A Riemannian Framework for Analysis of Human Body Surface

Emery Pierson, Mohamed Daoudi, Alice-Barbara Tumpach

A Riemannian Framework for Analysis of Human Body Surface

Emery Pierson
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
emery.pierson@univ-lille.fr

Mohamed Daoudi
IMT Lille Douai, Institut Mines-Télécom, Univ. Lille, Centre for Digital Systems, F-59000 Lille, France
Univ. Lille, CNRS, Centrale Lille, Institut Mines-Télécom, UMR 9189 CRIStAL, F-59000 Lille, France
mohamed.daoudi@imt-lille-douai.fr

Alice-Barbara Tumpach
Univ. Lille, CNRS, UMR 8524 Laboratoire Painlevé, F-59000 Lille, France
Wolfgang Pauli Institut, Vienna, Austria
alice-barbora.tumpach@univ-lille.fr

Abstract

We propose a novel framework for comparing 3D human shapes under the change of shape and pose. This problem is challenging since 3D human shapes vary significantly across subjects and body postures. We solve this problem by using a Riemannian approach. Our core contribution is the mapping of the human body surface to the space of metrics and normals. We equip this space with a family of Riemannian metrics, called Ebin (or DeWitt) metrics. We treat a human body surface as a point in a "shape space" equipped with a family of Riemmanian metrics. The family of metrics is invariant under rigid motions and reparametrizations; hence it induces a metric on the "shape space" of surfaces. Using the alignment of human bodies with a given template, we show that this family of metrics allows us to distinguish the changes in shape and pose. The proposed framework has several advantages. First, we define family of metrics with desired invariant properties for the comparison of human shape. Second, we present an efficient framework to compute geodesic paths between human shape given the chosen metric. Third, this framework provides some basic tools for statistical shape analysis of human body surfaces. Finally, we demonstrate the utility of the proposed framework in pose and shape retrieval of human body.

1. Introduction

Human shape analysis is an important area of research with a wide applications in vision, graphics, virtual reality, product design and avatar creation. In this paper, we seek a framework for human shape analysis which provides: (i) a shape metric, for pairwise quantification of shape differences; (ii) generating deformations; and (iii) a shape summary, a compact representation of human shapes in terms of the center (mean of human shapes). Human bodies vary significantly across subjects and body postures. These variations make human body shape analysis a challenging problem.

2. Related Work

Past literature has tackled this fundamental issue with a varying degree of success, depending on the ultimate application. The main tasks in human shape analysis can be divided into representing, comparing, deforming and summarizing human shapes. For representing and comparing human bodies, a common theme in the literature has been to represent human surfaces by certain geometrical features, such as HKS [24], WKS [1] and ShapeDNA [21]. These methods provide reasonable results in classification and clustering of human shapes, but they do not study deformations nor develop statistical analysis of shapes [22], [25], [13]. Our approach falls within the class of elastic shapes analysis. In this section we cover methods from this family that are more closely related to ours, and refer to recent surveys [20] and [16] for an extensive review and comparison of such descriptors.

A particularly elegant mathematical approach to the problem of comparing surfaces is to consider the quotient of the space of embeddings of a fixed surface S into $\mathbb{R}^3$ by the action of the orientation-preserving diffeomorphisms of S and the group of Euclidean transformations, and provide

this quotient with the structure of an infinite-dimensional manifold. We can then define and use Riemannian metrics on this manifold to measure the distance between two given shapes as well as to interpolate between them by computing a geodesic that joins them ([25], [13]). However, the computational costs of this approach are high. Another recent approach is that of *square root normal fields* or SRNF in which different embeddings and immersions of the surface S modulo translations are described by points in a Hilbert space, and both rotations in $\mathbb{R}^3$ as well as reparametrizations of the surfaces translate into orthogonal transformations in the Hilbert space ([11]). However, the SRNF map is neither injective nor surjective and there exist different shapes having the same SRNF. In addition, as observed by Su el al. [22], the resulting distance can be viewed as an extrinsic distance obtained by embedding the space of parametrized surfaces in a linear space. However, one can observe that the articulation of the human body enables it to adopt a great variety of poses with very small changes to the intrinsic geometry (lengths of curves, angles, and areas) of the surface that models it and big changes in shape may lead to small change in extrinsic geometry.

Most of the above approaches use a spherical parameterization of 3D objects, while we propose in this paper to use a human template as a parametrization, and take some advantages of the recent developments of static and dynamic human datasets such as SMPL and FAUST.

2.1. Main Contributions

In this paper, we present a comprehensive Riemannian framework for analyzing human bodies, in the process of dealing with the change in shape and pose. Unlike some past works, instead of using a general parameterization of human body surfaces, we propose to use a human template and to align the human surfaces to this template. The human body surface is represented by the normal and the induced surface metric. Using the metric on the space of normals and the Ebin metric on the space of Riemannian metrics, a family of metrics is proposed to compare a human body shape and pose. To our best knowledge, this is the first demonstration of the use of this metric in human body shape analysis. We will show also for the first time, that this family of metrics takes into account the intrinsic and extrinsic geometry of human bodies. Additionally, we present an efficient framework to compute geodesic between given human body surfaces under the chosen metric. We provide some basic tools for statistical shape analysis of human body surfaces. These tools help us to compute an average human body. To evaluate our approach, we conduct extensive experiments on multiple datasets. The experimental results show that the proposed family of Riemannian metrics classifies correctly the shapes and the poses. The experimental results show also that our proposed framework provides better geodesics than the state-of-the-art Riemannian framework.

3. Mathematical Framework and Background

3.1. Notation

Given a reference human being $\mathcal{T}$ (also called a template in the sequel), we will represent a human shape S with an embedding $f : \mathcal{T} \to \mathbb{R}^3$ such that the image $f(\mathcal{T})$ equals S. The map f is an embedding onto a human shape $f(\mathcal{T})$. The function f is also called a *correspondence* between the template $\mathcal{T}$ and the human shape $f(\mathcal{T})$.

Recall that a map $f : \mathcal{T} \to \mathbb{R}^3$ is an *embedding* when: (1) f is smooth, in particular small variations on the template $\mathcal{T}$ correspond to small variations on the human shape $f(\mathcal{T})$ (2) f is an immersion, i.e. at each point of the human shape $f(\mathcal{T})$ one can define the normal (resp. tangent) space to the surface of the human body as subspace of $\mathbb{R}^3$, and (3) f is an homeomorphism onto its image, i.e. points on $f(\mathcal{T})$ that look close in $\mathbb{R}^3$ are images of close points in $\mathcal{T}$. We define the space of all registered human shapes as

$$\mathcal{H} := \{f : \mathcal{T} \to \mathbb{R}^3, f \text{ is an embedding}\}.$$

It is often called the *pre-shape space* since human bodies with the same shape but different correspondences with the template may correspond to different points in $\mathcal{H}$. The set $\mathcal{H}$ is a manifold, as an open subset of the linear space $\mathcal{C}^\infty(\mathcal{T}, \mathbb{R}^3)$ of smooth functions from $\mathcal{T}$ to $\mathbb{R}^3$. The tangent space to $\mathcal{H}$ at f, denoted by $T_f\mathcal{H}$, is therefore just $\mathcal{C}^\infty(\mathcal{T}, \mathbb{R}^3)$.

The shape preserving transformations can be expressed as group actions on $\mathcal{H}$. The group $\mathbb{R}^3$ with addition as group operation acts on $\mathcal{H}$, by *translations* : $(v, f) \mapsto f + v$, for $v \in \mathbb{R}^3$ and $f \in \mathcal{H}$. The group $\mathrm{SO}(3)$ with matrix multiplication as group operation acts on $\mathcal{H}$, by *rotations* : $(O, f) \mapsto Of$, for $O \in \mathrm{SO}(3)$ and $f \in \mathcal{H}$. Finally, the group $\Gamma := \mathrm{Diff}^+(\mathcal{T})$ consisting of diffeomorphisms which preserve the orientation of $\mathcal{T}$ acts also on $\mathcal{H}$, by *reparameterization* : $(\gamma, f) \mapsto f \circ \gamma^{-1}$, for $\gamma \in \mathrm{Diff}^+(\mathcal{T})$ and $f \in \mathcal{H}$. The use of γ^{-1}, instead of γ, ensures that the action is from left and, since the action of $\mathrm{SO}(3)$ is also from left, one can form a joint action of $G := \mathrm{Diff}^+(\mathcal{T}) \times \mathrm{SO}(3)$ on $\mathcal{H}$. In this paper, the translation group is taken care of by using a translation-independant metric. Therefore, in the following we will focus only on the reparameterization group Γ and on the rotation group $\mathrm{SO}(3)$.

3.2. Shape Space as Quotient Space

Since we are only interested in human body shape and not in the correspondence with the template, we would like to identify transformations that can be related through a human preserving transformation. This is accomplished using

the notion of group action and orbits under those group actions.

Given a group G acting on $\mathcal{H}$, the elements in $\mathcal{H}$ obtained by following a fix registered human body $f \in \mathcal{H}$ when acted on by all elements of G is called the *G-orbit* of f or the *equivalence class* of f under the action of G, and will be denoted by $[f]$. In particular, when G is the reparameterization group $\Gamma := \text{Diff}^+(\mathcal{T})$, the orbit of $f \in \mathcal{H}$ is characterized by the human shape $f(\mathcal{T}) = S$, i.e. the elements in $[f] = \{f \circ \gamma^{-1} \text{ for } \gamma \in \Gamma\}$ are all possible registrations of S. The set of orbits of $\mathcal{H}$ under a group G is called the *quotient space* and will be denoted by $\mathcal{H}/G$. The quotient space of interest in this paper is called *shape space* and is defined as follows.

Definition 3.1 *The shape space $\mathcal{S}$ is the set of (oriented) human bodies in $\mathbb{R}^3$, which are diffeomorphic to $\mathcal{T}$, modulo rotation. It is isomorphic to the quotient space of the pre-shape space $\mathcal{H}$ by the human motion-preserving group $G := \text{Diff}^+(\mathcal{T}) \times SO(3)$: $\mathcal{S} = \mathcal{H}/G$.*

In this paper, each human body surface is aligned to a given template $\mathcal{T}$ (See Appendix).

4. Shape Space as Section of a Fiber Bundle

In this paper, each human body surface is aligned to a given human template (SMPL template). As illustrated in Figure 1, the geometric features of the template are aligned with geometric feature of the human surface (for instance, the finger tips of the template correspond to the finger tips of the other humans bodies).

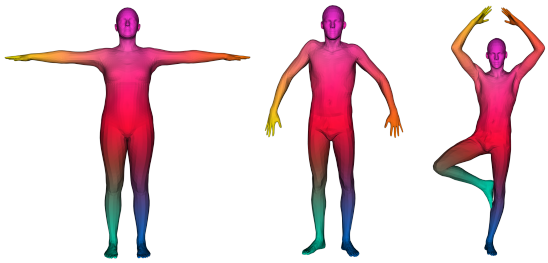


Figure 1: Alignment with the template: The 3 meshes in different poses are displayed with different color on extremities. This validate the choice to work on this particular section of the fiber bundle $\Pi : \mathcal{H} \to \mathcal{H}/G$

Mathematically the choice of a preferred alignment with the template is called a section $\mathcal{S}_0$ of the fiber bundle $\Pi : \mathcal{H} \to \mathcal{H}/G$. A section of Π is a (smooth) map assigning to each equivalence class $[f_0] \in \mathcal{H}/G$ a representative $f_0 \in \mathcal{H}$ in this class, i.e. such that $\Pi(f_0) = [f_0]$. This notion is illustrated in Figure 2. The section we are using, i.e. the correspondence with the template, is smooth, thanks to the geometric alignment as explained above. An illustration of the section is displayed in Figure 2.

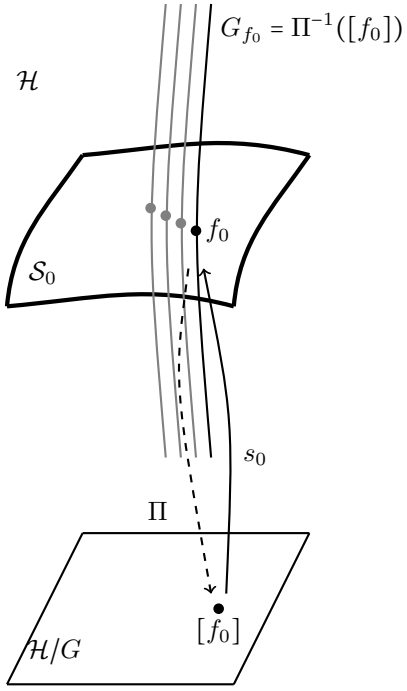


Figure 2: Section of the fiber bundle $\Pi : \mathcal{H} \to \mathcal{H}/G$: the one-to-one correspondance with the template mesh allows us to work on the corresponding section $\mathcal{S}_0$ as a shape space. The correspondance initially gives the section for $\text{Diff}^+(\mathcal{T})$, but with Procrustes analysis, the section for $\text{SO}(3)$ comes straightforwardly.

In this paper, we pull-back the Riemannian metrics that are defined on shape space (see in section 4) on the preferred section $\mathcal{S}_0$ given by the correspondence with the template.

5. Riemannian Analysis of Human Shapes on $\mathcal{S}_0$

Next, we describe our approach to construct the metric between two elements of $\mathcal{S}_0$ and the "optimal" deformation from one human surface to another. Since human surfaces are represented as elements of $\mathcal{S}_0$, a natural formulation of "optimal" is to consider the two corresponding elements in $\mathcal{S}_0$ and to construct a geodesic connecting them in $\mathcal{S}_0$.

5.1. Elastic Riemannian Metric

Consider a parameterized surface $f : \mathcal{T} \to \mathbb{R}^3$. Denote by $g = f^* \bar{g}$ the pull-back of the Euclidian metric $\bar{g}$ of $\mathbb{R}^3$ and by n_f the unit normal vector field (Gauss map) on $S = f(\mathcal{T})$.

The metric g and the normal vector field n are defined using derivatives of f according to:

$$g = \begin{pmatrix} \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial u} \right\rangle & \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v} \right\rangle \\ \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v} \right\rangle & \left\langle \frac{\partial f}{\partial v}, \frac{\partial f}{\partial v} \right\rangle \end{pmatrix} \text{ and } n_f = \frac{f_u \times f_v}{\|f_u \times f_v\|},$$

where f_u and f_v are the derivatives of f with respect to the local coordinates (u, v) on $\mathcal{T}$. We consider the following relationship between parameterized surfaces on one hand and the product space of metrics and normals on the other :

$$\begin{array}{rccc} \Phi : & \mathcal{S}_0 & \longrightarrow & \mathrm{Met}(\mathcal{T}) \times \mathcal{C}^\infty(\mathcal{T}, \mathbb{S}^2) \\ & f & \longmapsto & (g, n). \end{array}$$

It follows from the fundamental theorem of surface theory (see Bonnet's Theorem in [7] for the local result, Theorem 3.8.8 in [12] or Theorem 2.8-1 in [5] for the global result) that two parameterized surfaces f_1 and f_2 having the same representation (g, n) differ at most by a translation (and rotation for g). This theorem implies that we can represent a surface by its induced metric $g = f^* \bar{g}$ and the unit normal field $n = n_f$, for the purpose of analyzing its shape. We will not loose any information about the shape of a surface f if we represent it by the pair (g, n). The induced metric g captures the intrinsic shape, while the normal n captures the extrinsic geometry of shape. The numerical computation of the metric g is given in the appendix.

5.2. The Manifold of Metrics on $\mathcal{T}$ and its Geodesic Distance

The space of positive-definite Riemannian metrics on $\mathcal{T}$ will be denoted by $\mathrm{Met}(\mathcal{T})$. Once we have selected a Riemannian metric for a human body, it is a point in the infinite-dimensional manifold $\mathrm{Met}(\mathcal{T})$. We will equip the infinite-dimensional space of all Riemannian metrics with a diffeomorphism-invariant Riemannian metric, called the Ebin (or DeWitt) metric [8, 6], as suggested by [22]. The Riemannian metric on the tangent space is defined by:

$$((\delta g, \delta g))_g = \int_{\mathcal{T}} \mathrm{Tr}(g^{-1}\delta g_0 g^{-1}\delta g_0) + \lambda \mathrm{Tr}(g^{-1}\delta g)^2 \mu g \quad (1)$$

with $\delta g_0 = \delta g - \frac{1}{2}\mathrm{Tr}(g^{-1}\delta g)g$ is called the traceless part of δg.

The space of the metrics on $\mathcal{T}$ with respect to the geodesic distance of the DeWitt metric is not metrically complete for any choice of λ. As suggested by [22], it is possible to dermine a metric completion of $\mathcal{T}$ for any choice of λ denoted in the following by $\overline{\mathrm{Met}}(\mathcal{T})$. The following theorem, from [22], presents the geodesic distance for the metrics for any choice of λ.

Theorem 5.1 *Let $g_1, g_2 \in \overline{\mathrm{Met}}(\mathcal{T})$. The square of the geodesic distance for the family of metrics is*

$$d^\lambda(g_1, g_2)^2 = \int_M d_{\frac{\lambda}{\mathrm{Sym}}}(g_1(x), g_2(x))^2 \, dx$$

where

$$d_{\frac{\lambda}{\mathrm{Sym}}}(g_1(x), g_2(x))^2 = 16\lambda\left(s_1^2(x) - 2s_1(x)s_2(x)\cos(\theta(x)) + s_2^2(x)\right)$$

with

$$s_1(x) = \sqrt[4]{\det(g_1(x))}, \quad s_2(x) = \sqrt[4]{\det(g_2(x))}$$
$$\theta(x) = \min\left\{\pi, \frac{\sqrt{\lambda^{-1}\,\mathrm{tr}(K_0^2(x))}}{4}\right\}$$
$$K(x) = \begin{cases} 0 \text{ if either } g_1(x) \text{ or } g_2(x) \text{ is degenerate} \\ g_1(x)\log\left(g_1(x)^{-1}g_2(x)\right) \text{ else} \end{cases}$$
$$K_0(x) = K(x) - \mathrm{tr}\left(g_1^{-1}(x)K(x)\right)g_1(x)$$

Theorem 5.2 *Let a, λ, c, three positive real numbers. We equip the space $\mathrm{Met}(\mathcal{T}) \times C^\infty\left(\mathcal{T}, \mathbb{S}^2\right)$ with the following Riemannian metric:*

$$(((\delta g, \delta n), (\delta g, \delta n)))_{g,n} = a\int_{\mathcal{T}} Tr(g^{-1}\delta g_0 g^{-1}\delta g_0) + \lambda Tr(g^{-1}\delta g)^2 \mu g + c\int \langle \delta n, \delta n\rangle \mu g \quad (2)$$

Let $f_1, f_2 \in \mathcal{H}$ such that $\Phi(f_1) = (g_1, n_1), \Phi(f_2) = (g_2, n_2)$, then the square of the distance $d_{\mathcal{H}}$ between f_1 and f_2, with parameters a, λ, c, is given by

$$d_{\mathcal{H}}^{a,\lambda,c}(f_1, f_2)^2 = a d^\lambda(g_1, g_2)^2 + c\int_{\mathcal{T}} d_{\mathbb{S}^2}(n_1(x), n_2(x))^2 dx \quad (3)$$

where d^λ is given by Theorem 5.1 and $d_{\mathbb{S}^2}(n_1(x), n_2(x)) = \arccos\langle n_1(x), n_2(x)\rangle$ is the geodesic distance on $\mathbb{S}^2$.

The proof of this result is straightforward from the results of [22]. The equation 3 defines a 3-parameter (a, λ and c) familly of Riemannian metrics.

5.3. Computation of Geodesics

As mentioned above, an important advantage of our Riemannian approach over many past papers is its ability to compute not only the distance between two human surfaces but also the geodesics or the deformations between shapes. The computation of geodesics requires the minimization of an energy. In [25] the path-straightening method is used to find critical points of the energy functional. Starting with an arbitrary path, the method consists of iteratively deforming (or "straightening") the path in the opposite direction of the gradient, until the path converges to a geodesic. The problem would then be a problem of optimization on the set of vertices of the shape. However, this can lead to numerical instabilities. We will use another, more stable approach [23]. In this approach, after choosing a time step $\frac{1}{T}, T \in \mathbb{N}$, the path is set to the linear path (initialization) on which we add a sum of deformations:

$$f(t_0) = f_0, \quad f(t_T) = f_1$$
$$f(t_i) = (1-t_i) f_0 + t_i f_1 + \sum_j \alpha_{ij} \mathcal{D}_j \tag{4}$$

Where $\mathcal{D}_j$ is an orthogonal basis of $N_{\mathcal{D}}$ plausible deformations gathered beforehand. The computation of the geodesic requires the minimization of the energy functional $E(\alpha)$, defined by:

$$E(\alpha) = \int_0^1 \left(\left(\frac{d\Phi(f(t))}{dt}, \frac{d\Phi(f(t))}{dt} \right) \right)_{\Phi(f(t))} dt \tag{5}$$

with $\alpha \in \mathbb{R}^{(T-2)*N_{\mathcal{D}}}$ the vector containing all α_{ij} presented in equation 4, and $((.,.))_{\Phi(f(t))}$ being the pullback Riemannian metric of Φ on the space $\mathrm{Met}(\mathcal{T}) \times C^{\infty}(\mathcal{T}, \mathcal{S}^2)$.

To find the optimal coefficients α, similar to [23], we employ the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method [9], implemented in the SciPy library [29] where we calculate the gradient using the automatic differentiation feature of PyTorch library [19].

Basis Deformations In [13], [14], [22], [25], spherically parameterization of 3D objects is used and spherical harmonics are computed to define the set of deformations. However, human surfaces will require a large number of basis elements to achieve high accuracy and capture all the human surface details. In addition, in the case of human shapes, we are using a human template as a parametrization and there are several publicly available dynamic human shapes that can be used to build a PCA basis of deformations.

In our case to build such real deformations, we use the publicly Dynamic FAUST dataset [3], which contains motions registered to the template $\mathcal{T}$. 10 individuals (5 males, 5 females) perform 14 different motions, sampled at the rate of 60 frame per second. Given a set of motions, we collect deformations by gathering differences from the sequences. Let $(m_1, ..., m_T) \in \mathcal{S}_0$ be a motion available in the dataset. We define the small deformations that we collect from the motions as the family $(m_{n\tau+\tau} - m_{n\tau})_n$, with τ being a time interval chosen manually, fixed to 10 frames ($\simeq$160 ms). Thus, given a set of training samples, we can compute its PCA basis. In our experiments, the number of PCA basis elements required is of the order of 100.

6. Statistical Analysis of Human Shapes

For future statistical analysis, we are interested in defining a notion of "mean" for a given set of human shape. Now that we have defined all of the required mathematical tools for comparing human shapes, we will show how these tools can be used to compute statistics such a mean of a set of human shapes. Let $f_1, \dots f_n$ be a set of human shapes. The mean of a set of human shapes is the human shape that is as close as possible to all of the human shapes in the set of human shapes, under the distance metric defined by Equation 3. This is known as the Karcher mean and is defined as the human shape that minimizes the sum of squared distances to all of the human shape in the given human shape. Let $\{f_1, \dots, f_n\} \in \mathcal{S}_0$, the Karcher mean can be defined as:

$$\bar{f} = \arg\min_{\mu \in \mathcal{S}_0} \sum_{i=1}^{n} d(f_i, \mu)^2 \tag{6}$$

In order to find the Karcher mean one can define the following functional:

$$\mathcal{V} : \mathcal{S}_0 \to \mathbb{R}, \mathcal{V}(f) = \sum_{i=1}^{n} d(f_i, f)^2 \tag{7}$$

That is differentiable with the distance previously computed. We initialize the Karcher mean as f_1 and set it to be the sum of f_1 with a linear combination of deformations:

$$\bar{f} = f_1 + \sum_j \beta_j \mathcal{D}_i$$

The functional to minimize becomes:

$$\mathcal{W}(\beta) = \mathcal{V}(f_1 + \sum_i \beta_i \mathcal{D}_i)$$

Algorithm 1: Computation of Geodesics

Input : the source and target surfaces f_1 and f_2, a, λ, c the parameter of the elastic metric
Output: f_{geo}: the geodesic connecting f_1 and f_2
1: Initialize $\alpha_{ij} = 0$ and $f(t_i)$ by linear path;
2: Define the energy functional $E(\alpha)$ in an automatic differentiation framework (PyTorch here), that computes the gradient value $\nabla_\alpha E$ along the functional value;
3: Minimize E with respect to α with a BFGS implementation (SciPy *BFGS* or *L-BFGS-B*), that uses the gradient $\nabla_\alpha E$;
4: Set the geodesic to be: $f_{\text{geo}}(t_i) = t_i f_0 + (1-t_i) f_1 + \sum_j \alpha_{ij} \mathcal{D}_j$;
5: **return** *the final geodesic* f_{geo}

7. Experiments

7.1. Assessment of the Family of Elastic Metrics

To further assess the pertinence of the family of elastic metrics defined in Equation. 3 in human shape and pose analysis, we measured pairwise distances of the metric on the registrations present in the FAUST dataset [3]. It contains 10 individuals (5 males, 5 females) in 10 different

Algorithm 2: Karcher Mean of Human Shapes

Input : $f_1, \dots f_n$ a set of human body, a, λ, c the parameter of the elastic metric

Output: $\bar{f}$: Karcher mean

1: Initialize $\alpha = 0$ and $\bar{f} = f_1$ by the first shape in the set;

2: Define the Karcher mean functional $\mathcal{W}^{a,\lambda,c}(\beta)$ in an automatic differentiation framework (PyTorch here) that computes the gradient value $\nabla_\beta \mathcal{W}$ along the functional value;

3: Minimize $\mathcal{W}$ with respect to β with a BFGS implementation (SciPy *BFGS* or *L-BFGS-B*), that uses the gradient $\nabla_\beta \mathcal{W}$;

4: Set the Karcher mean to be: $\bar{f} = f_1 + \sum_i \beta_i \mathcal{D}_i$;

5: **return** *Karcher mean*

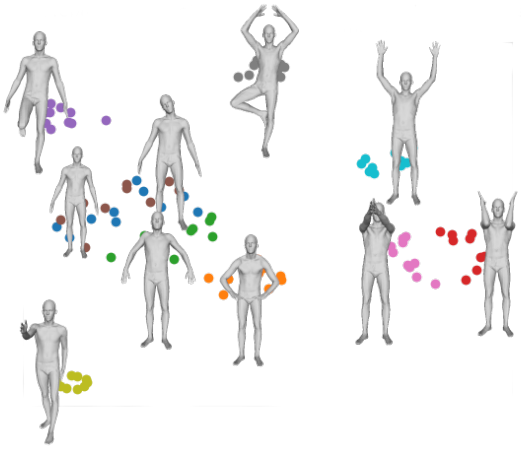


Figure 3: 2D visualization of the FAUST dataset by our method using t-SNE algorithm based on the metric from equation 3. The metric parameters are set to $a = 1, \lambda = 0.0001, c = 0$. Each color represents a class of pose and a class representative is also displayed.

poses. We present in Figure 3 and 4 2D visualizations of the dataset using the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm [26].

The Figure 3 clearly evidences that the 3D human with similar poses belong to very close distributions. These results show the assumption that given $a = 0, \lambda = 0, c = 1$ (normal field L_2 metric), the metric is preserved under shape change, and could be used in pose and motion analysis application [18, 28]. The figure 4 shows that 3D human with similar shape belong to very close distribution. These results states the assumption that given $a = 1, \lambda = 0.0001, c = 0$, the metric is preserved under pose change, and could be used in many shape analysis application approaches [20] and [16].

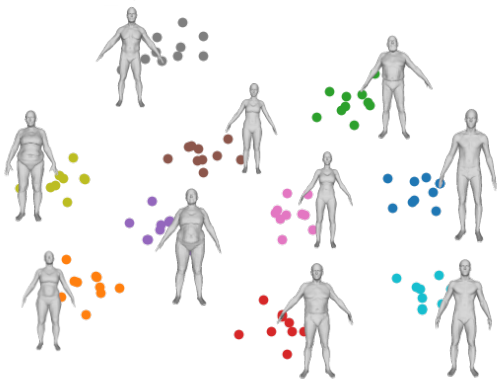


Figure 4: 2D visualization of the FAUST dataset by our method using t-SNE algorithm based on the metric from equation 3. The metric parameters are set to $a = 0, \lambda = 0, c = 1$. Each color represents a class of shape and a class representative is also displayed.

7.2. Geodesics and Karcher Mean

We performed a number of experiments using human surfaces of same and different persons under a variety of pose and shape, and studied the resulting geodesic paths. Some examples are shown in Figures 5. Figures show human surfaces f_1 and f_2 of the same person under different pose.

Figure 5 shows the geodesic path between f_1 (shown in far left) and f_2 (shown in far right). Drawn in between are human surfaces denoting equally spaced points along the geodesic path. In terms of the Riemannian metric chosen, these paths denote the optimal deformations in going from the first human body to the second and the path lengths quantify the amount of deformations. For this experiment, we also provide a curve of the energy, available right to the paths, which shows that the energy decreases smoothly with time.

For the first path, the change in the pose induces small changes in shape. We thus want to minimize the shape change along the path, which would set the extrinsic parameters $c = 0$. We find that $a = 1, \lambda = 1$ gives the best visual results.

The second path is a path with change in shape. We thus want to minimize the pose change along the path, which would set parameters $a = \lambda = 0$, and the normal parameter $c = 1$.

The geodesic computation were made on a computer setup with Intel(R) Xeon(R) Bronze 3204 CPU @ 1.90GHz, and a Nvidia Quadro RTX 4000 8GB GPU. The computation time of the different geodesics took less than 5

mins.

An example of using Karcher mean to compute the average human body is shown in Figure 6. Figure 6 consists of five human body in the same pose with varying shape.

We also compare the results obtained with our method to the results using linear geodesic path, SRNF and SMPL descriptors.

1. The linear geodesic path defined by:

$$\begin{aligned} f(t_0) &= f_0, \quad f(t_T) = f_1 \\ f(t_i) &= (1-t_i) f_0 + t_i f_1 \end{aligned} \tag{8}$$

2. The SRNF geodesic path is also visualized. This representation has been used to analyze human shapes with interesting results [14, 23]. The SRNF is a pointwise representation based on $q = \sqrt{A}n$, where $A = \|f_u \times f_v\|$ is the area, and n the normal field. We compute the geodesic for the $SRNF$ representation with the same method as presented in this section.

   As shown in Figures 7(a) and (b), the linear interpolation and SRNF lead to unnatural deformations for human paths. The deformation between surfaces contains many artifacts and degeneracies.

3. SMPL body model [17] : The SMPL model is a human blend shape model. The human shape is presented as a function of β, θ, with θ being the parameters of human body pose, as a cartesian product of axis angle rotation of skeletal joints (21 joints), in axis-angle representation, which lives in $\mathbb{R}^{21*3} = \mathbb{R}^{63}$. β are the parameters of the human body shape being the coefficients of linear combination of Principal Component Analysis (PCA) shape decomposition (10 components). After fitting SMPL model to the FAUST dataset, we can compute the corresponding geodesic, using the resulting shapes of the linear path in the SMPL space, see Figure 7.

In all examples, our approach provides better results.

8. Application to Pose and Shape Retrieval

Here, we demonstrate how the proposed metric can be exploited for 3D human retrieval. Given a 3D human, we look for the similar 3D human in a database.

8.1. Evaluation Metrics and Comparisons

We test the usefulness of the family of metrics (Equation 3) in 3D human shape and pose retrieval.

Evaluation metrics: We use three evaluation measures. For all measures a high score implies better results.

1. **Nearest neighbor (NN):** It equals one if the nearest neighbor is of the same class of the query, 0 otherwise. This statistic provides an indication of how well a nearest neighbor classifier would perform.

2. **First-tier (FT), Second-tier (ST):** the percentage of models in the query's class C that appear within the top K matches, K depending on query's class size. For a class with $|C|$ members, $K = |C| - 1$ for the first tier, and $K = 2 \times (|C| - 1)$ for the second tier.

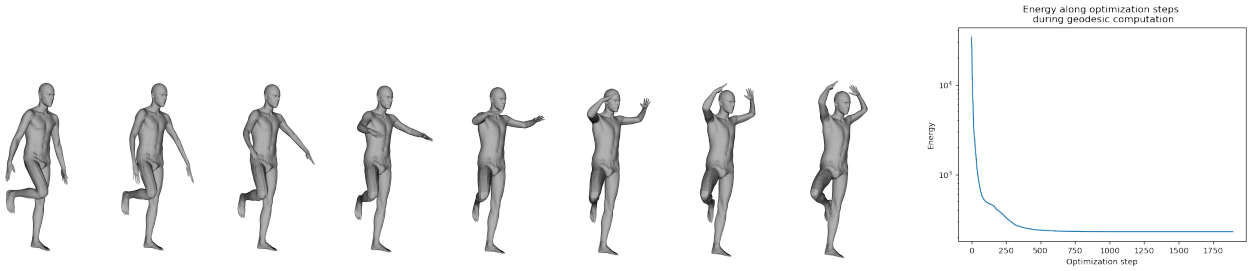
Comparisons: We propose four methods for comparison with our method.The first method GDVAE [2] is a point cloud variational autoencoder which is trained to disantengle the intrinsic and extrinsic informations of a given shape in the latent space, and propose a latent vector that decomposes in an intrinsic and extrinsic part. We used the FAUST meshes as input of their available trained network and gathered their extrinsic latent vectors, which lives in $\mathbb{R}^{12}$, along with their intrinsic latent vectors, which were for human pose retrieval and shape retrieval respectively. The network has been trained on the SURREAL dataset [27] .
The second method proposed by Zhou et al. [30] is a mesh autoencoder based on Neural3DMM [4] graph neural network structure. They disantengle the shape and pose in the latent space. We apply the FAUST meshes on their available network, trained on AMASS dataset, and use the pose latent vector, which lives in $\mathbb{R}^{112}$ as a descriptor for comparison. For human shape, the Area Projection Transform [10] which won the human shape retrieval challenge [20] is presented. It has been designed for a different goal here, since it is parameterization invariant. We also compare to the SRNF metric that showed reliable results for pose retrieval. Finally we use both shape and pose representation from the SMPL body model for the respective retrieval tasks.

8.2. Experimental Results

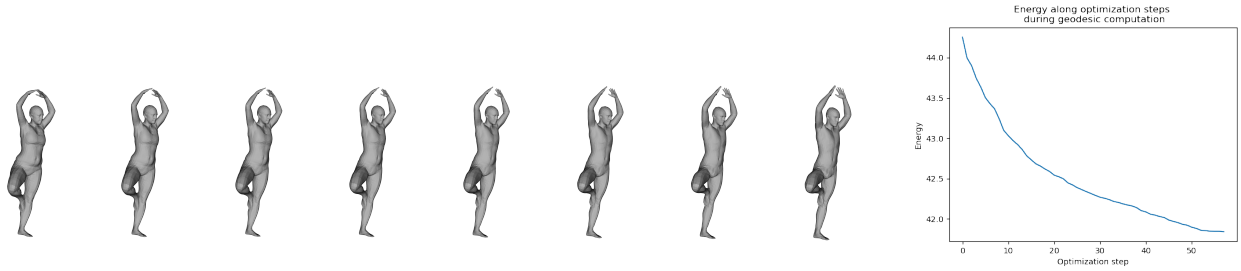
In this section, we perform evaluations of our method in FAUST dataset. We evaluate on pose and shape retrieval. The scores displayed in evaluation tables are the mean scores computed over the dataset. The evaluation results in Table 1 demonstrate that our method outperforms the previous state of the art shape retrieval methods in term of NN criteria. The Table 2 shows that the proposed approach provides the best results on pose retrieval in term of FT and ST criteria. We also find that for shape retrieval, the best parameters are $a = 1, \lambda << a$. The computation times for each pairwise distance were ≃70 ms and ≃80 ms for pose and shape retrieval respectively.

9. Acknowledgments

This work was supported by the ANR project Human4D ANR-19-CE23-0020. This work was also partially supported by the French State, managed by National Agency



(a)



(b)

Figure 5: Examples of geodesic path between f_1 and far left and f_2 far right: (a) with metric parameters (a=1, λ = 1, c=0), (b) with metric parameters (a=0, λ = 0, c=1). The corresponding energy evolution during optimization are displayed on the right. Computation time was respectively 3min31s and 10.6s.

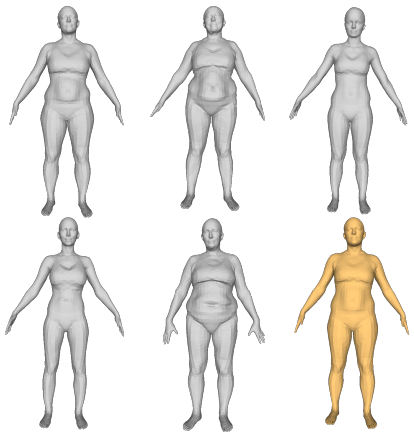


Figure 6: Karcher mean (yellow) for a five different people relatively to the distance with metric parameters $(a = 0, \lambda = 0, c = 1)$.

Repr.	NN	FT	ST
GDVAE intrinsic [2]	27	24.8	46.2
Zhou et al. shape[30]	42	24.8	42.8
SMPL shape vector	98	72.4	86.7
APT [10]	96	86.5	96.2
Metric (1, 0.0001, 0)	**100**	**94.8**	**97.1**

Table 1: FAUST dataset results for shape retrieval

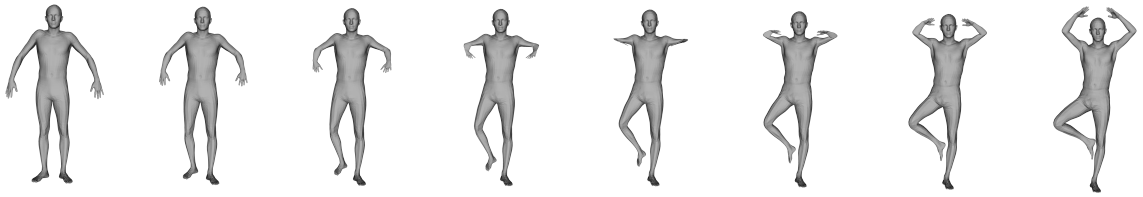
Repr.	NN	FT	ST
GDVAE extrinsic [2]	60	38.0	54.2
Zhou et al. pose[30]	82	69.2	83.4
SMPL pose vector	80	84.4	95.2
SRNF	73	77.7	94.4
Metric (0, 0, 1)	**85**	**88.3**	**97.6**

Table 2: FAUST dataset results for pose retrieval

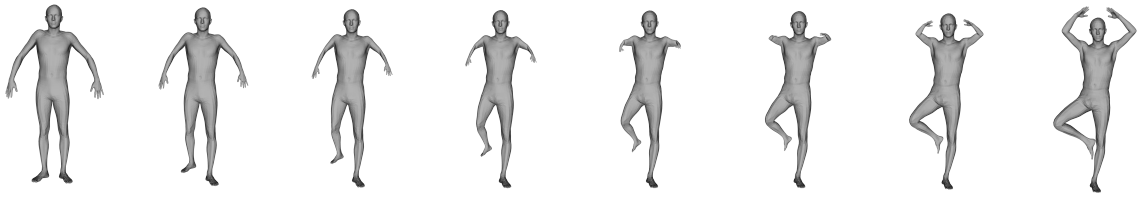
for Research (ANR) under the Investments for the future program with reference ANR-16-IDEX-0004 ULNE.
The authors thank Eric Klassen, Martin Bauer and Zhe Su from Department of Mathematics, Florida State University, for the discussion on the implementation of geodesic computation. We are grateful to Bruno Levy from Inria Nancy Grand-Est research center for the discussion about the computation of g on a triangulated mesh.

10. Conclusion

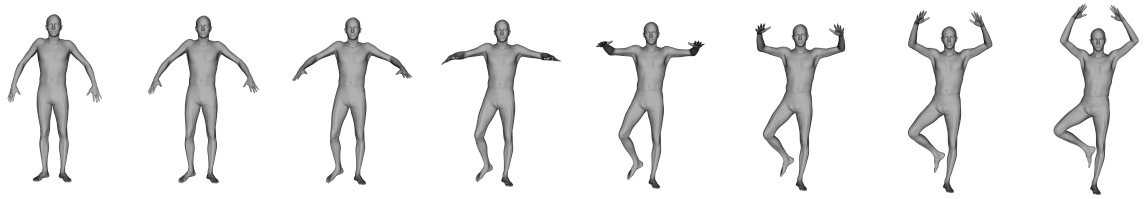
In this paper we have proposed a novel Riemannian framework which allows not only to compute a metric between human bodies under pose and shape changes, but also provides a geodesic path between human bodies, and statistical tools (eg. mean of human shape). We have proposed for the first time the use of a family of Riemannian metrics in human shape analysis, and we have shown that they



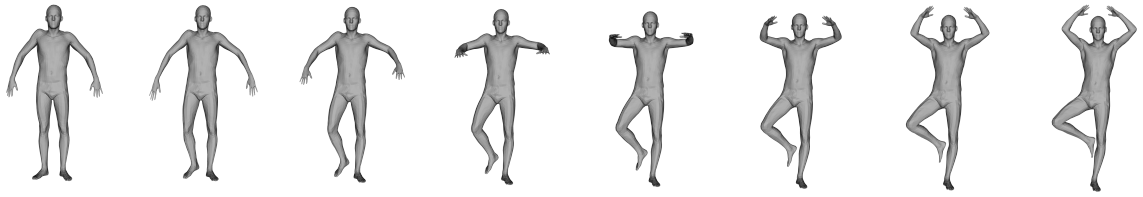
(a) Linear geodesic path



(b) Geodesic computed with SRNF



(c) Geodesic computed in SMPL space



(d) Geodesic computed with our approach, metric parameters are set to $a = 1, \lambda = 1, c = 0$. Computation time was 3min10s.

Figure 7: Comparison of our approach with different frameworks. We observe that the linear initial path is challenging, while the SRNF path induces distortion in the shape. Finally although the SMPL geodesic is able to keep the shape, we argue that the path of our approach is the most natural path compared to the one proposed by SMPL: the natural deformations between the source and target shape would indeed bend more the elbow. In addition, in the SMPL fitting step the target and source shapes are modified.

are able to classify correctly human body shape and pose. In addition, the comparison of the geodesic paths obtained by our framework to linear, SRNF and SMPL descriptors, shows qualitatively that our framework provides the best results. We have also demonstrated that the proposed metrics can be exploited in human shape and pose retrieval. We have shown that the proposed method outperforms state-of-art shape and pose methods in term of NN, FT and ST criteria on FAUST dataset. In future work, we will develop more statistical tools such as covariance shapes and PCA to capture shape variability in a given set of human shapes.

Appendix

A. Sample shapes

A.1. SMPL Template

The Figure 8 shows the used SMPL [17] template $\mathcal{T}$.

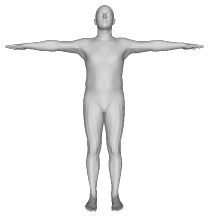


Figure 8: SMPL template

A.2. Examples from FAUST dataset

The Figure 9 shows some examples of human shapes from FAUST dataset [3]. One can see that there is a significant variability in shape and pose.

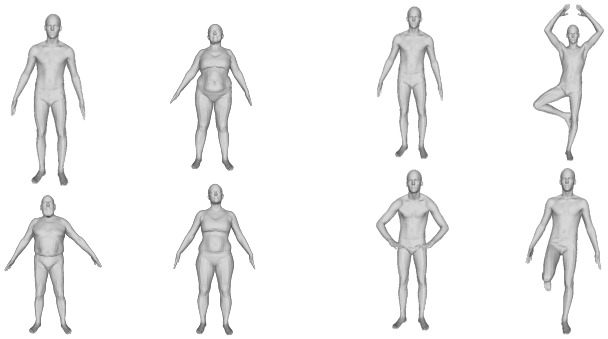


Figure 9: Human shapes from the FAUST dataset.

B. Numerical Computation of the First Fundamental Form g

The computation of the first fundamental g is given by:

$$g = \begin{pmatrix} \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial u} \right\rangle & \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v} \right\rangle \\ \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v} \right\rangle & \left\langle \frac{\partial f}{\partial v}, \frac{\partial f}{\partial v} \right\rangle \end{pmatrix}$$

It is always computed relatively to a given parameterization. Here the parameterization is given relatively to the template.

The first step is to describe a canonical frame for each template triangle t_i^T: Let p_1, p_2, p_3 be the 3 corners of the triangle. We set the origin of the frame to 0. We design the local coordinates by (u, v) in the plane denoted by the triangle, that we define with : the coordinates of the triangle corners being $(0, 0), (u_2, 0), (u_3, v_3)$.

Now we need to find the f that maps the triangle t_i in the template to the triangle t_i^m in the destination mesh. Let q_1, q_2, q_3 be the 3 corners of the triangle.

In a given (u, v) of t_i, the parameterization is of the form:

$$f(u, v) = \lambda_1 q_1 + \lambda_2 q_2 + \lambda_3 q_3$$

In order to compute the derivatives, we need to derive f. The derivative is given by:

$$\frac{\partial f(u,v)}{\partial u} = \frac{\partial \lambda_1}{\partial u} q_1 + \frac{\partial \lambda_2}{\partial u} q_2 + \frac{\partial \lambda_3}{\partial u} q_3$$

$$\frac{\partial f(u,v)}{\partial v} = \frac{\partial \lambda_1}{\partial v} q_1 + \frac{\partial \lambda_2}{\partial v} q_2 + \frac{\partial \lambda_3}{\partial v} q_3$$

Since the triangle are in correspondence, only the λ_i – the barycentric coordinates – depend on (u, v). Given (u, v), those λ_i are defined as the solution of the following equation (see [15] for similar calculations):

$$\begin{pmatrix} 0 & u_2 & u_3 \\ 0 & 0 & v_3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$$

The solution is straightforward and given by:

$$\frac{\partial f(u,v)}{\partial u} = \frac{1}{u_2}(q_2 - q_1)$$

$$\frac{\partial f(u,v)}{\partial v} = \frac{u_3}{v_3 u_2}(q_1 - q_2) + \frac{1}{v_3}(q_3 - q_1)$$

The values of u_2 is simply the length of the first edge of the triangle (l_1). u_3 and v_3 are the projection of first and second edge on the u-axis and v-axis. So v_3 is the height H of the triangle (relative to first edge as the basis), and $\frac{u_3}{v_3}$ is $\tan(\theta)$, where θ is the angle between first edge and second edge:

$$\frac{\partial f(u,v)}{\partial u} = \frac{1}{l_1}(q_2 - q_1)$$

$$\frac{\partial f(u,v)}{\partial v} = \frac{\tan\theta}{l_1}(q_1 - q_2) + \frac{1}{H}(q_3 - q_1)$$

References

[1] Mathieu Aubry, Ulrich Schlickewei, and Daniel Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1626–1633, 2011. 9876

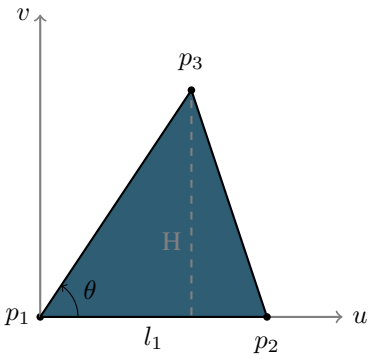


Figure 10: Ordered triangle with the corresponding local reference frame. The quantities l_1, H, θ are useful in the computation of the derivatives.

[2] Tristan Aumentado-Armstrong, Stavros Tsogkas, Allan Jepson, and Sven Dickinson. Geometric disentanglement for generative latent shape models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 9882, 9883

[3] 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. 9880, 9885

[4] 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. *arXiv:1905.02876 [cs]*, Aug. 2019. arXiv: 1905.02876. 9882

[5] P. G. Ciarlet. *An Introduction to Differential Geometry with Applications to Elasticity*, volume 78-79. Kluwer Academic Publishers, 2005. 9879

[6] Bryce S. DeWitt. Quantum theory of gravity. i. the canonical theory. *Phys. Rev.*, 160:1113–1148, Aug 1967. 9879

[7] M. P. do Carmo. *An Introduction to Differential Geometry with Applications to Elasticity*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1976. 9879

[8] D. G. Ebin. The manifold of Riemannian metrics, in : Global analysis, berkeley, calif., 1968. *Proc. Sympos. Pure Math.*, 15:11–40, 1970. 9879

[9] R. (Roger) Fletcher. *Practical methods of optimization*. Chichester ; New York : Wiley, 1987. 9880

[10] A. Giachetti and C. Lovato. Radial Symmetry Detection and Shape Characterization with the Multiscale Area Projection Transform. *Computer Graphics Forum*, 31(5):1669–1678, 2012. 9882, 9883

[11] 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. 9877

[12] W. Klingenberg. *Eine Vorlesung in Differentialgeometrie*. Springer Verlag, Berlin, 1973. 9879

[13] 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. 9876, 9877, 9880

[14] Hamid Laga, Qian Xie, Ian H. Jermyn, and Anuj Srivastava. Numerical inversion of SRNF maps for elastic shape analysis of genus-zero surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(12):2451–2464, 2017. 9880, 9882

[15] Bruno Lévy. *Géométrie Numérique*. Habilitation à diriger des recherches, Institut National Polytechnique de Lorraine - INPL, Feb. 2008. 9885

[16] Zhouhui Lian, Afzal Godil, Benjamin Bustos, Mohamed Daoudi, Jeroen Hermans, Shun Kawamura, Yukinori Kurita, Guillaume Lavoué, Hien Van Nguyen, Ryutarou Ohbuchi, Yuki Ohkita, Yuya Ohishi, Fatih Porikli, Martin Reuter, Ivan Sipiran, Dirk Smeets, Paul Suetens, Hedi Tabia, and Dirk Vandermeulen. A comparison of methods for non-rigid 3d shape retrieval. *Pattern Recognit.*, 46(1):449–461, 2013. 9876, 9881

[17] 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, Oct. 2015. 9882, 9885

[18] Guoliang Luo, Frederic Cordier, and Hyewon Seo. Spatio-temporal segmentation for the similarity measurement of deforming meshes. *The Visual Computer*, 32(2):243–256, Feb. 2016. 9881

[19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 9880

[20] D. Pickup, X. Sun, P. L. Rosin, R. R. Martin, Z. Cheng, Z. Lian, M. Aono, A. Ben Hamza, A. Bronstein, M. Bronstein, S. Bu, U. Castellani, S. Cheng, V. Garro, A. Giachetti, A. Godil, L. Isaia, J. Han, H. Johan, L. Lai, B. Li, C. Li, H. Li, R. Litman, X. Liu, Z. Liu, Y. Lu, L. Sun, G. Tam, A. Tatsuma, and J. Ye. Shape Retrieval of Non-rigid 3D Human Models. *International Journal of Computer Vision*, 120(2):169–193, 2016. 9876, 9881, 9882

[21] Martin Reuter, Franz-Erich Wolter, and Niklas Peinecke. Laplace-beltrami spectra as 'shape-dna' of surfaces and solids. *Comput. Aided Des.*, 38(4):342–366, 2006. 9876

[22] 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 (CVPR) Workshops*, June 2020. 9876, 9877, 9879, 9880

[23] Zhe Su, Martin Bauer, Stephen C. Preston, Hamid Laga, and Eric Klassen. Shape Analysis of Surfaces Using General Elastic Metrics. *arXiv:1910.02045 [math]*, Oct. 2019. arXiv: 1910.02045. 9879, 9880, 9882

[24] Jian Sun, Maks Ovsjanikov, and Leonidas J. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Comput. Graph. Forum*, 28(5):1383–1392, 2009. 9876

[25] 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, Jan. 2016. arXiv: 1506.03065. 9876, 9877, 9879, 9880

[26] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *JMLR*, 9(86):2579–2605, 2008. 9881

[27] Gül Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J. Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In *CVPR*, 2017. 9882

[28] Christos Veinidis, Antonios Danelakis, Ioannis Pratikakis, and Theoharis Theoharis. Effective descriptors for human action retrieval from 3d mesh sequences. *Int. J. Image Graph.*, 19(3):1950018:1–1950018:34, 2019. 9881

[29] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. 9880

[30] Keyang Zhou, Bharat Lal Bhatnagar, and Gerard Pons-Moll. Unsupervised shape and pose disentanglement for 3d meshes. In *European Conference on Computer Vision (ECCV)*, August 2020. 9882, 9883