) in our examples. In this work, we adapted the Python implementation used in $H2\_SurfaceMatch$[4] which itself relies on the *PyKeops* library (Feydy et al., 2020) for efficient evaluation and automatic differentiation of kernel functions on the GPU. We emphasize that such varifold metrics derive from distances between continuous surfaces which are independent of their parametrization. In practice, when considering discrete surface meshes, this typically leads to those metrics being approximately insensitive to variations in mesh sampling, at least for a certain range of kernel scale σ. We illustrate this property empirically with the example of Figure A2.

A.4 Computational cost

As stated in the paper, our pipelines are optimization based. We provide a substantial comparison for the different approaches.

All the other approaches require significant training costs compared to BaRe-ESA which requires less than one hour, cf Table A1. On the other hand, BaRe-ESA, ARAPReg and 3d-Coded require additional optimization for the latent code retrieval, which we found takes approximately the same time for all three methods. The optimization cost is driven by the mesh invariant costs – varifold or Chamfer – which have n^2 complexity, where n is the number of vertices. LIMP is the only method that does not require optimization, but the network behaves notably bad when the poses are unseen as showed in the experiments. For the interpolation problem our method requires approximately 90 seconds if the latent codes are already available, whereas it takes approximately the same time as one latent code retrieval if they are not available. All timing results were obtained using a standard home PC with a Intel 3.2 GHz CPU and a GeForce GTX 2070 1620 MHz GPU.

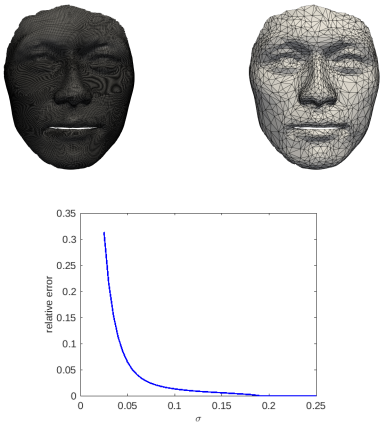

Fig. A2: Empirical illustration of the varifold distances approximate invariance to mesh sampling. Top row: a triangular mesh of a human face with 57,836 faces (left) and its downsampled version with 2000 faces (right). Bottom plot: the relative error in varifold norm $d_{Var}(q, q')/\|q'\|_{Var}$ between the full surface and the downsampled one, as a function of the kernel scale σ. One can see that this relative error remains close to 0 for scales larger than $\sigma = 0.1$ but increases for smaller kernels. Note for reference that the surface diameter is normalized to 1 while the average diameter of the mesh triangles in the original and downsampled mesh are respectively 9.4×10^{-3} and 4.6×10^{-2}.

[4] https://github.com/emmanuel-hartman/H2_SurfaceMatch

A.5 Description of state-of-the-art methods

We propose a detailed description of the state-of-the-art method we use as baselines. We selected deep learning methods that builds a flat latent space for human shape deformations. They describe as follows:

- Learning Latent Shape Representations with Metric Preservation (LIMP) is a deep learning method modeling deformations of shapes using a variational auto encoder with geodesic constraints. The encoder part use a PointNet architecture, which makes it invariant to parameterization. The decoder part is a Multi Layer Perceptron. The geometric constraints are used a loss functions during the training process. The latent space is divided in an extrinsic part and an intrinsic part and the loss is applied on the interpolation in those dimensions. The intrinsic part is constrained using the computation of full geodesic matrix, which make the training process particularly heavy.
- As-Rigid-As-Possible Regularization (ARAP) is a deep learning method modeling deformations of shapes using an auto-decoder architecture. The latent codes and the decoder are learned altogether. During the training, an As-Rigid-As-Possible loss is imposed such that the decoder directions are similar to the ARAP ones. This procedure also makes the training procedure heavy. In order to make it parameterization invariant, we replace the L^2 metric by the varifold distance, as an alternative to our Riemannian latent space.
- 3D correspondences by deep deformation (3D Coded) is a deep learning method modeling deformations of shapes using a variational auto encoder. Similarly to LIMP, the encoder part use a PointNet architecture, which makes it invariant to parameterization. The decoder uses a Multi Layer Perceptron to deform a template mesh, but no constraint is imposed on the interpolation of latent variables. By taking advantage of a high number of training samples (> 200000), they obtained state-of-the-art results for human shape correspondence.
- Functional Automatic Registration Method for 3D Human Bodies (FARM) is a functional-maps based approach for human body registration. The approach consists of multiple stages that enhance the initial mesh structure of a human body scan to propose a valid final functional map, based on a set of 15 landmark extracted automatically from the scan, between a given human body scan and a human body template. A final step of registration between the SMPL body model and the obtained correspondence is proposed.

In the paper, all those methods are trained using the same training set as Bare-ESA, from Dynamic FAUST and reported parameters from the respective papers.

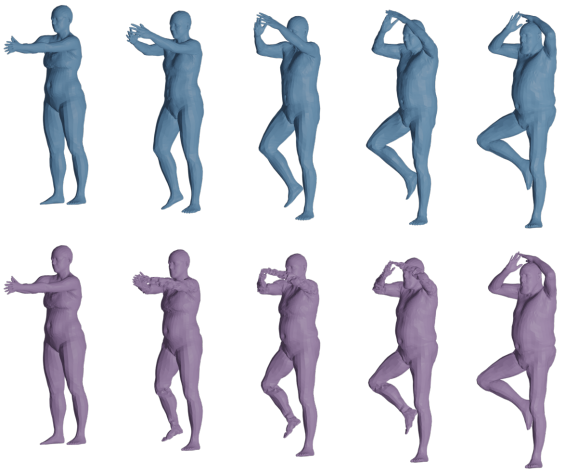

Fig. B3: First line: optimal deformation calculated using the basis informed ESA of the present article. Second line: optimal deformation calculated using a standard H^2-matching.

Appendix B

B.1 Comparison to the framework of Hartman et al. (2023b)

In Figure B3 we compare BaRe-ESA to the unrestricted method of Hartman et al. (2023b). Note, that BaRE-ESA is significantly cheaper to compute as we reduced the dimension of the minimization problem – the latent space dimension will be in the order of 100s, while the dimension of the unrestricted method is on the order of 10000s. More importantly, one can observe that BaRe-ESA leads to significantly more natural deformations, cf. the movement of the arms in Fig. B3.

References

Anuj Srivastava and Eric P Klassen. *Functional and shape data analysis*, volume 1. Springer, 2016.

Nicolas Charon and Alain Trouvé. The varifold representation of nonoriented shapes for diffeomorphic

registration. *SIAM journal on Imaging Sciences*, 6 (4):2547–2580, 2013.

Emmanuel Hartman, Emery Pierson, Martin Bauer, Nicolas Charon, and Mohamed Daoudi. Bare-esa: A riemannian framework for unregistered human body shapes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14181–14191, October 2023a.

Paul Audain Desrosiers, Yasmine Bennis, Mohamed Daoudi, Boulbaba Ben Amor, and Pierre Guerreschi. Analyzing of facial paralysis by shape analysis of 3d face sequences. *Image Vis. Comput.*, 67:67–88, 2017.

Dafei Qin, Jun Saito, Noam Aigerman, Thibault Groueix, and Taku Komura. Neural face rigging for animating and retargeting facial meshes in the wild. In *ACM SIGGRAPH 2023 Conference Proceedings, SIGGRAPH 2023, Los Angeles, CA, USA, August 6-10, 2023*, pages 68:1–68:11, 2023.

Naima Otberdout, Claudio Ferrari, Mohamed Daoudi, Stefano Berretti, and Alberto Del Bimbo. Sparse to dense dynamic 3d facial expression generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 20353–20362, 2022a.

Bailin Deng, Yuxin Yao, Roberto M. Dyke, and Juyong Zhang. A survey of non-rigid 3d registration. *Comput. Graph. Forum*, 41(2):559–589, 2022. doi: 10.1111/CGF.14502.

Yan Zhang, Mohamed Hassan, Heiko Neumann, Michael J. Black, and Siyu Tang. Generating 3d people in scenes without people. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 6193–6203, 2020.

Ulf Grenander and Michael I Miller. Computational anatomy: An emerging discipline. *Quarterly of applied mathematics*, 56(4):617–694, 1998.

M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61:139–157, 2005.

L. Younes. *Shapes and diffeomorphisms*, volume 171. Springer, 2019.

Laurent Younes. Computable elastic distances between shapes. *SIAM Journal on Applied Mathematics*, 58(2):565–586, 1998.

Michel A Audette, Frank P Ferrie, and Terry M Peters. An algorithmic overview of surface registration techniques for medical imaging. *Medical image analysis*, 4(3):201–217, 2000.

Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (TOG)*, 31 (4):1–11, 2012.

Sebastian Kurtek, Eric Klassen, John C. Gore, Zhaohua Ding, and Anuj Srivastava. Elastic geodesic paths in shape space of parameterized surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(9): 1717–1730, 2012.

Ian H Jermyn, Sebastian Kurtek, Hamid Laga, and Anuj Srivastava. Elastic shape analysis of three-dimensional objects. *Synthesis Lectures on Computer Vision*, 12(1):1–185, 2017.

Zhe Su, Martin Bauer, Eric Klassen, and Kyle Gallivan. Simplifying transformations for a family of elastic metrics on the space of surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 848–849, 2020a.

Alice Barbara Tumpach, Hassen Drira, Mohamed Daoudi, and Anuj Srivastava. Gauge Invariant Framework for Shape Analysis of Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):46–59, 2016.

Hamid Laga, Marcel Padilla, Ian H Jermyn, Sebastian Kurtek, Mohammed Bennamoun, and Anuj Srivastava. 4d atlas: Statistical analysis of the spatio-temporal variability in longitudinal 3d shape data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

Marc Vaillant and Joan Glaunes. Surface matching via currents. In *Biennial international conference on information processing in medical imaging*, pages 381–392. Springer, 2005.

Martin Bauer, Nicolas Charon, Philipp Harms, and Hsi-Wei Hsieh. A numerical framework for elastic surface matching, comparison, and interpolation. *Int. J. Comput. Vis.*, 129(8):2425–2444, 2021.

Emmanuel Hartman, Yashil Sukurdeep, Eric Klassen, Nicolas Charon, and Martin Bauer. Elastic shape analysis of surfaces with second-order sobolev metrics: a comprehensive numerical framework. *International Journal of Computer Vision*, 131(5):1183–1209, 2023b.

S. Arguillère, E. Trélat, A. Trouvé, and L. Younes. Shape deformation analysis from the optimal control viewpoint. *Journal de Mathématiques Pures et Appliquées*, 104(1):139–178, 2015.

Barbara Gris, Stanley Durrleman, and Alain Trouvé. A sub-riemannian modular framework for diffeomorphism-based analysis of shape ensembles. *SIAM Journal on Imaging Sciences*, 11(1): 802–833, 2018.

Benjamin Charlier, Jean Feydy, David W. Jacobs, and Alain Trouvé. Distortion Minimizing Geodesic Subspaces in Shape Spaces and Computational Anatomy. In *VipIMAGE 2017*, pages 1135–1144, 2018.

Dai-Ni Hsieh, Sylvain Arguillère, Nicolas Charon, and Laurent Younes. Diffeomorphic shape evolution coupled with a reaction-diffusion PDE on a growth potential. *Quarterly of Applied Mathematics*, 80(1), 2022.

Nicolas Charon and Laurent Younes. Shape spaces: From geometry to biological plausibility. *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*, pages 1929–1958, 2023.

Emery Pierson, Mohamed Daoudi, and Alice-Barbara Tumpach. A riemannian framework for analysis of human body surface. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2991–3000, January 2022.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Giorgos Bouritsas, Sergiy Bokhnyak, Stylianos Ploumpis, Michael Bronstein, and Stefanos Zafeiriou. Neural 3d morphable models: Spiral convolutional networks for 3d shape representation learning and generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7213–7222, 2019.

Clément Lemeunier, Florence Denis, Guillaume Lavoué, and Florent Dupont. Representation learning of 3d meshes using an autoencoder in the spectral domain. *Computers & Graphics*, 107:131–143, 2022.

Qixing Huang, Xiangru Huang, Bo Sun, Zaiwei Zhang, Junfeng Jiang, and Chandrajit Bajaj. Arapreg: An as-rigid-as possible regularization loss for learning deformable shape generators. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5815–5825, 2021.

Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. 3d-coded : 3d correspondences by deep deformation. In *ECCV*, 2018a.

Naima Otberdout, Claudio Ferrari, Mohamed Daoudi, Stefano Berretti, and Alberto Del Bimbo. Sparse to dense dynamic 3d facial expression generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20385–20394, 2022b.

Thomas Besnier, Sylvain Arguillère, Emery Pierson, and Mohamed Daoudi. Toward mesh-invariant 3d generative deep learning with geometric measures. *Comput. Graph.*, 115:309–320, 2023.

Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 77–85, 2017.

Giovanni Trappolini, Luca Cosmo, Luca Moschella, Riccardo Marin, Simone Melzi, and Emanuele Rodolà. Shape registration in the time of transformers. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 5731–5744, 2021.

Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusionnet: Discretization agnostic learning on surfaces. *ACM Transactions on Graphics (TOG)*, 41(3):1–16, 2022.

Ruben Wiersma, Ahmad Nasikun, Elmar Eisemann, and Klaus Hildebrandt. Deltaconv: anisotropic operators for geometric deep learning on point clouds. *ACM Transactions on Graphics (TOG)*, 41 (4):1–10, 2022.

Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. 3d-coded: 3d correspondences by deep deformation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 230–246, 2018b.

Luca Cosmo, Antonio Norelli, Oshri Halimi, Ron Kimmel, and Emanuele Rodolà. Limp: Learning latent shape representations with metric preservation priors. In *Computer Vision–ECCV 2020: 16th*

European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, pages 19–35. Springer, 2020.

Sanjeev Muralikrishnan, Siddhartha Chaudhuri, Noam Aigerman, Vladimir G. Kim, Matthew Fisher, and Niloy J. Mitra. GLASS: geometric latent augmentation for shape spaces. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2022.

Matan Atzmon, David Novotny, Andrea Vedaldi, and Yaron Lipman. Augmenting implicit neural shape representations with explicit deformation fields. *arXiv preprint arXiv:2108.08931*, 2021.

Garvita Tiwari, Dimitrije Antić, Jan Eric Lenssen, Nikolaos Sarafianos, Tony Tung, and Gerard Pons-Moll. Pose-ndf: Modeling human pose manifolds with neural distance fields. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 572–589, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-20065-6.

Oren Freifeld and Michael J Black. Lie bodies: A manifold representation of 3d human shape. *ECCV (1)*, 7572:1–14, 2012.

Peter W Michor and David Mumford. Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms. *Doc. Math*, 10:217–245, 2005.

Martin Bauer, Martins Bruveris, Philipp Harms, and Peter W Michor. Vanishing geodesic distance for the riemannian metric with geodesic equation the kdv-equation. *Annals of Global Analysis and Geometry*, 41:461–472, 2012.

Martin Bauer, Philipp Harms, and Peter W Michor. Sobolev metrics on shape space of surfaces. *Journal of Geometric Mechanics*, 3(4):389, 2011.

Zhe Su, Martin Bauer, Stephen C Preston, Hamid Laga, and Eric Klassen. Shape analysis of surfaces using general elastic metrics. *Journal of Mathematical Imaging and Vision*, 62:1087–1106, 2020b.

Irène Kaltenmark, Benjamin Charlier, and Nicolas Charon. A general framework for curve and surface comparison and registration with oriented varifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3346–3355, 2017.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, October 2015.

Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: shape completion and animation of people. In *ACM SIGGRAPH 2005 Papers*, pages 408–416. 2005.

Martin Rumpf and Benedikt Wirth. Discrete geodesic calculus in shape space and applications in the space of viscous fluidic objects. *SIAM Journal on Imaging Sciences*, 6(4):2581–2602, 2013.

Federica Bogo, Javier Romero, Gerard Pons-Moll, and Michael J. Black. Dynamic FAUST: Registering human bodies in motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Piscataway, NJ, USA, June 2014. IEEE.

R. Marin, S. Melzi, E. Rodolà, and U. Castellani. Farm: Functional automatic registration method for 3d human bodies. *Computer Graphics Forum*, 39 (1):160–173, 2020.

Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J Black. Generating 3d faces using convolutional mesh autoencoders. In *Proceedings of the European conference on computer vision (ECCV)*, pages 704–720, 2018.

Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. *ACM Trans. Graph.*, 36(6):194–1, 2017.

Javier Romero, Dimitrios Tzionas, and Michael J Black. Embodied hands: modeling and capturing hands and bodies together. *ACM Transactions on Graphics (TOG)*, 36(6):1–17, 2017.

Sanjeev Muralikrishnan, Chun-Hao Paul Huang, Duygu Ceylan, and Niloy J Mitra. Bliss: Bootstrapped linear shape space. *arXiv preprint arXiv:2309.01765*, 2023.

Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 605–613, 2017.

M. Eisenberger, D. Novotny, G. Kerchenbaum, P. Labatut, N. Neverova, D. Cremers, and A. Vedaldi. Neuromorph: Unsupervised shape interpolation and correspondence in one go. In

IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

S. Melzi, R. Marin, E. Rodolà, U. Castellani, J. Ren, A. Poulenard, P. Wonka, and M. Ovsjanikov. Matching Humans with Different Connectivity. In Silvia Biasotti, Guillaume Lavoué, and Remco Veltkamp, editors, *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Association, 2019. ISBN 978-3-03868-077-2. doi: 10.2312/3dor.20191070.

Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J Black. Keep it smpl: Automatic estimation of 3d human pose and shape from a single image. In *European conference on computer vision*, pages 561–578. Springer, 2016.

Mohamed Omran, Christoph Lassner, Gerard Pons-Moll, Peter Gehler, and Bernt Schiele. Neural body fitting: Unifying deep learning and model based human pose and shape estimation. In *2018 international conference on 3D vision (3DV)*, pages 484–494. IEEE, 2018.

Haitao Yang, Bo Sun, Liyan Chen, Amy Pavel, and Qixing Huang. Geolatent: A geometric approach to latent space design for deformable shape generators. *ACM Transactions on Graphics (TOG)*, 42(6): 1–20, 2023.

Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. doi: 10.1109/TSMC.1979.4310076.

Ping-Sung Liao, Tse-Sheng Chen, Pau-Choo Chung, et al. A fast algorithm for multilevel thresholding. *J. Inf. Sci. Eng.*, 17(5):713–727, 2001.

Yakov Eliashberg and Leonid Polterovich. Bi-invariant metrics on the group of hamiltonian diffeomorphisms. *Internat. J. Math*, 4(5):727–738, 1993.

Ian H. Jermyn, Sebastian Kurtek, Eric Klassen, and Anuj Srivastava. Elastic shape matching of parameterized surfaces using square root normal fields. In *ECCV (5)*, pages 804–817, 2012.

B. S. DeWitt. Quantum theory of gravity. I. The canonical theory. *Phys. Rev.*, 160 (5):1113–1148, 1967.

David G Ebin. The manifold of Riemannian metrics, in: Global analysis, berkeley, calif., 1968. In *Proc. Sympos. Pure Math.*, volume 15, pages 11–40, 1970.

Brian Clarke. The metric geometry of the manifold of Riemannian metrics over a closed manifold. *Calculus of Variations and Partial Differential Equations*, 39(3-4):533–545, 2010.

Olga Gil-Medrano and Peter W Michor. The Riemannian manifold of all Riemannian metrics. *Quarterly Journal of Mathematics (Oxford)*, 42:183–202, 1991.

Keenan Crane. Discrete differential geometry: An applied introduction. *Notices of the AMS, Communication*, pages 1153–1159, 2018.

Alec Jacobson, Daniele Panozzo, et al. libigl: A simple C++ geometry processing library, 2018. https://libigl.github.io/.

Tong Wu, Liang Pan, Junzhe Zhang, Tai WANG, Ziwei Liu, and Dahua Lin. Balanced Chamfer Distance as a Comprehensive Metric for Point Cloud Completion. In *Advances in Neural Information Processing Systems*, volume 34, pages 29088–29100. Curran Associates, Inc., 2021.

Jean Feydy, Joan Glaunès, Benjamin Charlier, and Michael Bronstein. Fast geometric learning with symbolic matrices. *Advances in Neural Information Processing Systems*, 33, 2020.