Published in final edited form as:

IEEE Trans Pattern Anal Mach Intell. 2023 February ; 45(2): 1335–1352. doi:10.1109/TPAMI.2022.3163720.

4D Atlas: Statistical Analysis of the Spatiotemporal Variability in Longitudinal 3D Shape Data

Hamid Laga,
Information Technology Discipline, Murdoch University, Murdoch, 6150 (Australia), with the Harry Butler Institute, Murdoch University, Murdoch, 6150 (Australia)

Marcel Padilla,
Technische Universität Berlin, Germany

Ian H. Jermyn,
Durham University

Sebastian Kurtek [Senior Member, IEEE],
Ohio State University, US

Mohammed Bennamoun [Senior Member, IEEE],
University of Western Australia, Perth, WA 6009, Australia

Anuj Srivastava [Fellow, IEEE]
Florida State University, US

Abstract

We propose a novel framework to learn the spatiotemporal variability in longitudinal 3D shape data sets, which contain observations of objects that evolve and deform over time. This problem is challenging since surfaces come with arbitrary parameterizations and thus, they need to be spatially registered. Also, different deforming objects, hereinafter referred to as *4D surfaces*, evolve at different speeds and thus they need to be temporally aligned. We solve this spatiotemporal registration problem using a Riemannian approach. We treat a 3D surface as a point in a shape space equipped with an elastic Riemannian metric that measures the amount of bending and stretching that the surfaces undergo. A 4D surface can then be seen as a trajectory in this space. With this formulation, the statistical analysis of 4D surfaces can be cast as the problem of analyzing trajectories embedded in a nonlinear Riemannian manifold. However, performing the spatiotemporal registration, and subsequently computing statistics, on such nonlinear spaces is not straightforward as they rely on complex nonlinear optimizations. Our core contribution is the mapping of the surfaces to the space of Square-Root Normal Fields (SRNF) where the $\mathbb{L}^2$ metric is equivalent to the partial elastic metric in the space of surfaces. Thus, by solving the spatial registration in the SRNF space, the problem of analyzing 4D surfaces becomes the problem of analyzing trajectories embedded in the SRNF space, which has a Euclidean structure. In this paper, we develop the building blocks that enable such analysis. These include: **(1)** the spatiotemporal registration of arbitrarily parameterized 4D surfaces even in the presence of large elastic deformations and large variations in their execution rates; **(2)** the computation of geodesics

H.Laga@murdoch.edu.au .

between 4D surfaces; **(3)** the computation of statistical summaries, such as means and modes of variation, of collections of 4D surfaces; and **(4)** the synthesis of random 4D surfaces. We demonstrate the performance of the proposed framework using 4D facial surfaces and 4D human body shapes.

Keywords

Dynamic surfaces; Elastic metric; Square-Root Normal Field; Statistical summaries; Shape synthesis and generation; 4D surface; Human4D; Face4D; Motion; Growth

1 Introduction

Shape, an essential property of natural and man-made 3D objects, deforms over time as a result of many internal and external factors. For instance, anatomical organs such as bones, kidneys, and subcortical structures in the brain deform due to natural growth or disease progression; human faces deform as a consequence of talking, executing facial expressions, and aging. Similarly, actions and motions such as walking, jumping, and running are the result of a deformation, over time, of the human body shape. The ability to understand and model **(1)** the typical deformation patterns of a class of 3D objects, and **(2)** the variability of these deformations within and across object classes has many applications. For example, in medical diagnosis and biological growth modeling, one is interested in measuring the intensity of pain from facial deformations [1], and in distinguishing between normal growth and disease progression using the deformation of body shape over time. In computer vision and graphics, the ability to statistically model such spatiotemporal variability can be used to summarize collections of 3D animations, and simulate animations and motions. Similar to 3D morphable models [2], these tools can also be used in a generative model for synthesizing large corpora of labeled longitudinal 3D shape data, e.g., 4D faces, for training deep neural networks.

This paper proposes a novel framework for the statistical analysis of longitudinal 3D shape data composed of objects that deform over time. Each object is represented as a closed manifold surface. We refer to an object captured at different points in time, e.g., a 3D human face performing a facial expression or speaking a sentence, or a 3D human body shape growing or performing actions, as a *4D (or 3D + t) surface*. Given a set of such 4D surfaces, our goals are to:

- Compute the mean deformation pattern, *i.e.,* the statistical mean 4D surface. For example, the same person can smile in different ways. Similarly, different people smile differently. The goal is to learn, based on observed longitudinal shape data, the typical smile.
- Compute the main directions of variation, analogous to Principal Component Analysis (PCA) for modeling 3D shape variability [3], [4], but here we focus on modeling variability in 4D surface collections.
- Characterize a population of 4D surfaces using statistical models.
- Synthesize new 4D surfaces by sampling from these statistical models.

We refer to these tasks as the process of constructing a 4D atlas. Achieving this goal requires solving important fundamental challenges. In fact, 3D objects such as faces, human body shapes, and anatomical organs, which come with arbitrary parameterizations, exhibit large elastic deformations within the same subject and across different subjects. This makes their spatial registration, *i.e.,* finding one-to-one correspondences between each pair of shapes, very challenging. In the case of 4D surfaces, there is an additional temporal variability due to different execution rates (speeds) of evolution within and across subjects. For instance, a walking action can be executed at variable speeds even by the same person. Thus, the statistical analysis of the spatiotemporal variability in samples of 4D surfaces requires efficient spatiotemporal registration of these samples. *Spatial registration* refers to the process of finding a one-to-one correspondence between two 3D surfaces of the same individual, captured at different points in time, or of different individuals. *Temporal registration* refers to the problem of finding the optimal time warping that aligns 4D surfaces, e.g., walking actions, performed at different execution rates.

In this paper, we treat a 4D surface as a trajectory in a high-dimensional nonlinear space. We then formulate the problem of analyzing the spatiotemporal variability of 4D surfaces as the statistical analysis of elastic trajectories, where elasticity corresponds to variations in the execution rates of the 4D surfaces. However, performing statistics on trajectories embedded in nonlinear spaces of high dimension is computationally expensive since it relies on nonlinear optimizations. Our core contribution in this paper is the mapping of the surfaces to the space of Square-Root Normal Fields (SRNF) [4], [5], which has a Euclidean structure (see Section 3.1—in particular, the $\mathbb{L}^2$ metric in the space of SRNFs is equivalent to the partial elastic metric in the space of surfaces), meaning that the problem of analyzing 4D surfaces becomes the problem of analyzing trajectories, or curves, embedded in the Euclidean space of SRNFs.

This paper develops the building blocks that enable such analysis. We then use these building blocks to compute statistical summaries, such as means and modes of variation of collections of 4D surfaces, and for the automatic synthesis of novel 4D surfaces. We demonstrate the utility and performance of the proposed framework using 4D facial surfaces from the VOCA dataset [6], 4D human body shapes from the Dynamic FAUST (DFAUST) dataset [7], and dressed 4D human body shapes from the CAPE dataset [8]. Our approach is, however, general and applies to all spherically-parameterized surfaces. In summary, the main contributions of this paper are as follows.

- We represent 4D surfaces as trajectories in the space of SRNFs, which has a Euclidean structure (Section 3.1). This key contribution enables the usage of standard computational tools for the analysis and modeling of 4D surfaces (Section 3.2).
- We propose efficient algorithms for the spatiotemporal registration of 4D surfaces and the computation of geodesics between such 4D surfaces, even in the presence of large elastic deformations and significant variation in execution rates (Sections 3.2.2 and 3.2.3).

- The framework does not explicitly or implicitly assume that the correspondences between the surfaces are given. It simultaneously solves for the spatial and temporal registrations, and for the 4D geodesics that are optimal under the proposed metrics.
- We develop computational tools for **(1)** computing summary statistics of 4D surfaces and **(2)** synthesizing 4D surfaces from formal statistical models (Section 4).

The remainder of this paper is organized as follows. We first discuss related work in Section 2. Section 3 describes the proposed mathematical framework. Section 4 discusses its application to various statistical analysis tasks. Section 5 presents the results and discusses the performance of the proposed framework. Section 6 summarizes the main findings of this paper and discusses future research directions. The Supplementary Material includes more technical details, additional results, and further performance analyses.

2 Related work

We classify the state-of-the-art into two categories. Methods in the first category focus on cross-sectional shape data (Section 2.1). Methods in the second category focus on longitudinal shape data (Section 2.2).

2.1 Statistical models of cross-sectional 3D shape data

Modeling shape variability in 2D and 3D objects has been studied extensively in the literature. Early methods use Principal Component Analysis (PCA) to characterize the shape space of objects. Initially introduced for the analysis of planar shapes, the active shape model of Cootes *et al.* [9] has been extended to 3D faces [10] and 3D human bodies [3]; see [2] for a detailed survey. These methods represent 3D objects as discrete sets of landmarks, e.g., vertices, which are assumed to be in correspondence across a population of objects, and use standard Euclidean metrics for their comparison. Thus, they are limited to 3D objects that undergo small elastic deformations.

To handle large nonlinear variations, e.g., elastic deformations such as the bending and stretching observed in 3D human body shapes, Anguelov *et al.* [11] introduced SCAPE, which represents body shape and pose-dependent shape in terms of triangle deformations instead of vertex displacements. Hasler *et al.* [12] learn two linear models: one for pose and one for body shape. Loper *et al.* [13] introduce SMPL, a vertex-based linear model for human body shape and pose-dependent shape variation. This model, which has been extensively used in the literature, has also been adapted to other types of objects such as animals [14] and human body parts [15]. While these models can capture large variations, they exhibit two fundamental limitations. **First**, they rely on separate models for pose-independent shape, pose-dependent shape, and pose. Thus, they are limited to specific classes of objects, e.g., human bodies. Changing the target application, e.g., to animals [14] or infants [16], requires redefining the model. **Second**, they either assume a given registration between the surfaces of the 3D objects or solve for registration separately by matching vertices across the surfaces using an unrelated optimization criterion. To address this problem, some methods, e.g., [17], inspired by the minimum description length approach, jointly learn the statistical model and the registration of the 3D scans used for training.

Recently, there has been a growing interest in analyzing variability in 3D shape collections using tools from differential and Riemannian geometry [4], [5], [18], [19], [20], [21], [22]; see [23] for a detailed survey. The work most relevant to ours is the Square-Root Normal Field (SRNF) representation introduced in [5]. In this work, parameterized surfaces are compared using a partial elastic Riemannian metric defined as a weighted sum of a bending term and a stretching term. More importantly, Jermyn *et al.* [5] show that by carefully choosing the weights of these two terms, the complex partial elastic metric reduces to the $\mathbb{L}^2$ metric in the space of SRNFs. Thus, by treating shapes of objects as points in the SRNF space, a straight line between two points in this space is equivalent to the geodesic (or shortest) curve in the original space of surfaces under the partial elastic metric, and represents the optimal deformation between them. As a result, one can perform statistical analysis in the SRNF space using standard vector calculus, and then map the results back to the space of surfaces (for visualization), using the approach of Laga *et al.* [4]. Another important property of SRNFs is that both registration and optimal deformation (geodesic) are computed jointly, using the same partial elastic metric.

One of the fundamental problems in statistical shape analysis is correspondence and registration; see [24]. Past methods do not define a shape space and a metric that enable the computation of geodesics and statistics. Also, correspondence methods that are based on the intrinsic properties of surfaces, e.g., Generalized Multidimensional Scaling [25], spectral descriptors [26], or functional maps (which rely on the availability of descriptors) [27], [28], are primarily suited for surfaces that deform in an isometric manner. They also require landmarks to resolve symmetry ambiguities.

2.2 Statistical models for longitudinal shape data

As stated in [7], we live in a 4D world of 3D shapes in motion. With the availability of a variety of range sensing devices that can scan dynamic objects at high temporal frequency, there is a growing interest in capturing and modeling the 4D dynamics of objects [29], [30], [31]. For instance, Wand *et al.* [29] and Tevs *et al.* [31] propose methods to reconstruct the deforming geometry of time-varying point clouds. Li *et al.* [32] use sequences of 4D scans to learn a statistical 3D facial model. This model, referred to as FLAME, has been later used by Cudeiro *et al.* [6] to capture, learn, and synthesize 3D speaking styles. Bogo *et al.* [7] build a 4D human data set by registering a 3D human template to sequences of 3D human scans performing various types of actions. These methods focus on the 3D reconstruction of deforming objects. The literature on the statistical analysis of their spatiotemporal variability is rather limited.

Early works focused on longitudinal 2D shape data. For instance, Anirudh *et al.* [33] represent the contour of planar shapes that evolve as trajectories on a Grassmann manifold. They then use the Transported Square-Root Vector Field (TSRVF) representation for their rate-invariant analysis. This approach was later extended to the analysis of the trajectories of sparse features or landmarks measured on the surface of a deforming 3D object. Akhter *et al.* [34] introduced a bilinear spatiotemporal basis to model the spatiotemporal variability in 4D surfaces. The approach treats surfaces as N discrete landmarks and uses the $\mathbb{L}^2$ metric and PCA in $\mathbb{R}^{4N}$ for their analysis. Thus, the approch is not suitable for highly articulated shapes that undergo large articulated and elastic motion (e.g., human bodies). The approach also assumes that the landmarks are in correspondence, both spatially and temporally.

Anirudh *et al.* [33] and Ben Amor *et al.* [35] represent human body actions using dynamic skeletons. By treating each skeleton, represented by a set of landmarks, as a high-dimensional point on Kendall's shape space [36], motions become trajectories in a high-dimensional Euclidean space. Thus, one can use the rich literature on the statistical analysis of high-dimensional curves [37] to build a framework for the statistical analysis of human motions and actions. This approach, however, has two fundamental limitations. **First**, the $\mathbb{L}^2$ metric on Kendall's shape space is not suitable for large articulated motions. **Second**, skeletons and landmarks do not capture surface elasticity, and thus, cannot be used to model growth processes and surface deformations due to motion. While this can be addressed by using two separate models, one for shape and another for motion, it will fail to capture motion-dependent shape variations.

Using the LDDMM framework [38], Debavelaere *et al.* [39] and Bone *et al.* [40] represent a 4D surface as a flow of deformations of the 3D volume around each surface and then encode deformations as geodesics on a Riemannian manifold. However, in general, natural deformations do not correspond to geodesics but can be arbitrary paths on the shape space. Also, deforming 3D volumes is expensive in terms of computation and memory requirements. Finally, this approach relies on manually-specified landmarks to efficiently register the 3D volumes. Our approach, which can handle large articulated and elastic motions, works directly on surfaces, does not assume that deformations are (piecewise) geodesics, and does not rely on landmarks for the spatiotemporal registration.

3 Mathematical framework

In this section, we describe the proposed mathematical framework for the spatiotemporal registration and comparison of 4D surfaces. Section 4 discusses its application to various statistical analysis tasks. A 4D surface, where the fourth dimension refers to time, is a 3D surface that evolves over time. Examples of such 4D surfaces include facial expressions (e.g., a smiling face), a human body shape performing an action such as walking or jumping, or an anatomical organ that evolves over time due to natural growth or disease progression. A 4D surface can be represented as a path $\alpha(t)$, $t \in [0, 1]$ such that $\alpha(0)$ and $\alpha(1)$ are the initial and final surfaces, respectively, and $\alpha(t)$, $0 < t < 1$ are the intermediate surfaces. The main challenges posed by the statistical analysis of such 4D surfaces are two-fold. **First**, surfaces within the same 4D surface and across different 4D surfaces come with arbitrary poses and registrations. **Second**, 4D surfaces can have different execution rates, e.g., two smiling expressions performed at different speeds. Thus, to compare and perform statistical analysis on samples of 4D surfaces, we first need to spatiotemporally register them.

We solve the spatiotemporal registration problem using tools from differential geometry. We treat surfaces as points in a Riemannian shape space equipped with an elastic metric that captures shape differences using bending and stretching energies. We then formulate the elastic registration problem, *i.e.,* the problem of computing spatial correspondences, as that of finding the optimal rotation and reparameterization that align one surface onto another. This enables comparing and spatially registering surfaces, even in the presence of large elastic deformations (Section 3.1).

With this representation, a 4D surface becomes a time-parameterized trajectory in the above-referenced Riemannian shape space. Thus, the problem of analyzing 4D surfaces is reduced to the problem of analyzing curves. Similar to surfaces, we define a space of curves equipped with a Riemannian metric, which quantifies the amount of elastic deformation, or time warping, needed to align two 4D surfaces (Section 3.2).

3.1 The elastic shape space of surfaces

Fig. 1 overviews the proposed spatial registration framework. We consider a surface as a function f of the form:

$$f : \Omega \rightarrow \mathbb{R}^3; \quad s \mapsto f(s) = (X(s),\ Y(s),\ Z(s)), \tag{1}$$

where Ω is a parameterization domain and $s \in \Omega$ is the parameter in this domain. The choice of Ω depends on the nature of the surfaces of interest. When dealing with closed surfaces of genus-0, Ω is a sphere, *i.e.,* $\Omega = \mathbb{S}^2$, and $s = (u, v)$, where $u \in [0, \pi]$ and $v \in [0, 2\pi[$ are the spherical coordinates. In practice, surfaces come as unregistered triangular meshes, which we map to a spherical domain using the spherical parameterization algorithm of [41].

To remove shape-preserving transformations, we first translate the surfaces so that their center of mass is located at the origin, and then scale them to have unit surface area. The space of such normalized surfaces, denoted by $\mathcal{F}$, is called the *preshape space.*

Having removed translation and scale, we still need to account for rotations and reparameterizations. Those are handled algebraically. For any surface $f \in \mathcal{F}$ and for any rotation $O \in SO(3)$, Of and f have equivalent shapes. Similarly, any reparameterization of a surface with an orientation-preserving diffeomorphism preserves its shape. Let Γ be the space of all orientation-preserving diffeomorphisms of Ω. Then, $\forall\ \gamma \in \Gamma$, f and $f \circ \gamma$, *i.e.,* the reparameterization of f using γ, have the same shape. (Here, $\circ$ refers to the composition of two functions.) Note that reparameterizations provide dense correspondences across surfaces. If one wants to put a surface f_2 in correspondence with another surface f_1, then we need to find a rotation O^* and a reparameterization γ^* such that $O^*(f_2 \circ \gamma^*)$ is as close as possible to f_1. This is precisely the process of 3D surface registration. It is defined mathematically as:

$$\left(O^*,\ \gamma^*\right) = \underset{O \in SO(3),\, \gamma \in \Gamma}{\operatorname{argmin}} d_{\mathcal{F}}(f_1,\ O(f_2 \circ \gamma)), \tag{2}$$

where $d_{\mathcal{F}}$ is a distance in $\mathcal{F}$.

3.1.1 SRNF representation of surfaces—For efficient registration and comparison of surfaces, the distance measure, or metric, $d_{\mathcal{F}}$ should quantify interpretable shape differences, *i.e.,* the amount of bending and stretching one needs to apply to one surface to deform

it into another. It should also be simple enough to facilitate efficient computation of correspondences and geodesic paths. Jermyn *et al.* [5] introduced a partial elastic metric that measures differences between surfaces as a weighted sum of the amount of bending and stretching that one needs to apply to a surface to align it to another. In this approach, bending is measured in terms of changes in the orientation of the unit normal vectors, while stretching is measured in terms of changes in the infinitesimal surface areas. More importantly, Jermyn *et al.* [5] showed that by using a special representation of surfaces, called the Square-Root Normal Field (SRNF), the complex partial elastic metric reduces to the simple $\mathbb{L}^2$ metric on the SRNF space.

Definition 3.1 (Square-Root Normal Field (SRNF)).: The SRNF map $H(f)$ of a surface $f \in \mathcal{F}$ is defined as the normal vector field of the surface scaled by the square-root of the local area around each surface point:

$$\begin{aligned} H &: \mathcal{F} \to \mathcal{C}_h \\ f &\mapsto H(f) = h, \text{ such that } h(u, v) = \frac{\mathbf{n}(u, v)}{\|\mathbf{n}(u, v)\|_2^{\frac{1}{2}}}, \end{aligned} \tag{3}$$

where $\mathcal{C}_h$ is the space of all SRNFs, $\mathbf{n} = \frac{\partial f}{\partial u} \times \frac{\partial f}{\partial v}$ is the normal field to f and $\| \cdot \|_2$ is the Euclidean norm in $\mathbb{R}^3$.

The SNRF representation of surfaces has nice properties that make it suitable for the various analysis tasks at hand:

- It is translation invariant. Also, the SRNF of a rotated surface is simply the rotation of the SRNF of that surface, *i.e.,* $H(Of) = OH(f)$.
- $\forall \gamma \in \Gamma$, $H(f \circ \gamma) = \sqrt{|J_\gamma|}(h \circ \gamma) \equiv h * \gamma$, where J_γ is the Jacobian of γ and $|\cdot|$ is its determinant.
- Under the $\mathbb{L}^2$ metric on the space of SRNFs, the action of Γ is by isometries, *i.e.,* $\forall\ \gamma \in \Gamma$ and $\forall f_1, f_2 \in \mathcal{F}$, $\|h_1 - h_2\| = \|h_1 * \gamma - h_2 * \gamma\|$, where $h_i = H(f_i)$, $i = 1, 2$.
- The space of SRNFs is a subset of $\mathbb{L}^2(\Omega, \mathbb{R}^3)$. In addition, the $\mathbb{L}^2$ metric in $\mathcal{C}_h$ is equivalent to the partial elastic metric in the space of surfaces. As such, geodesics in $\mathcal{F}$ become straight lines in the SRNF space $\mathcal{C}_h$; see Fig. 1.
- Currently, there is no analytical expression for the inverse SRNF map, and in fact, the injectivity and surjectivity of the SRNF remain open questions. However, Laga *et al.* [4] showed that, for a given SRNF of a valid surface, one can always numerically estimate the original surface, up to translation [4].

The last three properties are critical for comparison and atlas construction of 4D surfaces. One can perform elastic registration of surfaces using the standard $\mathbb{L}^2$ metric in the space of SRNFs, which is computationally very efficient compared to using the complex elastic metric in the space of surfaces (Section 3.1.2). Further, temporal evolutions of surfaces can be interpreted as curves in the Euclidean space of SRNFs, making them amenable to statistical analysis. Thus, the problem of constructing 4D atlases becomes the problem of statistical analysis of elastic curves in the space of SRNFs using standard statistical tools developed for Euclidean spaces. After analysis, the results can be mapped back to the original space of surfaces using efficient SRNF inversion procedures [4] (Section 3.2).

3.1.2 Spatial elastic registration of surfaces—Under the SRNF representation, the elastic registration problem in Eqn. (2) can be reformulated using the $\mathbb{L}^2$ metric on $\mathscr{C}_h$, the space of SRNFs, instead of the complex partial elastic metric on the preshape space $\mathscr{F}$. Let f_1 and f_2 be two surfaces in the preshape space $\mathscr{F}$, and h_1 and h_2 their SRNFs. Then, the rotation and reparameterization that optimally register f_2 to f_1 are given by:

$$\left(O^*, \gamma^*\right) = \underset{O \in SO(3), \gamma \in \Gamma}{\operatorname{argmin}} \|h_1 - O(h_2 * \gamma)\|, \tag{4}$$

where * is the composition operator between an SRNF and a diffeomorphism $\gamma \in \Gamma$. This joint optimization over $SO(3)$ and Γ can be solved by alternating, until convergence, between the two marginal optimizations (this is allowed due to the product structure of $SO(3) \times \Gamma$) [42]:

- Assuming a fixed parameterization, solve for the optimal rotation using Procrustes analysis via Singular Value Decomposition (SVD).
- Assuming a fixed rotation, solve for the optimal reparameterization using a gradient descent algorithm.

To solve for the optimal reparameterization, we represent the space Γ of diffeomorphisms γ, which are functions on the sphere, using gradients of the spherical harmonic basis $\{B_i\}_{i=1,\dots,n}$. This way, every $\gamma \in \Gamma$ can be written as a weighted sum of the harmonic basis gradients: $\gamma = \sum_{i=1}^{n} a_i B_i$. Thus, the search for the optimal diffeomorphism is reduced to the search for the optimal weights $\{a_i\}$. This procedure is described in detail in Section 2.1 of the Supplementary Material.

Although this approach converges to a local optimum, in practice, it can be used in a very efficient way. Since a 4D surface α is a sequence of discrete realizations $\mathscr{F}$, $i = 0, \cdots, n$, with $t_0 = 0$ and $t_n = 1$, one can perform the elastic registration sequentially. Let $\beta = H(\alpha)$ be the SRNF map of the 4D surface α, *i.e.,* $\forall\ t \in [0, 1]$, $\beta(t) = H(\alpha(t))$. Also, let α_0 be a reference surface randomly chosen from the population of surfaces being analyzed, and β_0 its SRNF map (α_0 can be, for example, $\alpha(0)$). Then, the spatial registration procedure is as follows.

1. Find $O_0 \in SO(3)$ and $\gamma_0 \in \Gamma$ that register $\beta(t_0)$ (the start point of the SRNF path) to the SRNF of the reference surface β_0, by solving Eqn. (4).
2. For $i = 0, \dots, n$,
   - $\beta(t_i) \leftarrow O_0(\beta(t_i) * \gamma_0)$ and $\alpha(t_i) \leftarrow O_0 \alpha(t_i) \circ \gamma_0$.
3. For $i = 0, \dots, n$,
   - Find, by solving Eqn. (4), $O_i \in SO(3)$ and $\gamma_i \in \Gamma$ that register $\beta(t_i)$ to $\beta(t_{i-1})$.

- $\beta(t_i) \leftarrow O_i(\beta(t_i) * \gamma_i)$ and $\alpha(t_i) \leftarrow O_i\alpha(t_i) \circ \gamma_i$.

The first step ensures that, when given a collection of 4D surfaces α_j, $j = 1, \cdots, n$, the surfaces $\alpha_j(0)$, $j = 1, \cdots, n$ are registered to each other. The subsequent steps ensure that $\forall t$, $\alpha_j(t)$ is registered to $\alpha_j(0)$. This sequential approach is efficient since, in general, elastic deformations between two consecutive frames in a 4D surface are relatively small. In what follows, we assume that all surfaces within a 4D surface and across 4D surfaces are correctly registered, *i.e.,* they have been normalized for translation and scale, and optimally rotated and reparameterized using the approach described in this section.

3.2 The shape space of 4D surfaces

Under the setup of Section 3.1, a 4D surface becomes a curve $\alpha: [0, 1] \rightarrow \mathcal{F}$. However, since $\mathcal{F}$ is endowed with the partial elastic metric, which is non-Euclidean, we propose to further map the 4D surfaces to the SRNF space, which has a Euclidean structure. Thus, 4D surfaces become curves of the form $\beta: [0, 1] \rightarrow \mathcal{C}_h$. With this representation, all statistical tasks are carried out in $\mathcal{C}_h$ under the $\mathbb{L}^2$ metric with results mapped back to the space of surfaces $\mathcal{F}$ for visualization.

3.2.1 TSRVF representation of SRNF trajectories—Let α be a curve (path) in $\mathcal{F}$ and β its image under the SRNF map, *i.e.,* $\forall\ t \in [0, 1]$, $\beta(t) = H(\alpha(t))$; β is also a curve, but in $\mathcal{C}_h$. Let $\mathcal{M}_{\mathcal{F}}$ be the space of all paths in $\mathcal{F}$, and $\mathcal{M}_h$ be the space of all paths in $\mathcal{C}_h$: $\mathcal{M}_h = \{\beta: [0, 1] \rightarrow \mathcal{C}_h \mid \beta = H(\alpha),\ \alpha \in \mathcal{M}_{\mathcal{F}}\}$.

To temporally register, compare, and summarize samples of such curves, we need to define an appropriate metric on $\mathcal{M}_{\mathcal{F}}$, or $\mathcal{M}_h$, that is invariant to the rate (or speed) of the 4D surfaces. For example, facial expressions that only differ in the rate of their execution should be deemed equivalent under such a metric. Let $\Xi = \{\xi: [0, 1] \rightarrow [0, 1]$ such that $0 < \dot{\xi} < \infty$, $\xi(0) = 0$ and $\xi(1) = 1\}$ denote all reparameterizations of the temporal domain $[0, 1]$. Here, $\dot{\xi} = \frac{d\xi}{dt}$. Then, for any $\xi \in \Xi$, $\beta \circ \xi$ and β only differ in the rate of execution and are thus equivalent. The function ξ is often referred to as a time warping of the domain $[0, 1]$. Temporal registration of two 4D surfaces α_1 and α_2 then becomes the problem of registering their corresponding curves β_1 and β_2 in $\mathcal{C}_h$. This requires solving for an optimal reparameterization $\xi^* \in \Xi$ that minimizes an appropriate distance $d(\cdot,\cdot)$ between β_1 and β_2:

$$\xi^* = \underset{\xi \in \Xi}{\operatorname{argmin}}\ d(\beta_1,\ \beta_2 \circ \xi). \tag{5}$$

The optimization over Ξ in Eqn. (5) ensures rate invariance. Thus, we are left with defining a distance $d(\cdot,\cdot)$ that is invariant to time warping of the temporal domain $[0, 1]$. To this end, we borrow tools from Srivastava *et al.* [37] for analyzing shapes of curves in $\mathbb{R}^n$, $n \geq 2$. The associated elastic metric defined therein is invariant to reparameterizations of curves, and quantifies the amount of bending and stretching of the curves in terms of changes in the orientations and lengths of their tangent vectors, respectively. However, instead of directly working with such a complex elastic metric, Su *et al.* [43] introduced the Transported

Square-Root Vector Field (TSRVF) representation, which simplifies the complex elastic metric into the simple $\mathbb{L}^2$ metric.

Definition 3.2 (Transported Square-Root Vector Field (TSRVF)[1]).: For any smooth trajectory $\beta \in \mathscr{M}_h$, the transported square-root vector field (TSRVF) is a parallel transport of a scaled velocity vector field of β to a reference point $c \in \mathscr{C}_h$ according to

$$Q(\beta)(t) = q(t) = \frac{\dot{\beta}(t)\big|_{\beta(t) \to c}}{\sqrt{\|\dot{\beta}(t)\|}}, \tag{6}$$

where $\dot{\beta} = \frac{\partial \beta}{\partial t}$ is the tangent vector field on β and $\|\cdot\|$ is the $\mathbb{L}^2$ metric on $\mathscr{C}_h$.

Note that the parallel transport $\dot{\beta}(t)\big|_{\beta(t) \to c}$ is performed along the geodesic from $\beta(t)$ to c. The TSRVF representation has nice properties that facilitate efficient temporal registration of 4D surfaces. Let β_1 and β_2 be two trajectories on $\mathscr{M}_h$, and let q_1 and q_2 be their respective TSRVFs.

- The elastic metric on the space of trajectories $\mathscr{M}_h$ reduces to the $\mathbb{L}^2$ metric on the space of their TSRVFs. Thus, one can use the $\mathbb{L}^2$ metric to compare two paths:

$$d(\beta_1,\ \beta_2) = \|q_1 - q_2\| = \left(\int_0^1 \|q_1(t) - q_2(t)\|^2 dt\right)^{\frac{1}{2}}, \tag{7}$$

  where $\|\cdot\|$ is again the $\mathbb{L}^2$ norm on $\mathscr{C}_h$.

- For any $\xi \in \Xi$, $Q(\beta \circ \xi) = (q \circ \xi)\sqrt{\dot{\xi}(t)} \equiv q \odot \xi$.

- Under the $\mathbb{L}^2$ metric, the action of the reparameterization group Ξ on the space of TSRVFs is by isometries, *i.e.,* $\|q_1 - q_2\| = \|(q_1 \odot \xi) - (q_2 \odot \xi)\|$, $\forall\ \xi \in \Xi$.

- Given a TSRVF q and an initial trajectory point, one can reconstruct the corresponding path β, such that $Q(\beta) = q$, by solving an ordinary differential equation [43].

As we will see next, these properties enable efficient temporal registration of trajectories and subsequent rate-invariant statistical analysis. In what follows, let $\mathbb{Q}$ denote the space of TSRVFs equipped with the $\mathbb{L}^2$ metric defined in Eqn. 7.

3.2.2 Temporal registration—

Under the TSRVF representation, the temporal registration problem in Eqn. (5), which involved optimization over Ξ, can now be reformulated using the standard $\mathbb{L}^2$ metric on the space of TSRVFs:

$$\xi^* = \underset{\xi \in \Xi}{\operatorname{argmin}} \|q_1 - q_2 \odot \xi\|. \tag{8}$$

[1] Although, in this paper, we consider curves β in the space $\mathscr{M}_h$, TSRVFs are general and can be defined on any curves, e.g., curves in $\mathbb{R}^d$, $d \geq 1$.

This problem can be solved efficiently using a Dynamic Programming algorithm [43], [44]. Then, the rate-invariant distance $d(\beta_1, \beta_2)$ between two trajectories β_1 and β_2 is given by:

$$d(\beta_1,\ \beta_2) = \inf_{\xi \in \Xi} \|q_1 - q_2 \odot \xi\| \,. \tag{9}$$

3.2.3 Geodesics between 4D surfaces—Let $\alpha_1,\ \alpha_2 \in \mathscr{M}_{\mathscr{F}}$ be two 4D surfaces. The pipeline to spatiotemporally register them and compute the geodesic path between them can be summarized as follows.

(1) Proposed spatial registration.: The goal is to spatially register the surfaces in α_1 and α_2 to the same reference surface, which can be any arbitrary surface. For simplicity, we choose it to be $\alpha_1(0)$, the first surface in the sequence α_1. The spatial registration can then be performed in two steps:

- Compute the SRNF maps: $\forall\ t \in [0, 1]$, $\beta_1(t) = H(\alpha_1(t))$ and $\beta_2(t) = H(\alpha_2(t))$.
- Spatially register β_1 and β_2, and thus α_1 and α_2, to the reference surface, using the algorithm described in Section 3.1.2.

For simplicity of notation, we also use β_1 and β_2 to denote the spatially-registered trajectories.

(2) Proposed temporal alignment.: β_1 and β_2 are elements of $\mathscr{M}_h$. We perform temporal registration in three steps:

- Map β_1 and β_2 to the TSRVF space $\mathscr{Q}: q_1 = Q(\beta_1)$ and $q_2 = Q(\beta_2)$.
- Find ξ^*, the optimal reparameterization that registers q_2 to q_1 by solving Eqn. (8).
- $q_2^* \leftarrow q_2 \odot \xi^*$ and $\beta_2^* \leftarrow \beta_2 \circ \xi^*$.

(3) Proposed geodesic computation.: Since $\mathscr{Q}$ is Euclidean, the geodesic path Λ_q between q_1 and q_2^* is a straight line:

$$\Lambda_q(\tau) = (1 - \tau)q_1 + \tau q_2^*,\ \tau \in [0,\ 1] \,. \tag{10}$$

Next, we map Λ_q back to $\mathscr{M}_h$ using the inverse TSRVF map, *i.e.,* $\forall\ \tau$, $\Lambda_\beta(\tau) = Q^{-1}(\Lambda_q(\tau))$. The computation of the inverse mapping uses the starting point on the trajectory and has a closed-form solution, making it computationally efficient. This is described in detail in [43]. After applying the inverse mapping to the entire geodesic path, we have $\Lambda_\beta(0) = \beta_1$, $\Lambda_\beta(1) = \beta_2$, and $\beta_\tau = \Lambda_\beta(\tau)$, $\tau \in (0,1)$, *i.e.,* a geodesic path between the SRNF curves β_1 and β_2.

(4) Visualization.: To visualize geodesic paths between 4D surfaces (and not their SRNFs), we need to further map all SRNFs on the trajectory $\Lambda_\beta(\tau)$ to their corresponding surfaces in $\mathscr{F}$. This is done using the inverse SRNF map, *i.e.,* $\forall\ \tau \in [0, 1]$, $t \in [0, 1]$, $\Lambda(\tau)(t) = H^{-1}(\Lambda_\beta(\tau)(t))$. Unlike the TSRVF map whose inverse can be computed analytically, inversion of the SRNF map, whose injectivity and surjectivity are yet to be determined, has to be accomplished numerically using the approach of Laga *et al.* [4].

Now, Λ is the geodesic path between the 4D surfaces α_1 and α_2^*, *i.e.,* $\Lambda(1) = \alpha_2^*$, and $\alpha_\tau = \Lambda(\tau)$ is a 4D surface at time τ along the geodesic path. Fig. 3 shows an example of a geodesic between two 4D surfaces representing talking faces. Each row corresponds to one 4D surface. The top row is the source, the bottom row is the target after optimal spatiotemporal registration, and the highlighted row in the middle corresponds to the mean 4D surface. The temporal registration is further illustrated in Fig. 4, where we show the source 4D surface, the target 4D surface before the spatiotemporal registration, and the target 4D surface after the spatiotemporal registration. Section 5 provides more examples of geodesics computed between various types of 4D surfaces.

4 Statistical analysis of 4D surfaces

Now that we have devised all of the required mathematical tools for comparing 4D surfaces, we shift our focus to how these tools can be used to build a 4D atlas from a sample of 4D surfaces. Let $\alpha_1, \cdots, \alpha_n$ be a set of 4D surfaces and $\beta_1, \cdots, \beta_n$ be their corresponding trajectories in $\mathscr{C}_h$. We assume that all of the surfaces, and their corresponding SRNFs, have been spatially registered to a common reference; see Section 3.1.2. We proceed to map all of the 4D surfaces to their corresponding TSRVFs, hereinafter denoted by $q_1, \cdots, q_n$, and compute statistics in that space. As before, all results are mapped at the end to the original space of surfaces $\mathscr{F}$ for visualization. We will use this framework to compute means and modes of variation, and to synthesize novel 4D surfaces by sampling from probability distributions fitted to a set of exemplar 4D surfaces.

Mean of 4D surfaces.

Intuitively, the mean of a collection of 4D surfaces is the 4D surface that is as close as possible to all of the 4D surfaces in the collection, under the specified distance measure (or metric). It is also called the Karcher mean and is defined as the 4D surface that minimizes the sum of squared distances to all of the 4D surfaces in the given sample. In other words, we seek to solve the following optimization problem, defined in the space of TSRVFs:

$$\bar{q} = \underset{q \in \mathscr{Q}}{\operatorname{argmin}} \sum_{i=1}^{n} \min_{\xi_i \in \Xi} \left\| q - q_i \odot \xi_i \right\|^2 . \quad (11)$$

Algorithm 3 in the Supplementary Material describes the proposed procedure for solving this optimization problem. It outputs the TSRVF Karcher mean $\bar{q}$, the optimal temporal reparameterizations ξ_i^*, $i = 1, \ldots, n$, and the temporally registered TSRVFs $q_i^* = q_i \odot \xi_i^*$; again, for simplified notation we simply use ξ_i and q_i to denote the optimal temporal reparameterizations and the temporally registered TSRVFs. The mean 4D surface can be obtained by TSRVF inversion of the mean TSRVF followed by SRNF inversion [4].

Principal directions of variation.

Since the TSRVF space is Euclidean, the principal directions of variation can also be computed in a standard way, *i.e.,* using the Singular Value Decomposition (SVD) of the covariance matrix. In the following, we assume that the TSRVFs are sampled using a finite set of points and appropriately vectorized. Let $K = \frac{1}{n-1}\sum_i (q_i - \bar{q})(q_i - \bar{q})^\top$ be the covariance matrix of the input sample, σ_i, $i = 1, \ldots, k$ its k-leading eigenvalues, and Σ_i, $i = 1, \ldots, k$ the corresponding eigenvectors. Then, one can explore the variability in the i-th principal direction using $q_\tau = \bar{q} + \tau\sqrt{\sigma_i}\Sigma_i$, where $\tau \in \mathbb{R}$. To visualize this principal direction of variation, we again use TSRVF inversion followed by SRNF inversion to compute the 4D surface α_τ, such that $Q(H(\alpha_\tau)) = \bar{q} + \tau\sqrt{\sigma_i}\Sigma_i$, $\tau \in \mathbb{R}$.

Random 4D surface synthesis.

Given the mean and the k-leading principal directions of variation, any TSRVF q of a 4D surface α can be approximately represented, in a parameterized form, as:

$$q = \bar{q} + \sum_{i=1}^{k} \tau_i\sqrt{\sigma_i}\Sigma_i, \tau_i \in \mathbb{R}. \tag{12}$$

Thus, to generate a random TSRVF, we only need to generate k random values $\tau_i \in \mathbb{R}$ and plug them into Eqn. (12). Then, to compute the corresponding random 4D surface, we apply the inverse TSRVF map followed by the inverse SRNF map. Also, by enforcing each τ_i to be within a certain range, e.g., $[-1, 1]$, we can ensure that the generated random 4D surfaces are similar to the given samples and thus plausible.

This procedure allows the generation of new random 4D surfaces. However, it does not offer any control over the generation process, which is entirely random. In many situations, we would like to control this process using a set of parameters. For instance, when dealing with 4D facial expressions, these parameters can be the degree of sadness, facial dimensions, etc. This type of control can be implemented using regression in the TSRVF space, a problem that we plan to explore in the future.

5 Results

This section demonstrates some results of the proposed framework and evaluates its performance. Section 5.1 focuses on spatiotemporal registration and geodesic computation between 4D surfaces. Section 5.2 focuses on the computation of statistical summaries while Section 5.3 focuses on the random synthesis of 4D surfaces. Finally, Section 5.4 provides an ablation study to demonstrate the importance of each component of the proposed framework. We use three data sets: **(1)** VOCA [6], which contains 4D facial scans, captured at 60fps, of 12 subjects speaking various sentences; **(2)** MPI DFAUST [7], which contains high-resolution 4D scans of 10 human subjects in motion, captured at 60fps, with a total of 129 dynamic performances; and **(3)** MPI 4D CAPE [8], which contains high-resolution 4D scans of 10 male and 5 female subjects in clothing. These data sets come as polygonal meshes with consistent triangulation and given registration across the meshes. We spherically parameterize them using Kurtek *et al.*'s implementation [18] of the spherical parameterization approach of [41]. We also apply randomly generated spatial diffeomorphisms to simulate non-registered surfaces. Our framework does not use, either explicitly or implicitly, the provided vertex-wise correspondences.

5.1 Spatiotemporal registration and 4D geodesics

We consider pairs of 4D facial expressions from the VOCA dataset. We first reparameterize each 4D surface using randomly generated time-warping functions to simulate facial expressions performed at different execution rates. We then apply the framework proposed in this paper to spatiotemporally register them. Fig. 4 shows an example of such spatiotemporal registration. In this example, we show **(a)** the source 4D surface, **(b)** the target 4D surface before spatiotemporal registration, and **(c)** the target 4D surface after spatiotemporally registering it to the source. We also highlight some key frames. As one can see, the original 4D surfaces differ significantly in their execution rates. The proposed spatiotemporal registration framework synchronizes the source and target expressions, thus enabling their comparison, interpolation and averaging. We also perform a similar experiment on the human body shapes in the DFAUST [7] and CAPE [8] data sets; see Figs. 5, 6, and 12(a)–(c). Compared to faces, human body shapes are very challenging to analyze since they perform complex articulated motions, which result in large bending and stretching of their surfaces.

4D geodesics.—Fig. 7 shows geodesics between 4D human body shapes. In this example, both the source and the target perform a punching action but at different rates. We show the geodesic before and after the spatiotemporal registration of the target 4D surface onto the source. Unlike the jumping action in Fig. 5, the left hand of the target surface does not perform the same action as the left hand of the source surface. Nevertheless, our framework can bring these two 4D surfaces as close as possible to each other. The Supplementary Material includes a video of the sequence and more examples of geodesics between 4D faces (from VOCA), 4D human bodies (from DFAUST), and clothed 4D human bodies (from CAPE).

Evaluation of the spatial registration.—We quantitatively evaluate the accuracy of the proposed spatial registration method and compare it to the latest functional map-based techniques such as MapTree [45] and Fast Sinkhorn filters [46]. Similar to our method, functional maps operate on clean manifold surfaces and do not use any form of (deep) learning. We take the surfaces of COMA [47], CAPE [8], and DFAUST [7] data sets, which come with ground-truth correspondences, and apply random spatial diffeomorphisms to them to simulate unregistered surfaces. We then compute the correspondence map between each pair of surfaces in the dataset. We measure the spatial registration error in terms of the geodesic distance, on the parameterization domain, between the ground-truth and the computed correspondence. Table 1 reports the mean, standard deviation, and median of the registration errors computed across all of the models in each data set. As one can see, the proposed SRNF-based spatial registration method significantly outperforms state-of-the-art algorithms [45], [46]. We refer the reader to the Supplementary Material, which includes visual examples of pairs of surfaces before and after spatial registration. It also includes additional spatial registration experiments using the quadruped animal data set of Kulkarni *et al.* [48].

An important property of the proposed approach is that it finds a one-to-one mapping between the source and target surfaces. This is not the case with functional map-based methods, which can map a point on the source to multiple points on the target. Thus, they cannot be used to compute geodesics and statistical summaries.

Evaluation of the temporal registration.—We perform an ablation study in which we evaluate the contribution of each component of the proposed temporal registration framework. We use the FLAME fitting framework [32] to generate random 4D facial surfaces with known ground-truth temporal registrations. We first generate two random SMPL [13] parameters, each corresponding to a 3D surface, and then linearly interpolate them to simulate a deforming 4D facial surface. Let α_i, $i \in \{1, ..., 5\}$ be the resulting 4D surfaces. Next, we generate 100 random temporal diffeomorphisms ξ_i; see Fig. 12(b) in the Supplementary Material.

Now, given a pair of 4D surfaces α_i and α_j, and for each pair of temporal diffeomorphisms ξ_k and ξ_l, $\alpha_i \circ \xi_k$ and $\alpha_j \circ \xi_l$ can be seen as a pair of 4D surfaces with different execution rates. We then compute the distance between:

- the perfectly registered 4D surfaces α_i and α_j; see the green curves in the plots (a) to (e) in Fig. 8. Note that, since these 4D surfaces correspond to different subjects performing different animations, then the original distance between them is not 0, but it is the lower bound.
- the perturbed 4D surfaces, *i.e.,* $\alpha_i \circ \xi_k$ and $\alpha_j \circ \xi_l$ before temporal registration; see Fig. 8(a);
- the registered 4D surfaces but without SRNF and without TSRVF; see Fig. 8(b). For this, we use a dynamic programming-based time warping algorithm;
- the registered 4D surfaces, without SRNF but with TSRVF; see Fig. 8(c);
- the registered 4D surfaces, with SRNF but without TSRVF; see Fig. 8(d);
- the registered 4D surfaces using the full framework, *i.e.,* with SRNF and with TSRVF; see Fig. 8(e).

Figs. 8(a) to (e) report statistics of these errors for each of the 20 pairs of 4D surfaces, but aggregated over the 100 random temporal diffeomorphisms. As one can see, the median distance between the 4D surfaces after registration using the full pipeline (Fig. 8(e)) is significantly lower than the ablated methods. The former is significantly closer to the ground-truth shown with green curves in Fig. 8 than the ablated methods.

Computation time.—Our approach runs entirely on the CPU. The Matlab implementation of the spatiotemporal registration process takes less than 31.43 seconds on 4.2 GHz Intel Core i7 with 32 GB of RAM. The visualization, which is needed when computing geodesics, means, and directions of variation, and when synthesizing random 4D surfaces, relies on the inversion of the SRNF maps. It requires 6 seconds per frame and a total of 30 minutes for the 300 temporal frames used in this paper. All of the experiments were performed using a spherical resolution of 256×256 for the DFAUST and CAPE data sets, and of 128×128 for the COMA data set.

5.2 Summary statistics

We now consider a set of unregistered 4D surfaces and compute their mean and principal directions of variation. Fig. 9 shows the 4D mean (highlighted with a blue box) computed from six 4D human shapes performing different types of actions. The figure also shows the input 4D surfaces after their spatiotemporal registration; see the video in the Supplementary Material for an illustration of the input 4D surfaces before spatiotemporal registration. Despite the large articulated motion, the large differences in the type of actions, and the significant differences in the execution rates of the 4D surfaces, our framework is able to co-register them and generate a plausible average 4D surface.

Fig. 10(a) and (b), on the other hand, show the mean and the first two principal directions of variation computed on input 4D facial surfaces. As we can see, the computed mean captures the main features of the dataset. The principal directions of variation further capture relevant variability in the given data. The Supplementary Material includes the input 4D surfaces prior to their registration. Please also refer to the videos in the Supplementary Material for additional results.

5.3 4D surface synthesis

Fig. 10(c) shows five 4D facial expressions randomly sampled from a Gaussian distribution with parameters estimated from the VOCA data set using the method described in Sec. 4. To ensure that the synthesized 4D surfaces are plausible, we only consider those that are within 1.5 standard deviations along each principal direction of variation. We refer the reader to the Supplementary Material for videos of all of the randomly generated 4D surfaces. The ability to synthesize novel 4D surfaces can benefit many applications in computer vision and graphics. It can be used to augment data sets for efficient training of deep learning models.

5.4 Ablation study

We undertake an ablation study to demonstrate the importance of each component of the proposed framework.

Importance of the SRNF representation.—In this experiment (Fig. 11), we take two challenging 3D human body models, which undergo a large articulated motion, perform their spatial registration using the proposed SRNF approach, and then compute their statistical mean using the $\mathbb{L}^2$ metric **(1)** in the original surface space (Fig. 11(a)) and **(2)** in the SRNF space (Fig. 11(b)). Fig. 11(a) shows that the articulated parts of the mean computed in the original surface space unnaturally shrink. This is predictable since, under the $\mathbb{L}^2$ metric, geodesics correspond to straight lines. However, in the SRNF space, the $\mathbb{L}^2$ metric is equivalent to the optimal bending and stretching of the surfaces. Thus, the computed mean is more natural; see Fig. 11(b).

Next, we consider two full 4D surfaces of deforming human body shapes (Fig. 12(a) and (b)) and show their mean 4D surface obtained:

- with the SRNF representation, with spatial registration, and **without** temporal registration (Fig. 12(d)),
- with the SRNF representation, with spatial registration, and with temporal registration (Fig. 12(e)),
- **without** SRNF representation, with spatial registration, and **without** temporal registration (Fig. 12(f)), and
- **without** SRNF representation, with spatial registration, and with temporal registration (Fig. 12(g)).

The last two cases are equivalent to a linear interpolation in the original surface space, after spatial registration. In all cases, we perform the spatial registration using the SRNF framework.

First, we can see that the temporally-aligned target 4D surface (Fig. 12(c)) is very close to the source 4D surface in Fig. 12(a). We observe that the right hands became fully synchronized. As such, the mean 4D surface obtained after temporal registration (Fig. 12(e)) is fully synchronized with the source and the aligned target, unlike the mean 4D surface in Fig. 12(d), which has been obtained without temporal registration. **Second**, in the mean 4D surfaces obtained without the SRNF framework (Figs. 12(f) and (g)), we can observe that the parts that undergo large articulated motion (e.g., the arms) unnaturally shrink. This shrinkage is stronger in Fig. 12(f) since the mean is obtained without temporal registration. The bottom row of Fig. 12 shows a zoom-in on the time frame highlighted in Fig. 12(a) to (g).

Finally, we quantitatively evaluate the importance of the SRNF representation by comparing the expressive power of PCA on the original space of spatially registered surfaces and on the space of SRNFs. We randomly divide a data set into a training set and a testing set, using an 80% – 20% split. We then fit a PCA model to the training set (both in the original space and in the space of SRNFs), project each surface in the test set onto the PCA model, reconstruct it, and measure the error between the original and the reconstructed surfaces. Let $P = \{p_1, \ldots, p_n\}$ be the set of points on the original surface (after centering the surface to its center of mass), and $Q = \{q_1, \ldots, q_n\}$ the corresponding points on the reconstructed surface (after centering the reconstructed surface to its center of mass). We measure the reconstruction error as:

$$E = \frac{1}{n}\left(\frac{\sum \|p_i - q_i\|_2^2}{\sum \|p_i\|_2^2}\right)^{0.5}. \quad (13)$$

Note that while the normalization in Equation (13) is not necessary, it allows comparison of the reconstruction errors on the same scale. We perform 10-fold cross-validation. Table 2 reports the mean, median, and standard deviation of the error over the test set and averaged over the 10 runs. In this experiment, we use 95% of the cumulative energy, *i.e.,* the ratio between the sum of the eigenvalues of the selected leading eigenvectors and the sum of all of the eigenvalues is 95%. As one can see, PCA on the SRNF space has a significantly lower reconstruction error than PCA in the original space of surfaces. This demonstrates that the former is more suitable to characterize variability in the shape of 3D objects that bend and stretch. Section 5 in the Supplementary Material provides an in-depth analysis of the expressive power of PCA in the SRNF space.

Importance of the TSRVF representation for 4D surfaces.

We perform a similar ablation study, but on 4D surfaces, to compare the expressive power of PCA on the original space of curves and on the space of TSRVFs. Table 3 shows that PCA error on the TSRVF space is lower than the error in the original space. We compute the reconstruction error using Equation (13), but this time the points are sampled from the curves. This experiment demonstrates that the former is more suitable to characterize variability in 4D surfaces. In this experiment, we use 95% of the cumulative energy, *i.e.,* the ratio between the sum of the eigenvalues of the selected leading eigenvectors and the sum of all of the eigenvalues is higher or equal to 95%. Section 5 in the Supplementary Material provides a more detailed analysis.

6 Conclusion

We have proposed a new framework for the statistical analysis of longitudinal 3D shape data (or 4D surfaces, *i.e.,* surfaces that deform over time), e.g., 3D human body shapes performing actions at different execution rates or 3D human faces pronouncing sentences at different speeds. Unlike traditional techniques, which only consider how features such as landmarks or measurements vary over time, the proposed framework considers the deformation of the entire surface of a 3D object. Our key contribution is in representing 4D surfaces as trajectories in the space of SRNFs, and the use of Transported Square-Root Vector Fields to analyze such trajectories statistically. The proposed framework can spatiotemporally register 4D surfaces, even in the presence of large elastic deformations and significant variation in the execution rates. It is also able to compute geodesics and summary statistics, which in turn can be used to randomly synthesize new, unseen 4D surfaces.

In contrast to SMPL-based representations, which are specialized for human body shapes, this paper's focus is on generic statistical models that are applicable to a wide range of object classes. Comparing generic statistical models vs. specialized ones such as SMPL (and its variants) is a very important problem, which requires an in-depth analysis that is well beyond the scope of this paper. We note that generic models can always be applied to specific classes of objects. However, each class of objects would require its specialised SMPL, e.g., one for faces, one for hands, one for the body, one for animals, etc.. They all lie on different subspaces and thus cannot be used for inter-class analysis. This, however, is not the case with our SRNF-based representation. In fact, although we have demonstrated the proposed 4D analysis framework on human body shapes and facial surfaces, it is general and can be applied to other classes of surfaces. Note also that our current implementation is limited to surfaces that are homeomorphic to a sphere, but we plan to extend the framework to higher-genus surfaces by exploring different parameterization methods, including mesh-based representations [49]. Also, the approach uses the numerical SRNF inversion procedure of Laga *et al.* [4], which is sometimes not accurate near the poles of the parameterization domain; we plan to improve its performance via the use of charts.

The framework deals with surfaces that bend and stretch but do not change in topology; as such, it does not apply to tree-like shapes, e.g., botanical trees or roots. However, the concept of representing deformations as trajectories in a shape space also applies to tree-shape spaces such as those used in [50], [51]. The framework is also limited to clean surfaces that are free of geometric and topological noise; as such, the proposed spatial registration method cannot be used to register partial scans to each other, or to register a template to partial scans. However, similar to statistical shape models such as 3D morphable models and SMPL, the proposed 4D atlas can be used as a prior; in conjunction with a data generation model, it can thereby be applied to noisy or partial data, e.g., to reconstruct entire 4D surfaces. The statistical analysis presented in this paper assumes that the population of the 4D surfaces follows a Gaussian distribution. We plan to extend the approach to other types of distributions, e.g., Gaussian Mixture Models, which can represent populations that follow multimodal distributions.

The proposed framework has various applications in computer vision, graphics, biology, and medicine. In computer vision, collecting large animations to train deep neural networks, e.g., for 3D reconstruction or action recognition [52], [53], is complex and time-consuming. Our framework can contribute to solving this problem by automatically synthesizing new samples from a small dataset. Our current implementation has only considered random synthesis, which is very important for populating virtual environments and for data augmentation to train deep learning networks. However, there are many situations where we would like to control this process using a set of parameters. For instance, when dealing with 4D facial expressions, these parameters can be the degree of sadness, facial dimensions, etc. This type of control can be implemented efficiently using regression in the TSRVF space. Finally, our framework can be used to statistically analyze how anatomical organs deform due to growth or disease progression.

The code for **(1)** the spherical parameterization, **(2)** the spatial registration of genus-0 surfaces using SRNFs, **(3)** the SRNF map inversion, and **(4)** the temporal registration of high-dimensional curves are available by request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement.

We would like to thank **(1)** the reviewers and the associate editor for their feedback and comments during the review process, and **(2)** the authors of [6], [7], [8], [12], [47] for making their data sets publicly available, and [45] and [46] for sharing the codes of MapTree and the Fast Sinkhorn Filters-based surface registration. This work is supported by the Australian Research Council Discovery Grants DP210101682 and DP220102197, Engineering and Physical Sciences Research Council UK (grant no. EP/V048104/1), and by NSF CCF 1740761, NSF CCF 1839252, NSF DMS 2015226, and NIH R37 CA214955.

Biographies

Hamid Laga is currently an Associate Professor at Murdoch University (Australia). His research interests span various fields of machine learning, computer vision, computer graphics, and pattern recognition, with a special focus on the 3D reconstruction, modeling and analysis of static and deformable 3D objects, and on machine learning for agriculture and health. He is the recipient of the Best Paper Awards at SGP2017, DICTA2012, and SMI2006.

Marcel Padilla is a doctoral researcher at the Technical University of Berlin. He works with the Discretization in Geometry and Dynamics research group under the supervision of Peter Schroder (Caltech) and Ulrich Pinkall (TU Berlin) through an Einstein visiting fellowship. His interest is the interplay between physical simulations, geometry processing, discrete exterior calculus and applications in graphics and animation.

Ian H. Jermyn Ian H. Jermyn is Professor of Statistics in the Department of Mathematical Sciences at Durham University. His research interests centre around geometrical questions in statistics, on methodologies for the representation and statistical modelling of geometric structure, and on the geometrical underpinnings essential to the success of statistical shape analysis. Methodologies include the Riemannian geometry, applied to the description of shape spaces, and random fields with long-range and higher-order interactions, used in particular to model objects with complex and varying topologies. He has also worked substantially on the application of these methodologies, in particular to image processing and computer vision.

Sebastian Kurtek is an Associate Professor in the Department of Statistics at The Ohio State University. He received his MS and PhD degrees in Biostatistics from Florida State University in 2009 and 2012, respectively. His research focuses on statistical modeling for shape and functional data and has been supported by multiple National Science Foundation and National Institutes of Health grants. He is a Senior Member of the IEEE.

Mohammed Bennamoun is Winthrop Professor at the University of Western Australia (UWA) and is a researcher in computer vision, machine/deep learning, robotics, and signal/speech processing. He has published 4 books, 1 edited book, 1 Encyclopedia article, 14 book chapters, 150+ journal papers, and 260+ conference publications. He was awarded 70+ competitive research grants, from the Australian Research Council, and numerous other Government, UWA and industry Research Grants. He delivered conference tutorials at major conferences, including: CVPR 2016, Interspeech 2014, IEEE ICASSP, and ECCV. He was also invited to give a Tutorial at an International Summer School on Deep Learning (2017).

Anuj Srivastava is a Professor of Statistics at the Florida State University in Tallahassee, FL. He obtained his MS and PhD degrees in Electrical Engineering from the Washington University in St. Louis in 1993 and 1996, respectively. After spending the year 1996–97 at the Brown University as a visiting researcher, he joined FSU as an Assistant Professor in 1997. His research is focused on pattern theoretic approaches to problems in image analysis, computer vision, and signal processing. Specifically, he has developed computational tools for performing statistical inferences on certain nonlinear manifolds. He is a Fellow of IEEE and IAPR.

References

[1]. Werner P, Lopez-Martinez D, Walter S, Al-Hamadi A, Gruss S, and Picard R, “Automatic recognition methods supporting pain assessment: A survey,” IEEE Transa. Affective Computing, 2019.

[2]. Egger B, Smith WA, Tewari A, Wuhrer S, Zollhoefer M, Beeler T, Bernard F, Bolkart T, Kortylewski A, Romdhani S et al. , "3D Morphable Face Models: Past, Present, and Future," ACM TOG, vol. 39, no. 5, pp. 1–38, 2020.

[3]. Allen B, Curless B, and Popović Z, "The space of human body shapes: Reconstruction and parameterization from range scans," ACM Transactions on Graphics, vol. 22, no. 3, pp. 587–594, Jul. 2003.

[4]. Laga H, Xie Q, Jermyn IH, and Srivastava A, "Numerical inversion of srnf maps for elastic shape analysis of genus-zero surfaces," IEEE PAMI, vol. 39, no. 12, pp. 2451–2464, 2017.

[5]. Jermyn I, Kurtek S, Klassen E, and Srivastava A, "Elastic shape matching of parameterized surfaces using square root normal fields," ECCV, vol. 5, no. 14, pp. 805–817, 2012.

[6]. Cudeiro D, Bolkart T, Laidlaw C, Ranjan A, and Black M, "Capture, learning, and synthesis of 3D speaking styles," IEEE CVPR, pp. 10 101–10 111, 2019. [Online]. Available: http://voca.is.tue.mpg.de/

[7]. Bogo F, Romero J, Pons-Moll G, and Black MJ, "Dynamic FAUST: Registering human bodies in motion," in IEEE CVPR, 2017.

[8]. Ma Q, Yang J, Ranjan A, Pujades S, Pons-Moll G, Tang S, and Black MJ, "Learning to dress 3d people in generative clothing," in IEEE CVPR, 2020, pp. 6469–6478.

[9]. Cootes TF, Taylor CJ, Cooper DH, and Graham J, "Active shape models-their training and application," Computer vision and image understanding, vol. 61, no. 1, pp. 38–59, 1995.

[10]. Blanz V and Vetter T, "A morphable model for the synthesis of 3d faces," in ACM Siggraph, 1999, pp. 187–194.

[11]. Anguelov D, Srinivasan P, Koller D, Thrun S, Rodgers J, and Davis J, "Scape: shape completion and animation of people," in ACM SIGGRAPH 2005 Papers, 2005, pp. 408–416.

[12]. Hasler N, Stoll C, Sunkel M, Rosenhahn B, and Seidel H-P, "A statistical model of human pose and body shape," in CGF, vol. 28, no. 2, 2009, pp. 337–346.

[13]. Loper M, Mahmood N, Romero J, Pons-Moll G, and Black MJ, "Smpl: A skinned multi-person linear model," ACM TOG, vol. 34, no. 6, pp. 1–16, 2015.

[14]. Zuffi S, Kanazawa A, Berger-Wolf T, and Black MJ, "Three-d safari: Learning to estimate zebra pose, shape, and texture from images," in IEEE CVPR, 2019, pp. 5359–5368.

[15]. Pavlakos G, Choutas V, Ghorbani N, Bolkart T, Osman AA, Tzionas D, and Black MJ, "Expressive body capture: 3d hands, face, and body from a single image," in IEEE CVPR, 2019, pp. 10 975–10 985.

[16]. Hesse N, Pujades S, Romero J, Black MJ, Bodensteiner C, Arens M, Hofmann UG, Tacke U, Hadders-Algra M et al. , "Learning an infant body model from rgb-d data for accurate full body motion analysis," in MICCAI, 2018, pp. 792–800.

[17]. Bolkart T and Wuhrer S, "A groupwise multilinear correspondence optimization for 3d faces," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 3604–3612.

[18]. Kurtek S, Srivastava A, Klassen E, and Laga H, "Landmark-guided elastic shape analysis of spherically-parameterized surfaces," CGF, vol. 32, no. 2pt4, pp. 429–438, 2013.

[19]. Kilian M, Mitra NJ, and Pottmann H, "Geometric modeling in shape space," in ACM SIGGRAPH, 2007.

[20]. Xie Q, Kurtek S, Le H, and Srivastava A, "Parallel transport of deformations in shape space of elastic surfaces," in IEEE ICCV, December 2013.

[21]. Xie Q, Jermyn I, Kurtek S, and Srivastava A, "Numerical inversion of srnfs for efficient elastic shape analysis of star-shaped objects," in ECCV, 2014, pp. 485–499.

[22]. Jermyn IH, Kurtek S, Laga H, and Srivastava A, "Elastic shape analysis of three-dimensional objects," Synthesis Lectures on Computer Vision, vol. 12, no. 1, pp. 1–185, 2017.

[23]. Laga H, "A survey on nonrigid 3D shape analysis," in Academic Press Library in Signal Processing, Volume 6. Elsevier, 2018, pp. 261–304.

[24]. Laga H, Guo Y, Tabia H, Fisher RB, and Bennamoun M, 3D Shape analysis: fundamentals, theory, and applications. John Wiley & Sons, 2018.

[25]. Bronstein AM, Bronstein MM, and Kimmel R, “Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching,” Proceedings of the National Academy of Sciences, vol. 103, no. 5, pp. 1168–1172, 2006.

[26]. Litman R and Bronstein AM, “Learning spectral descriptors for deformable shape correspondence,” IEEE PAMI, vol. 36, no. 1, pp. 171–180, 2013.

[27]. Ovsjanikov M, Ben-Chen M, Solomon J, Butscher A, and Guibas L, “Functional maps: a flexible representation of maps between shapes,” ACM TOG, vol. 31, no. 4, pp. 1–11, 2012.

[28]. Ovsjanikov M, Corman E, Bronstein M, Rodolà E, Ben-Chen M, Guibas L, Chazal F, and Bronstein A, “Computing and processing correspondences with functional maps,” in SIGGRAPH ASIA 2016 Courses, 2016, pp. 1–60.

[29]. Wand M, Jenke P, Huang Q, Bokeloh M, Guibas L, and Schilling A, “Reconstruction of deforming geometry from time-varying point clouds,” in SGP, 2007, pp. 49–58.

[30]. Beeler T, Hahn F, Bradley D, Bickel B, Beardsley P, Gotsman C, Sumner RW, and Gross M, “High-quality passive facial performance capture using anchor frames,” in ACM SIGGRAPH, 2011, pp. 1–10.

[31]. Tevs A, Berner A, Wand M, Ihrke I, Bokeloh M, Kerber J, and Seidel H-P, “Animation cartography?intrinsic reconstruction of shape and motion,” ACM TOG, vol. 31, no. 2, pp. 1–15, 2012.

[32]. Li T, Bolkart T, Black MJ, Li H, and Romero J, “Learning a model of facial shape and expression from 4d scans.” ACM Trans. Graph, vol. 36, no. 6, pp. 194–1, 2017.

[33]. Anirudh R, Turaga P, Su J, and Srivastava A, “Elastic functional coding of human actions: From vector-fields to latent variables,” in IEEE CVPR, 2015, pp. 3147–3155.

[34]. Akhter I, Simon T, Khan S, Matthews I, and Sheikh Y, “Bilinear spatiotemporal basis models,” ACM Transactions on Graphics (TOG), vol. 31, no. 2, pp. 1–12, 2012.

[35]. Amor BB, Su J, and Srivastava A, “Action recognition using rate-invariant analysis of skeletal shape trajectories,” IEEE PAMI, vol. 38, no. 1, pp. 1–13, 2015.

[36]. Dryden I and Mardia K, Statistical Shape Analysis. John Wiley & Son, 1998.

[37]. Srivastava A, Klassen E, Joshi S, and Jermyn I, “Shape analysis of elastic curves in euclidean spaces,” IEEE PAMI, no. 99, pp. 1–1, 2011.

[38]. Beg MF, Miller MI, Trouvé A, and Younes L, “Computing large deformation metric mappings via geodesic flows of diffeomorphisms,” IJCV, vol. 61, no. 2, pp. 139–157, 2005.

[39]. Debavelaere V, Durrleman S, Allassonnière S, and Initiative ADN, “Learning the Clustering of Longitudinal Shape Data Sets into a Mixture of Independent or Branching Trajectories,” IJCV, pp. 1–16, 2020.

[40]. Bône A, Colliot O, and Durrleman S, “Learning the Spatiotemporal Variability in Longitudinal Shape Data Sets,” IJCV, vol. 128, no. 12, pp. 2873–2896, 2020.

[41]. Praun E and Hoppe H, “Spherical parametrization and remeshing,” vol. 22, 2003, pp. 340–349.

[42]. Kurtek S, Klassen E, Ding Z, Jacobson S, Jacobson J, Avison M, and Srivastava A, “Parameterization-invariant shape comparisons of anatomical surfaces,” Medical Imaging, IEEE Transactions on, vol. 30, no. 3, pp. 849–858, 2011.

[43]. Su J, Kurtek S, Klassen E, Srivastava A et al. , “Statistical analysis of trajectories on Riemannian manifolds: bird migration, hurricane tracking and video surveillance,” The Annals of Applied Statistics, vol. 8, no. 1, pp. 530–552, 2014.

[44]. Robinson DT, Functional data analysis and partial shape matching in the square root velocity framework. The Florida State University, 2012.

[45]. Ren J, Melzi S, Ovsjanikov M, and Wonka P, “Maptree: recovering multiple solutions in the space of maps,” ACM Transactions on Graphics (TOG), vol. 39, no. 6, pp. 1–17, 2020.

[46]. Pai G, Ren J, Melzi S, Wonka P, and Ovsjanikov M, “Fast sinkhorn filters: Using matrix scaling for non-rigid shape correspondence with functional maps,” in CVPR, 2021.

[47]. Ranjan A, Bolkart T, Sanyal S, and Black MJ, “Generating 3D faces using convolutional mesh autoencoders,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 704–720.

[48]. Kulkarni N, Gupta A, Fouhey DF, and Tulsiani S, "Articulation-aware canonical surface mapping," in IEEE CVPR, 2020, pp. 452–461.

[49]. Bauer M, Charon N, Harms P, and Hsieh H-W, "A numerical framework for elastic surface matching, comparison, and interpolation," International Journal of Computer Vision, pp. 1–20, 2021.

[50]. Wang G, Laga H, Xie N, Jia J, and Tabia H, "The shape space of 3d botanical tree models," ACM TOG, vol. 37, no. 1, pp. 1–18, 2018.

[51]. Wang G, Laga H, Jia J, Miklavcic SJ, and Srivastava A, "Statistical analysis and modeling of the geometry and topology of plant roots," J. of Theoretical Biology, vol. 486, p. 110108, 2020.

[52]. Han X, Laga H, and Bennamoun M, "Image-based 3d object reconstruction: State-of-the-art and trends in the deep learning era," IEEE PAMI, 2019.

[53]. Laga H, Jospin LV, Boussaid F, and Bennamoun M, "A survey on deep learning techniques for stereo-based depth estimation," IEEE PAMI, 2020.

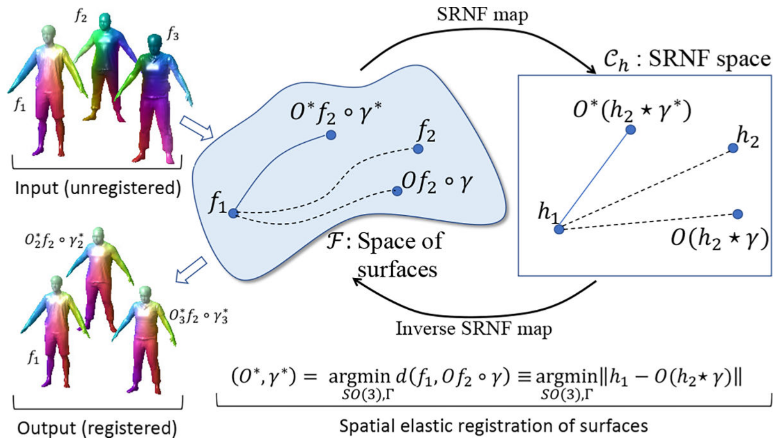


Fig. 1.

Overview of the proposed spatial registration framework. Surfaces are first mapped onto the space of Square-Root Normal Fields (SRNF) and spatially registered using the $\mathbb{L}^2$ metric, which is equivalent to the partial elastic metric in the original space of surfaces. 4D surfaces can then be treated as curves embedded in the $\mathbb{L}^2$ space of SRNFs. The operator $\star$ refers to the composition of functions in the SRNF space.

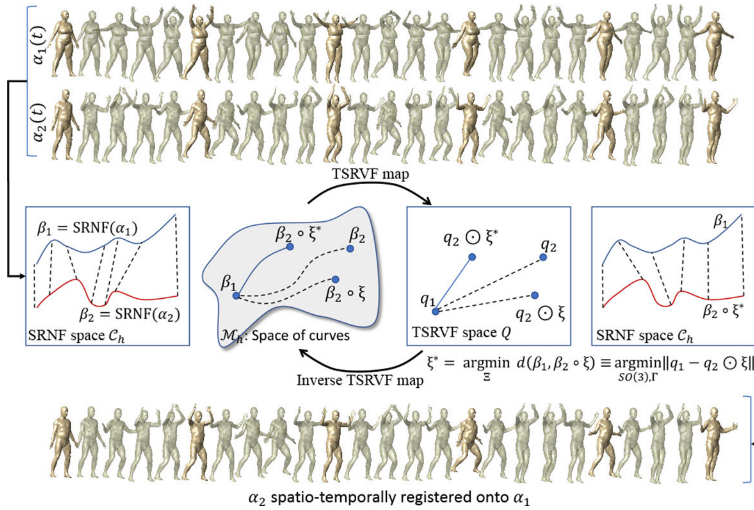


Fig. 2.

In the proposed temporal registration framework, 4D surfaces, represented as curves in the SRNF space, are first mapped to the space of Transported Square-Root Vector Fields (TSRVFs) for their temporal registration. Points in the TSRVF space are mapped back to the space of SRNFs and then to the original space of surfaces for visualization. The operator $\odot$ refers to the composition of functions in the TSRVF space.

Fig. 3.

Example of a geodesic between the source 4D surface (top row) and the target 4D surface (bottom row) after spatiotemporal registration. The highlighted row corresponds to the mean 4D surface. A video of the figure is included in the Supplementary Material.

Fig. 4.

Examples of the spatiotemporal registration of two facial expressions (4D faces). In each example, we show **(a)** the source 4D face, **(b)** the target 4D face, and **(c)** the target 4D face after spatiotemporal registration using the proposed framework. Note how the spatiotemporally registered target 4D surface became fully synchronised with the source 4D surface. The full video sequence is provided in the Supplementary Material.

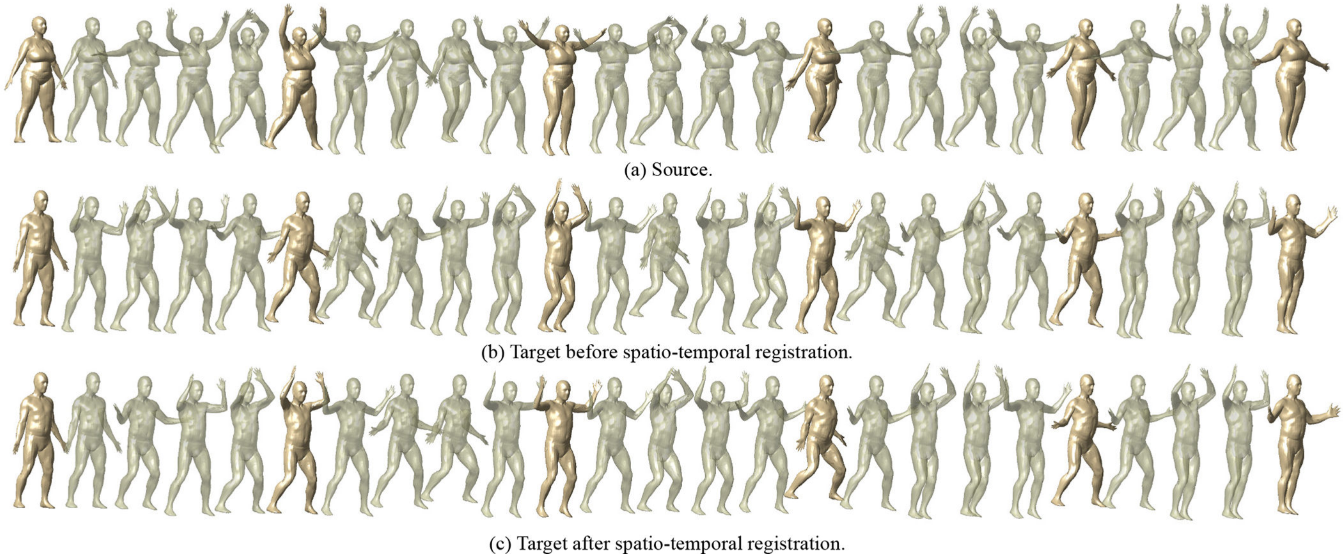


Fig. 5.
Example of the spatiotemporal registration, using the proposed algorithm, of two 4D human body shapes (from the DFAUST dataset) performing a jumping action at different speeds. Note how the spatiotemporally registered target 4D surface in (c) became synchronised with the source 4D surface in (a). The full video sequence is provided in the Supplementary Material.

Author Manuscript
Author Manuscript
Author Manuscript

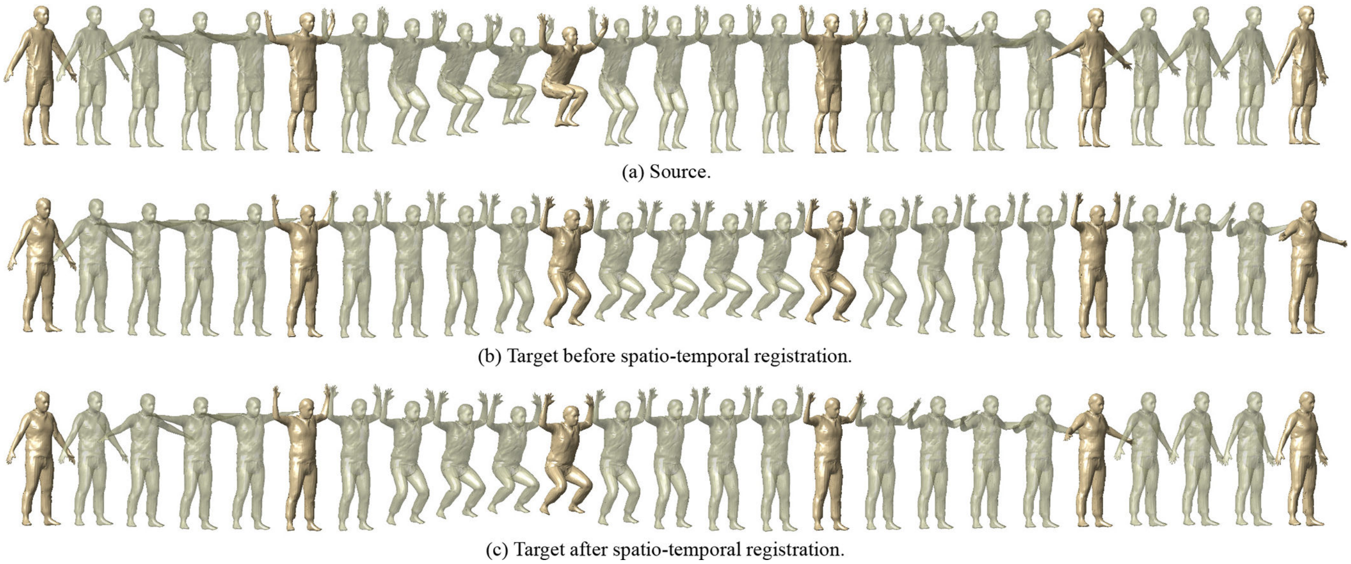
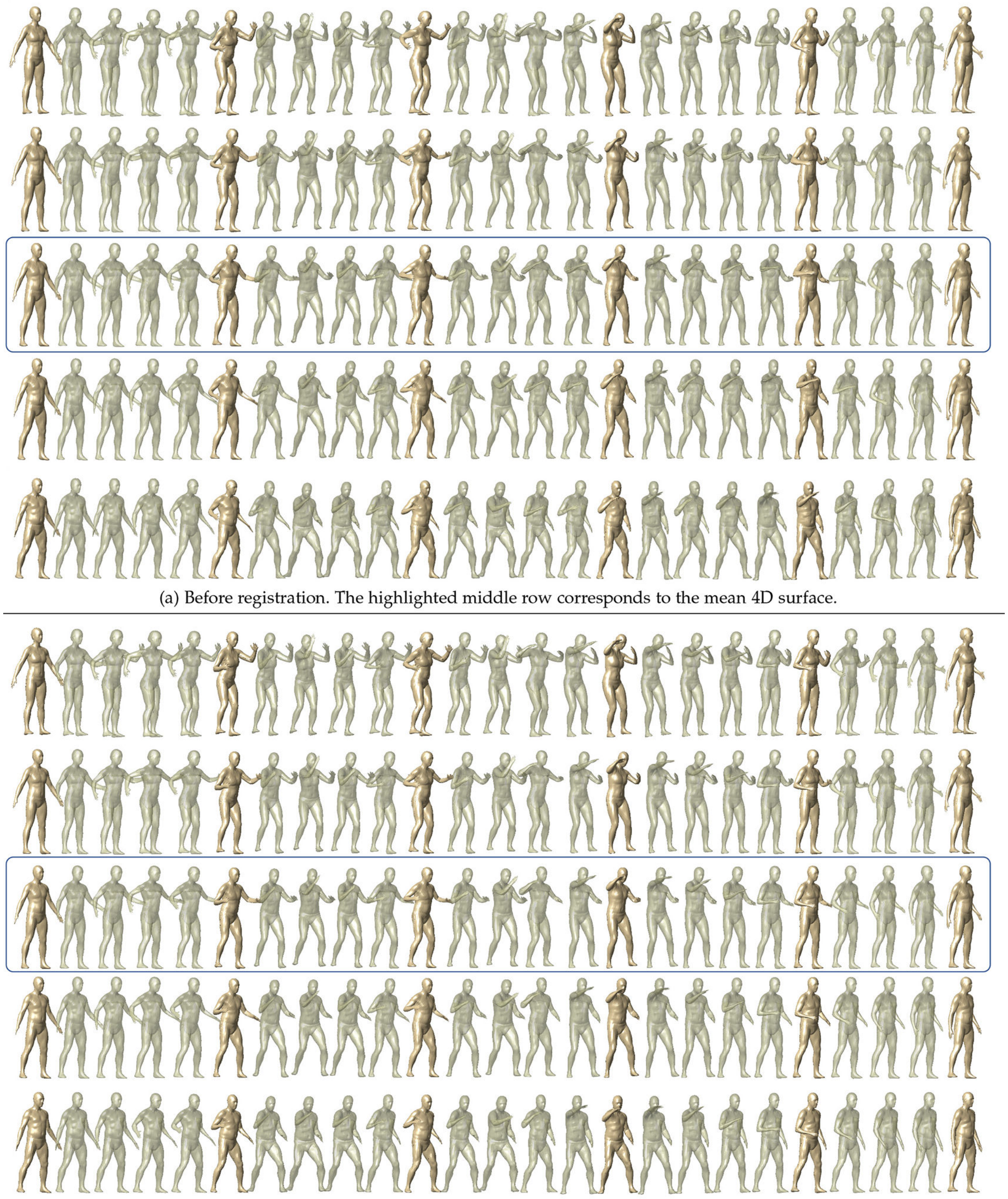


Fig. 6.

Example of the spatiotemporal registration, using the proposed algorithm, of two 4D body shapes with different clothing (from the CAPE dataset). Note how the spatiotemporally registered target 4D surface in (c) became fully synchronised with the source 4D surface in (a). The full video sequence is provided in the Supplementary Material.

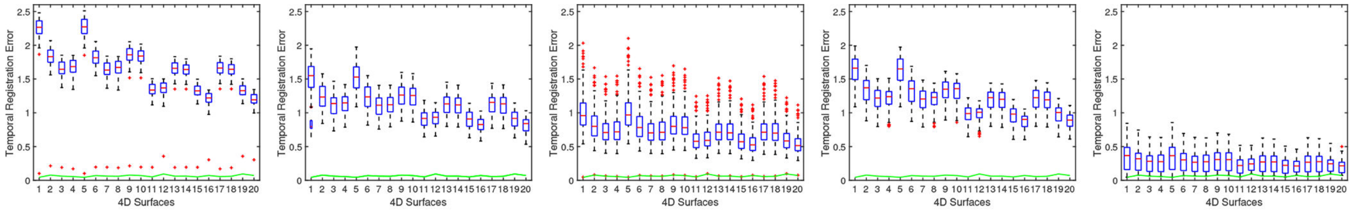


(a) Before registration. The highlighted middle row corresponds to the mean 4D surface.

(b) After registration. The highlighted middle row corresponds to the mean 4D surface.

Fig. 7.

Example of a geodesic between 4D surfaces corresponding to punching actions: **(a)** the 4D surfaces before registration and **(b)** after registration. In each example, we show the source 4D surface in the first row, the target 4D surface in the last row, and three intermediate 4D surfaces along the geodesic between the source and the target. Observe how misaligned are the highlighted frames before registration, and how synchronised they became after registration. A video illustrating these sequences is included in the Supplementary Material.



(a) Original no registration. (b) Registration no SRNF, no TSRVF. (c) Registration no SRNF, with TSRVF. (d) Registration with SRNF, no TSRVF. (e) Registration with SRNF and TSRVF.

Fig. 8.

Boxplots of errors between 20 pairs of 4D surfaces. In all of the plots, the red lines represent the median error and the boxes represent its spread. The green curve is the ground-truth distance, *i.e.,* the distance between the perfectly registered 4D surfaces. **(a)** Unregistered 4D surfaces generated using 100 random diffeomorphic transformations, **(b)** temporally registered surfaces, using dynamic programming-based time warping, without SRNF and without TSRVF, **(c)** temporally registered surfaces using TSRVF but without SRNF, **(d)** temporally registered surfaces using SRNF but without TSRVF, and **(e)** temporally registered surfaces using the full framework, *i.e.,* using TSRVF in the space of SRNFs.

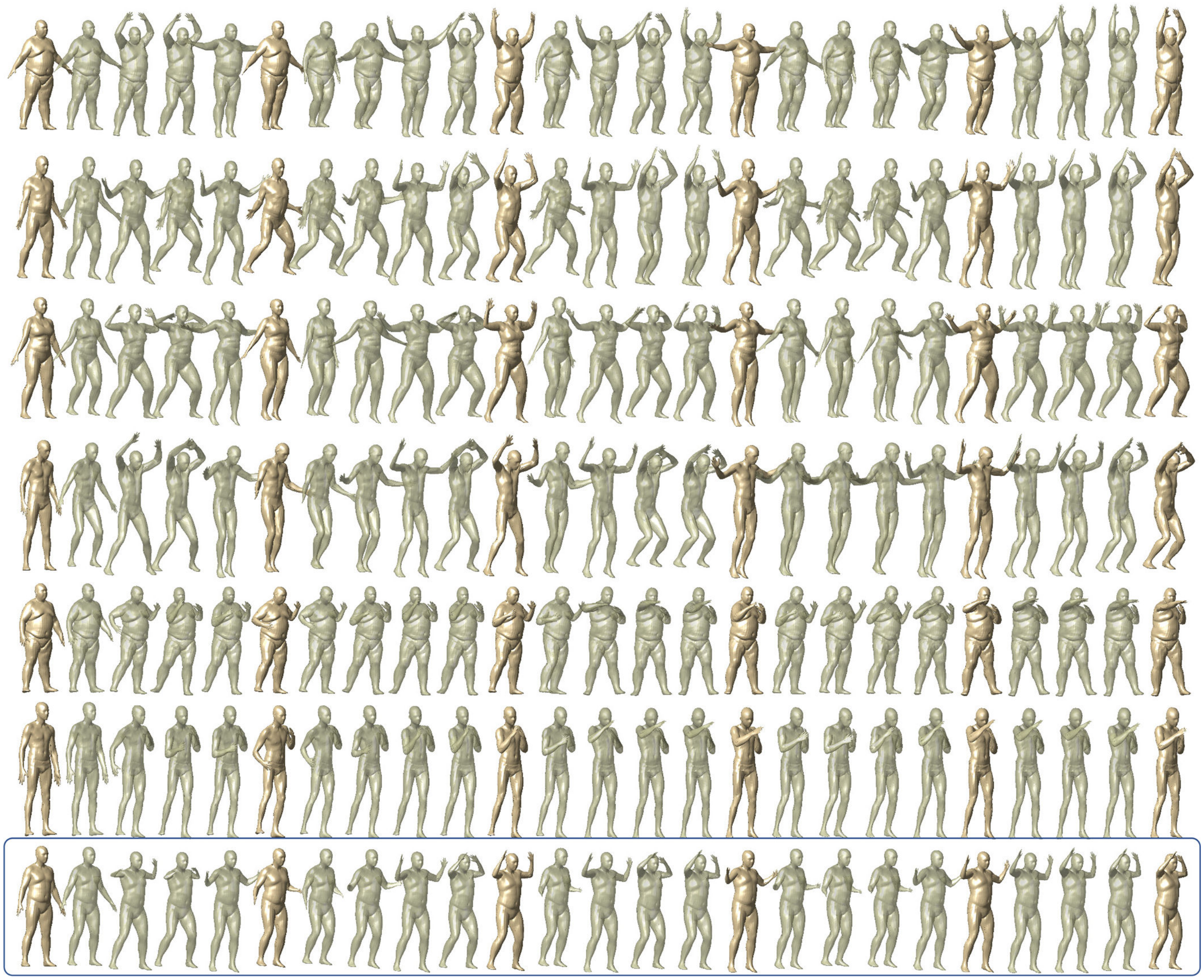
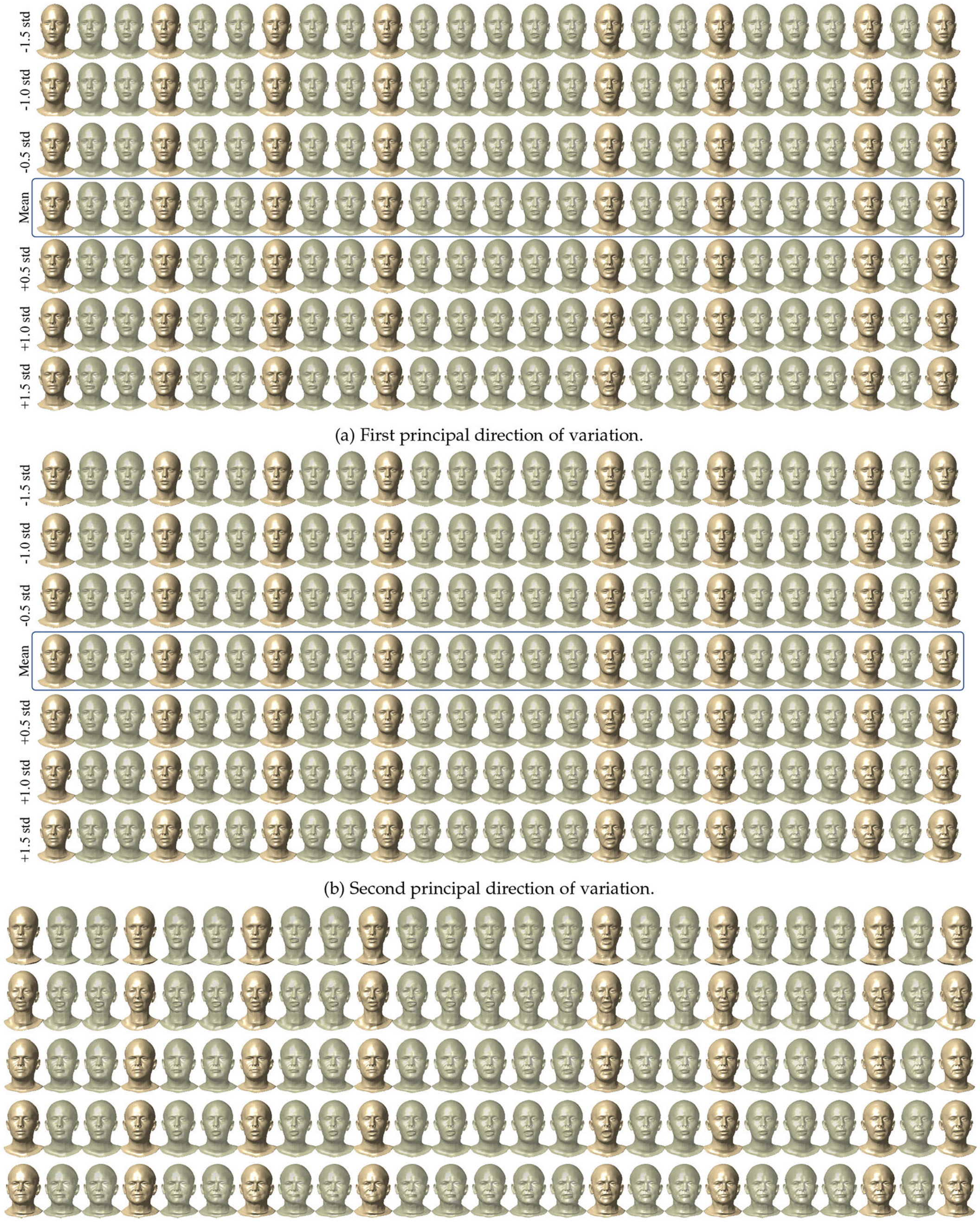


Fig. 9.

Co-registration of multiple 4D surfaces. In this example, we consider four human body shapes performing a jumping action (first four rows) and two others performing a punching action (rows 5 and 6). Here, we show the spatiotemporally co-registered 4D surfaces and the 4D mean (last row) computed using the proposed algorithm. The Supplementary Material includes the input 4D surfaces before their spatiotemporal registration. It also includes the full video sequences. The surfaces are from the DFAUST dataset.



(a) First principal direction of variation.

(b) Second principal direction of variation.

(c) Five randomly synthesized 4D faces.

Fig. 10.

First (a) and second (b) principal directions of variation (the mean 4D surface is highlighted in the middle). Each row corresponds to one 4D surface sampled between −1.5 to 1.5 times the standard deviation along the principal direction of variation. We refer the reader to the Supplementary Material and Video, which show the input 4D faces (before their spatiotemporal registration) and the animated sequences of the modes of variation as well as the randomly synthesized faces. The Supplementary Material and Video also include more modes of variation and randomly synthesized samples.

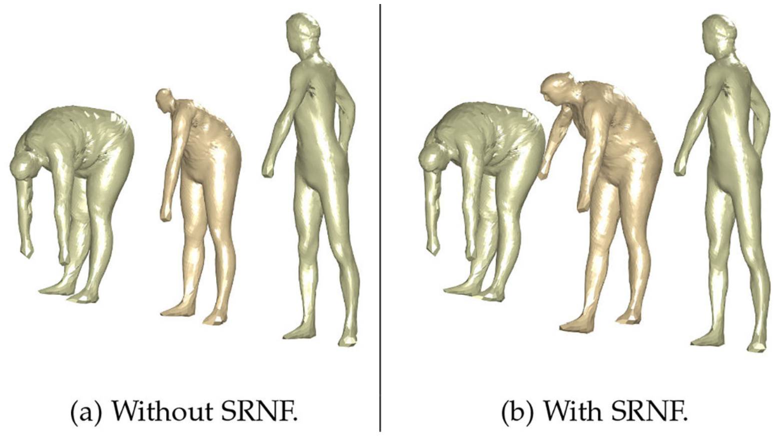


Fig. 11.

The mean shape between the left and right surfaces, computed **(a)** in the original surface space without the SRNF representation, and **(b)** in the SRNF space. In (a), the mean shape is distorted due to the use of the $\mathbb{L}^2$ metric in the original space of surfaces. In both cases, the spatial registration is performed using the proposed registration method.

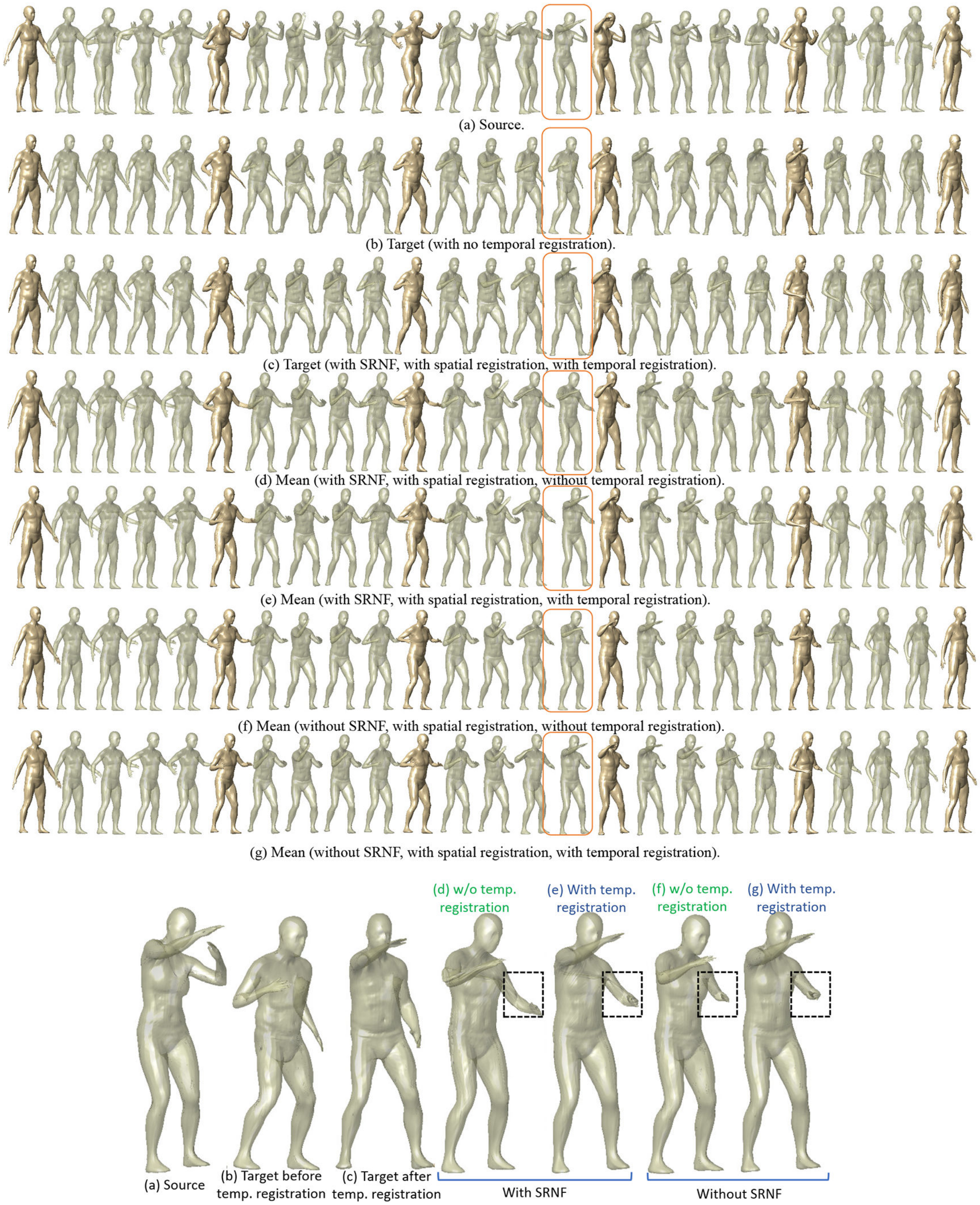


Fig. 12.

Illustration of the effect of the different components of the proposed framework on the quality of the computed mean 4D surface, which is the middle point along the geodesic between the source and target 4D surfaces. The bottom row is a zoom-in on the frame highlighted in (a) to (g). The 4D surfaces are from the DFAUST dataset. A video illustrating these sequences is included in the Supplementary Material.

TABLE 1

Comparison of the accuracy of the proposed spatial registration with state-of-the-art techniques such as MAP Tree [45] and Fast Sinkhorn filters [46], which are based on functional maps [28]. The computation time is in seconds. We use spherical maps of size 128 × 128 for COMA and 256 × 256 for CAPE and DFAUST.

	COMA [47]				CAPE [8]				DFAUST [7]			
	Mean	**Std**	**Median**	**Time**	**Mean**	**Std**	**Median**	**Time**	**Mean**	**Std**	**Median**	**Time**
MapTree [45]	1.6042	0.6956	1.6069	10.75	1.4447	0.7227	1.4483	9.59	1.5344	0.7065	1.5211	9.39
ICP-NN [46]	1.6028	0.6973	1.6056	4.13	1.4684	0.7087	1.5116	3.49	1.5185	0.6905	1.5082	3.65
ICP-Sinkhorn [46]	1.6019	0.6974	1.6053	7.05	1.4684	0.7037	1.5058	6.28	1.5275	0.6888	1.5227	7.03
Zoomout-NN [46]	1.5997	0.6902	1.6002	**2.49**	1.4743	0.7029	1.4781	**2.38**	1.5101	0.6922	1.4857	**2.67**
Zoomout-Sinkhorn [46]	1.6016	0.6908	1.5968	4.84	1.4737	0.6937	1.4925	4.66	1.5019	0.6948	1.4731	4.93
SRNF (ours)	**0.0012**	**0.0008**	**0.0003**	15.42	**0.0008**	**0.0008**	**0.0006**	25.32	**0.0008**	**0.0008**	**0.0007**	26.13

TABLE 2

Comparison of the expressive power of PCA on the original space of surfaces and on the space of SRNFs. The lower the values are, the better. These results are obtained using 95% of the cumulative energy.

	PCA on surfaces			**PCA on SRNFs**		
	Mean	**Std**	**Median**	**Mean**	**Std**	**Median**
DFAUST	0.083	0.022	0.083	**0.047**	0.012	**0.046**
VOCA	0.031	0.011	0.029	**0.009**	0.004	**0.008**
CAPE	0.098	0.035	0.097	**0.052**	0.019	**0.049**

TABLE 3

Comparison of the expressive power of PCA on the original space of curves and on the space of TSRVFs. The lower the values are, the better. These results are obtained using 95% of the cumulative energy; see Section 5 in the Supplementary Material.

	PCA on curves			**PCA on TSRVFs**		
	Mean	**Std**	**Median**	**Mean**	**Std**	**Median**
DFAUST	0.770	0.123	0.758	**0.602**	0.051	**0.617**
VOCA	0.690	0.127	0.662	**0.486**	0.318	**0.603**
CAPE	0.787	0.123	0.800	**0.500**	0.152	**0.471**