



Contact Preserving Shape Transfer: retargeting motion from one shape to another

Jean Basset, Stefanie Wuhrer, Edmond Boyer, Franck Multon

► To cite this version:

Jean Basset, Stefanie Wuhrer, Edmond Boyer, Franck Multon. Contact Preserving Shape Transfer: retargeting motion from one shape to another. Computers and Graphics, 2020, 89, pp.11-23. 10.1016/j.cag.2020.04.002 . hal-02613783

HAL Id: hal-02613783

<https://hal.science/hal-02613783>

Submitted on 20 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Contact preserving shape transfer: Retargeting motion from one shape to another

Jean Basset, Stefanie Wuhrer, Edmond Boyer, Franck Multon

► To cite this version:

Jean Basset, Stefanie Wuhrer, Edmond Boyer, Franck Multon. Contact preserving shape transfer: Retargeting motion from one shape to another. Computers and Graphics, Elsevier, 2020, 89, pp.11-23. 10.1016/j.cag.2020.04.002 . hal-02613783

HAL Id: hal-02613783

<https://hal.archives-ouvertes.fr/hal-02613783>

Submitted on 20 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Contact Preserving Shape Transfer: retargeting motion from one shape to another

Jean Basset^{a,*}, Stefanie Wuhrer^a, Edmond Boyer^a, Franck Multon^b

^aUniv. Grenoble Alpes, Inria, CNRS, Grenoble INP (Institute of Engineering Univ. Grenoble Alpes), LJK, 38000 Grenoble, France

^bUniv Rennes, Inria, CNRS IRISA, M2S, France

ARTICLE INFO

Article history:

Received May 20, 2020

Keywords: Motion retargeting, rigging-free animation, mesh deformation, shape transfer, surface-based retargeting, character animation

ABSTRACT

We present an automatic method that allows to retarget poses from a source to a target character by transferring the shape of the target character onto the desired pose of the source character. By considering shape instead of pose transfer our method allows to better preserve the contextual meaning of the source pose, typically contacts between body parts, than pose-based strategies. To this end, we propose an optimization-based method to deform the source shape in the desired pose using three main energy functions: similarity to the target shape, body part volume preservation, and collision management to preserve existing contacts and prevent penetrations. The results show that our method allows to retarget complex poses with several contacts to different morphologies, and is even able to create new contacts when morphology changes require them, such as increases in body size. To demonstrate the robustness of our approach to different types of shapes, we successfully apply it to basic and dressed human characters as well as wild animal models, without the need to adjust parameters.

1. Introduction

Animation studios have stored terabytes of animation files applied on various 3D characters with meticulous association between skeletal motion and 3D shape, created manually by skilled artists who designed rigged models and corresponding skeletal motions. Retargeting these existing skeletal motions to new characters automatically is a long-standing problem in computer animation [20] that remains challenging, especially when the animation contains close interactions and contacts.

Existing works to address the retargeting problem can be broadly classified into the following three categories. First, skeleton-based retargeting consists in adapting the joint angles

of the character in order to satisfy kinematic constraints either edited manually [20, 30] or automatically built based on geometric constraints between body parts [23]. With these approaches, it is difficult to prevent collisions or more generally respect distance constraints of the body shape. Second, surface-based retargeting considers surface deformations, typically mesh deformations, when transferring the pose of a source character to a target one [44]. In both categories, a key aspect is that the pose is encoded independently from the shape of the characters, which can lead to artifacts in particular for close body part interactions. A third category with data-driven approaches takes advantage of a database of models to learn correlations between shape and pose, and uses this for retargeting motions between characters [7, 26]. While this strategy yields improved results, the associated methods are yet specific to given classes of shapes, typically basic minimally dressed human bodies, and can hardly generalize to other models, such as dressed humans or animals. In essence, modeling the pose independently of the shape in a generic manner is still a challenging

*Corresponding author

e-mail: jean.basset@inria.fr (Jean Basset),
stefanie.wuhrer@inria.fr (Stefanie Wuhrer),
edmond.boyer@inria.fr (Edmond Boyer), fmulton@irisa.fr (Franck Multon)

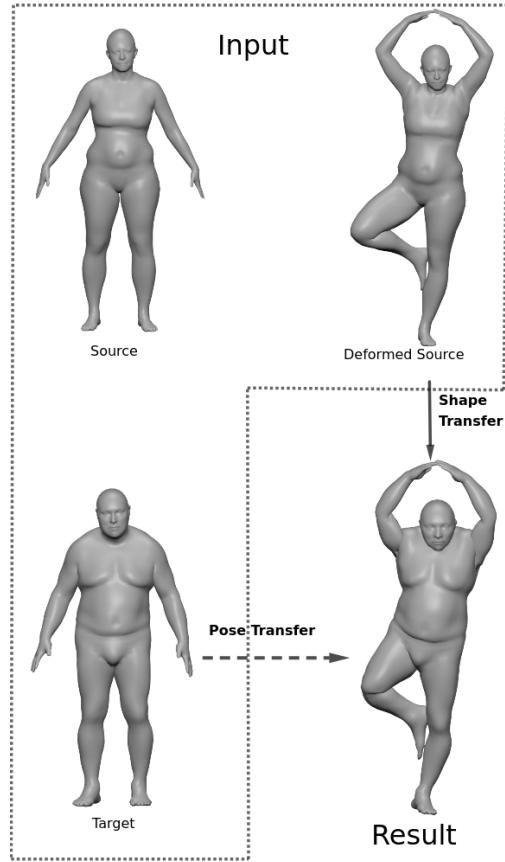


Fig. 1: Overview of the deformation transfer algorithm. Classical works [44] transfer the pose of the source to the target. We propose a novel approach, transferring the shape of the target to the deformed source character.

issue, in particular with body part contacts or close interactions. In this work, we propose an alternative method to solve the motion retargeting problem that is especially designed to handle contacts and interactions. The proposed approach is generic and can be applied to any class of shape, e.g. human or animal characters. In contrast to existing works, we explore transferring the shape of a target character to the desired pose of a source character, as shown in Figure 1. In particular, given as input a source character in both a standard and a deformed pose and a target character in a standard pose, we propose to morph the shape of the source character to make it similar to the target character, in the deformed pose, while maintaining a plausible posture w.r.t. contacts between surfaces. This way, we assume that the resulting target shape+pose better preserves the source pose contextual meaning, i.e. properties of the pose that are independent of the shape. As a consequence, surface contacts due to shape differences in body sizes can be handled by design with our approach.

To achieve shape transfer, we design a variational approach that optimizes three energy functions: similarity to the target shape, volume preservation of body parts, and collision management to preserve existing contacts and prevent penetrations. To allow for motion retargeting, continuity between subsequent poses of a motion sequence is encouraged in a post-process. We show experimentally that this approach can be used to retarget between a wide range of characters.

A preliminary version of this article appears in [8]. This new version includes improvements in collision detection and the adaptation to continuous motion. Furthermore, we present an extended experimental validation beyond basic human shapes by considering motion retargeting applications between basic human body shapes and humans dressed with different garments, as well as motion retargeting between different wild animals. Our experiments indicate that our method can be applied to different shapes as is, without the need to adjust parameters.

The paper is organized as follows. Section 2 summarizes the previous works addressing the problem of motion retargeting, and positions the paper in this state of the art. Section 3 gives an overview of our method and presents our energy function. Section 4 details this energy, and Section 5 presents our minimization procedure. We explain how our method is applied on continuous animations in Section 6. Section 7 presents an extended experimental validation, and finally we conclude the paper in Section 8.

2. Related Work

Previous works mostly explore three major directions to transfer a pose from a source to a target character: using a shape independent pose representation with the joint angles of a skeleton; directly modifying the character’s surface without the help of rigged skeletons; and data-driven methods that leverage a dataset of example shapes in different poses to learn the transfer.

2.1. Skeletal Motion Retargeting

With the popularization of marker-based motion capture systems in the early 90’s, the ability to transfer an animation from an actor to a rigged character rapidly gained success. Motion retargeting was then considered as solving kinematic constraints on joint positions and ensuring continuity using displacement maps [20, 33, 13]. Another approach consists in defining a morphology-independent representation [30, 22] with efficient constraint solvers [29], or in using an intermediate skeleton [40]. As these methods mainly consist in solving static kinematic constraints, postprocessing is needed to ensure continuity, for instance with recursive filters [20].

All these methods generally use predefined kinematic constraints that must be manually tuned. Automatic kinematic constraint detection in the source motion has been proposed [31] to automate the constraint editing problem. Most of these constraints consist in spatial relationship between body segments, which can be modeled as distance constraints [3] or as more generalized spatial relationship between joints [6]. These methods aim at transferring the topology of the body segments of the source motion to the target character, while using generalized inverse kinematics to solve all the corresponding constraints. This idea of modeling the topology between body segments has been extended by introducing an interaction mesh [23, 24]. A more recent work introduces egocentric planes to ensure that the instantaneous separating plane between each pair of body parts is transferred between the source and target motion [39]. This enables real-time motion transfer while preserving most

of the topology between body segments. However, all these methods require a rigged skeleton and cannot handle accurate constraints between body parts when they are not simplified body segments. Moreover, they assume that all the relative positions between body segments are preserved from source to target character, which is unlikely with shapes that differ significantly.

2.2. Surface Mesh Retargeting

Skeleton-based approaches have difficulties dealing with pose features that concern the surface of a character, typically incidence relationships. Another strategy in that respect consists in directly acting on the character surface instead of the joint angles of its skeleton, for instance by displacing the vertices of a surface mesh.

Deformation transfer methods follow this direction. They mostly encode the pose of the source character as a deformation of the source surface mesh and transfer this deformation to the target surface mesh [44, 52].

In order to better handle shape and pose constraints over the surface, recent works encode the pose as spatial relationships between points on the body surface. For instance Liu et al. [36] introduced the context graph, an extension of the interaction mesh proposed for skeleton joint centers [23] to body surfaces. In a context graph, nodes are placed on the body surface and edges of the graph encode distance constraints between nodes. Transferring the pose context consists then in enforcing such distances on the target surface given the source context graph. Using distance constraints to preserve context was also explored by Jin et al. [27], who proposed the Aura mesh, a volumetric mesh enclosing the body surface with a fixed offset. Spatial relationships are then modeled as the interpenetration of this Aura mesh.

Context graphs and Aura meshes both define an arbitrary distance threshold under which spatial relationships should be preserved, assuming hence that node distances below the threshold encode the contextual meaning of the pose. However, distances between body parts beyond this threshold can also embed contextual information. Moreover, some close interactions do not necessarily relate to the contextual meaning of the pose, but can result from intrinsic shape constraints, e.g. surface contacts due to corpulence. Additionally, both works consider simplified external meshes to model contacts and, consequently, can fail to capture fine contacts between the more detailed surface mesh of body segments.

A different strategy is the use of physics based constraints, such as balance [38]. Instead of enforcing distance constraints between body parts, Al Borno et al. [4] propose to apply physics-based forces on the characters. As a result, the pose adapts to the morphology of the target character, e.g. the extent of a kick motion depends on the corpulence. We follow a similar philosophy and allow as well poses on the target shape to adapt to the morphology.

2.3. Hybrid Approaches Combining Skeleton and Mesh

In order to benefit from both worlds a body of work combines skeleton and surface constraints in the pose transfer. Molla et

al. [39] uses a skeleton to model the pose with joint angles and adds constraints, i.e. egocentric mapping to preserve the topology between body parts, that are applied on a body surface approximation. Other methods use a complete surface mesh together with a skeleton [32, 25] to control the surface mesh deformation while preserving the coherence with the skeleton topology. By satisfying both skeletal and surface constraints, natural animations and poses can be generated. While deformation transfer is not the primal contribution of these works, the associated methods can be adapted to this purpose with convincing results as in [32]. Nevertheless, works in this category do not yet account for surface interactions and substantial collisions can appear when transferring poses between significantly different shapes.

2.4. Data Driven Approaches

With the recent availability of important datasets, data driven approaches have become more popular. They enable deformation transfer between shapes that can be very different in nature, e.g. humans and animals, by learning the semantic correspondences over a set of paired examples, for instance poses [7] or even animations [11]. In [10] Bouaziz et al. create blendshapes, i.e. base expressions of a target face, and estimate weights associated with each blendshape on the source to create the resulting target expression. Another recent work investigates deep neural networks for semantic deformation transfer [19], where training is performed on various poses of the source and target that do not need to be in correspondence. Mappings between semantically different poses of humans and animals can also be learned to interactively control animation generation [43]. All these methods neither require skeletons nor point-to-point correspondences between sources and targets. However, heavy pre-processing must be performed for every pair of source and target characters. Also these methods do not explicitly account for surface interactions.

In another vein, datasets of humans allow to create parametric models of human shapes and poses, e.g. [5, 9, 26]. Using these models, it is possible to combine the pose parameters of a source to the shape parameters of a target to effectively perform motion retargeting. However these models do not encode surface interactions and will not prevent collision artifacts, as shown for example in Figure 12.

Recent works have explored deep learning methods for motion retargeting, such as recurrent neural networks [47] or deep reinforcement learning [49]. An important number of these works focus on video based motion retargeting. They take as input a 2D video of a character performing a source motion, and generate a new video of a target character performing a similar motion. These works are based on the recent advances of generative networks such as GANs (e.g. [35, 12]) or VAE (e.g. [18]).

Data driven methods can give very satisfying results, but require usually important datasets to work and are intrinsically limited by such datasets. In the following, we propose instead a direct approach that only needs a source and a target mesh to perform motion retargeting.

3. Method Overview

Our goal is to make a target character reproduce the motion of a source character. To this aim we consider as inputs: the source character mesh in a *standard pose* (e.g. A-pose), the same source character in a flow of *deformed poses* we wish to duplicate with a target character, and the target character mesh in the standard pose, as shown in Figure 1. In this work, we consider the motion as a continuous sequence of static poses. Consequently, for each pose of the source character, the retargeting process should compute a deformed pose adapted to the target character, while preserving continuity in the resulting sequence. All the source and target meshes are assumed to be in correspondence through a single mesh graph. This mesh graph is further assumed to be segmented into body parts (see Section 4.2.1). Note that our method does not require a rigged skeleton.

We proceed by first retargeting each deformed input pose independently, and by subsequently smoothing the resulting animation to encourage continuity in a post-process. To retarget a static frame, starting from the source character in the deformed pose, our approach morphs its surface until its shape fits the target character shape, while preserving the surface contacts present in the source deformed pose. This way, we transfer shapes at the desired poses instead of transferring poses to the desired shapes, as traditional in the existing works.

Our method includes anyway pose deformation since body shape deformations impact the body pose. We follow the common hypothesis that pose deformations of the human body can be modeled as near-isometric [14]. Hence, all near-isometric deformations are attributed to pose, while non-isometric deformations are attributed to shape changes. In particular, we model pose deformations by applying rigid transformations to the mesh's body parts. Non-isometric shape deformations are applied directly to the mesh's vertices.

Input meshes are defined as $\mathcal{V} = (\mathbf{V}, E)$, where E is the set of edges of the mesh and $\mathbf{V} = (v_1, \dots, v_n)$ is the set of mesh vertices with v_i the 3D coordinates of vertex i . We define the rigid body part transformations $\Theta = \{R_P\}_{P \in BP}$, where R_P is the rotation associated with the body part $P \in BP$ and BP is the set of body parts. To perform shape transfer, we cast the problem as an optimization over the vertex positions \mathbf{V} and the rigid transformations Θ , and with respect to three energy terms that account for local and more global shape properties as well as surface contacts:

$$\underset{\mathbf{V}, \Theta}{\operatorname{argmin}} [\gamma_{Shape} E_{Shape}(\mathbf{V}) + \gamma_{Vol} E_{Vol}(\mathbf{V}) + \gamma_C E_C(\Theta)]. \quad (1)$$

The terms E_{Shape} and E_{Vol} penalize the discrepancy in shape and volume with the target character. As they correspond to non-isometric shape deformations, they are minimized w.r.t. \mathbf{V} . The term E_C penalizes collisions of surfaces and loss of contacts present in the source pose. This energy concerns the near-isometric pose deformations and is therefore minimized w.r.t. Θ . The weights γ_{Shape} , γ_{Vol} , and γ_C modulate the influence of each energy term.

The energy terms are detailed in Section 4. To facilitate the shape transfer, before optimizing expression 1, we compute the height of the source and target shapes using their provided standard poses, and pre-scale the deformed source mesh to the height of the target mesh. The method to iteratively minimize Energy 1 is presented in Section 5. Finally, the post-process used to adapt to motion sequences is explained in Section 6.

4. Shape and Pose Optimization

To perform the shape transfer from a source to a target character, we optimize the source shape and pose parameters through a set of energy functions to be minimized: a shape energy term transferring non-isometric deformations, a volume energy term to ensure volume preservation of the target shape, and a contact energy term to preserve relevant pose contacts and avoid penetration between the body parts. These energy terms are detailed in what follows.

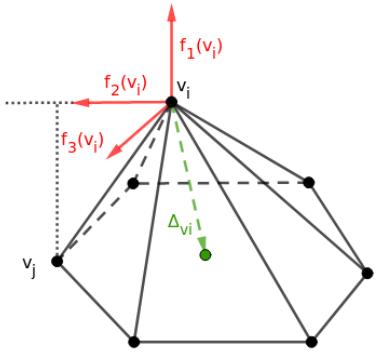
4.1. Local Shape Fidelity

As mentioned earlier pose deformations of the human body can be assumed near-isometric [14]. Hence, two meshes that represent the same shape in different poses should be near-isometric. This property is used in previous works to define shape preserving deformations, e.g. [34]. Building on a similar principle, we make the assumption that geometric shape features that are isometry-invariant encode shape information independently of the pose. Given the source shape in the correct pose, we seek therefore for a transformation that equals isometry-invariant features on both the deformed source and the target shapes. To this aim, we consider Laplacian offsets as local geometric features independent of the pose. They have already been successfully used in [50] as a shape representation for posture invariant shape analysis.

The key idea of this representation is to encode, for each vertex, the offsets w.r.t its neighboring vertices in a local coordinate frame. First the uniform Laplacian matrix L of the template mesh is computed. Since the input meshes are in correspondence with this template, they have the same connectivity graph and the Laplacian matrix is the same for all. This matrix is used to compute for each vertex v_i the Laplacian offset Δ_i as:

$$\begin{pmatrix} \Delta_1 \\ \vdots \\ \Delta_n \end{pmatrix} = L \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \sum_{v_j \in N_1(v_1)} \frac{1}{deg(v_1)} v_j - v_1 \\ \vdots \\ \sum_{v_j \in N_1(v_n)} \frac{1}{deg(v_n)} v_j - v_n \end{pmatrix}, \quad (2)$$

where $N_1(v_i)$ is the first ring neighborhood of v_i on \mathcal{V} . Δ_i is further expressed in a local coordinate system, defined at vertex v_i , to make it invariant to pose deformations. This coordinate system is composed of the normal vector of the surface at vertex v_i (called $f_1(v_i)$), a projection of a fixed vertex neighbour of v_i in the orthogonal plane of the normal (called $f_2(v_i)$), and their cross product (called $f_3(v_i)$). The three vectors are normalized to create the local coordinate system at vertex v_i (see Figure 2). It is invariant to translation and rotation of the neighbourhood

Fig. 2: Local frame (red) of Δ_i (green) at v_i .

of v_i . Thus, expressing Δ_i in this coordinate system makes the shape representation invariant to pose.

In a preliminary step of the algorithm, we compute the target shape representation $\Omega^T = \{\omega_1^T(v_i^T), \omega_2^T(v_i^T), \omega_3^T(v_i^T)\}_{v_i^T \in V^T}$ at each vertex of the target character in the standard pose, where $\omega_k^T(v_i^T)$ are the coordinates of Δ_i^T expressed in the local coordinate system of the target shape.

We subsequently deform the source shape in the desired pose so that it presents similar local offsets, thereby changing its local forms towards the target local forms. That is, for each vertex v_i on the deformed source shape we expect the Laplacian offset Δ_i to equal its counterpart Δ_i^T on the target shape, yielding the following shape energy term:

$$E_{Shape}(\mathbf{V}) = \sum_{v_i \in V} (\Delta_i - \Delta_i^T)^2. \quad (3)$$

A direct minimization of the above term results in a non-linear and complex optimization. In practice, given a fixed local configuration, i.e. $(f_1(v_i), f_2(v_i), f_3(v_i), v_j \in N_1(v_i))$, around v_i , expression 3 is minimized by moving v_i towards the optimal position \hat{v}_i :

$$\hat{v}_i = \sum_{v_j \in N_1(v_i)} \frac{v_j}{deg(v_i)} - (\omega_1^T(v_i^T)f_1(v_i) + \omega_2^T(v_i^T)f_2(v_i) + \omega_3^T(v_i^T)f_3(v_i)). \quad (4)$$

Hence we proceed iteratively in two steps: (i) vertices are moved in the optimal direction $\hat{v}_i - v_i$; (ii) local configurations are re-estimated. Details on the iterative solving are given in section 5.

4.2. Volume Preservation

By encoding the shape as described in the previous section we can enforce local shape properties. However, it is known that isometries do not encode the volume of the shape [34]. Hence, two isometric shapes can have drastically different volumes [14]. Consequently, the Laplacian offsets as used in expression 3 do not guarantee a global volume preservation between the deformed source and target shapes. We explain below how to remedy this issue.

Instead of preserving the full body volume, we propose to consider the volume of each body part independently. The reason for this is that the human body volume is not equally distributed among its body parts. Thus, preserving the volume globally leads to an under-constrained problem and can result in unnatural distribution of the volume such as inflated arms or faces. On the other hand, preserving the volume at a very local level constrains transformations to be rigid whereas body pose transformations can obviously be non-rigid. In our method we therefore take an in-between strategy and encode the volume at the body part level. While body part volumes do not always stay constant during body deformations, e.g. as a result of muscle deformation or breathing, we assume, in a first approximation, these variations to be negligible between the standard and source poses. Our method tries therefore to equal the volumes of body parts between the deformed source and the target characters. We expect this way a coherent volume distribution while allowing for non-rigid deformations.

4.2.1. Body Part Segmentation

The body part segmentation should ideally separate the body shape into elements that move rigidly. To this aim, we manually segment the template into 17 parts as depicted in figure 3a. Our experiments validate this segmentation however our approach could also consider other segmentations, with more or less details depending on the application. For instance fingers could be separated in the segmentation to better encode complex hand poses.

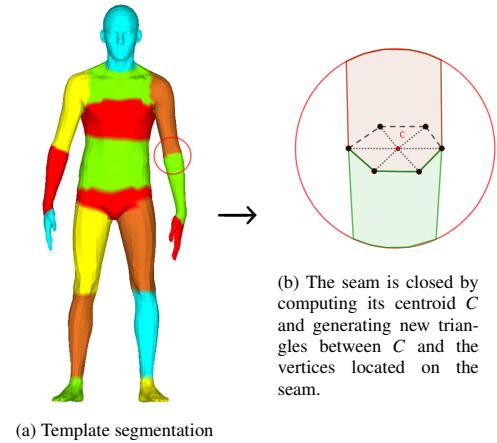


Fig. 3: Body part segmentation on the template and close up of a seam.

Given a segmentation, the volume of each body part can be computed as the sum of the signed volumes of the tetrahedrons formed by the body part's triangles and the origin O [51]. Let $\{v_i, v_j, v_k\}$ be an oriented triangle and O the origin, the signed volume of the tetrahedron $\{O, v_i, v_j, v_k\}$ writes:

$$V_{Oijk} = \frac{1}{6}(-x_k y_j z_i + x_j y_k z_i + x_k y_i z_j - x_i y_k z_j - x_j y_i z_k + x_i y_j z_k), \quad (5)$$

where (x_i, y_i, z_i) are the coordinates of v_i . Assuming body parts to be closed manifold meshes composed of triangles with

consistent orientation, the volume of a body part P is then $V_P = \sum_{T \in P} V_{O,T}$, where T denotes triangles. This requires body parts to be closed meshes. To this purpose, body parts are closed by computing the centroid of the seam between two neighboring body parts, and by generating triangles between this centroid and the vertices on the seam as shown in Figure 3b.

4.2.2. Volume Term

Given the body part segmentation, the volume energy term measures the discrepancy between body part volumes on the deformed source shape and on the target shape:

$$E_{Vol}(\mathbf{V}) = \sum_{P \in BP} (V_P - V_P^T)^2. \quad (6)$$

4.3. Contacts

The shape and volume terms previously presented help to accurately deform the source shape into the target shape. However, differences in morphology can make it difficult for the target character to correctly reproduce the source pose. For example, if the source is a thin character with arms close to the body, it will be challenging for a corpulent target character to reproduce the exact same pose, i.e. body parts presenting the exact same relative positions, since the arms would then intersect the thorax (see Figure 4a). Conversely, if a corpulent source character has a hand touching the belly, a thin target character reproducing the pose might see the hand floating in front of the belly, therefore losing this contextually significant contact. Consequently, it is important to note that changing the morphology of a character while keeping the same pose can result in an impossible or a contextually different pose.

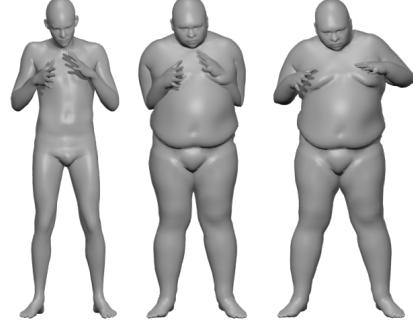
To address this issue, our method includes a contact energy term. This term aims to maintain the contextual posture, i.e. the contacts, of the source character by adapting its pose. In particular, we aim to maintain all contacts present in the source pose, while not introducing inter-penetrations. Our contact term builds on the contact loss presented in [21]. It is composed of a repulsion term that increases when surface inter-penetrations occur, and an attraction term that increases when a contextual contact is lost:

$$E_C(\mathbf{V}) = \gamma_r E_r(\mathbf{V}) + \gamma_a E_a(\mathbf{V}), \quad (7)$$

where E_r and E_a are the repulsion and attraction term respectively, with associated weights γ_r and γ_a . E_C helps correcting the pose of the result w.r.t. self interactions of the surface of the mesh and w.r.t interaction with the ground, as shown in Figure 4.

4.3.1. Repulsion Term

Body part segmentation enables us to follow the rigid members of the human body during the deformation (4.2.1). As such, if the source and target characters have correct poses, no inter-penetration should appear inside a same body part. We thus test only inter-penetrations between a vertex and all body parts but the one it belongs to. The repulsion term also considers collisions with the ground and is defined as



(a) Left to right: source pose, transfer result without the contribution of a repulsion term, and with such a contribution (target from Figure 5c). Notice on the right the arms that do not penetrate the torso anymore, and the wider gap between legs to avoid thigh colliding.



(b) Left to right: source pose, transfer result without the contribution of an attraction term, and such a contribution (target from Figure 6b). Notice on the right the foot that does not penetrate the leg thanks to the repulsion term, but requires anyway the attraction term to keep the contact present in the source pose.

Fig. 4: Contribution of the contact energy terms.

$$E_r(\mathbf{V}) = \gamma_r \sum_{P \in BP} \sum_{\substack{v_i \in \mathbf{V} \setminus P \\ v_i \in Int(P)}} d(v_i, P)^2 + \gamma_{rg} \sum_{\substack{v_i \in \mathbf{V} \\ v_i \in Int(G)}} d(v_i, G)^2, \quad (8)$$

where BP is the body part set, G is the ground, $Int(X)$ is the interior of object X , and $d(v, X)$ is the minimum distance between the vertex v and the object X ; $d(v, X) = \inf_{w \in X} \|v - w\|_2$. The effect of the repulsion term is illustrated in Figure 4a.

In a previous version of this work [8] raycasting with bounding box culling was used to detect collisions between vertices and body parts. This is a computationally costly method since a ray is sent from every tested vertex to every face of the body part. In this paper, we use the point-tetrahedron collision test with spatial hashing method described in [45]. Given a tetrahedral mesh, this method defines a hash-function that maps every object (vertices and tetrahedrons) to a 1D index. The function is designed such that objects mapped to the same index are the ones located in the same region in 3D space, and must be tested for collision. This allows to significantly reduce the number of collision tests to be performed, and has been applied in real-time animation pipelines [28].

Employing this collision test requires a tetrahedral mesh, which we create by adding the centroid of the body part to each of its triangles. We do the same to triangles created at the seams in Section 4.2.1 to have a closed mesh. This approx-

imation holds for convex objects, which is the case for most of our segmented body parts. However, hands and feet are not convex, due to fingers and toes, and our approximation could consequently lead to important errors. In our experiments, we mostly use poses from SMPL that do not encode the movement of the fingers or toes, so the approximation did not generate artifacts. If more detailed finger or toe poses are required, it is possible to add each phalanx of the fingers and toes to the body part segmentation to make the approximation more robust.

Although this method has not been designed for elongated tetrahedra, such as those needed to represent some long body parts, we obtained significantly higher performance than the raycasting approach. For example, for the mesh with important self collision shown in Figure 4 (center), the new contact detection algorithm takes 1 second to compute the repulsion energy, whereas the raycasting method takes 5 seconds. More examples and details about computation times are given in Section 7.2.

4.3.2. Attraction Term

Our contact energy term should also preserve contacts between surfaces present in the source's desired pose. These contacts give important semantic meaning to the pose. In a preliminary step, we encode those contacts in the source. We define a contact threshold proportional to the height of the character. Vertices that are under this threshold distance from a surface are considered in contact with the surface. As for interpenetrations, we consider that no important contacts should appear inside a same body part, and thus only encode contacts between different body parts. For each vertex under the contact threshold distance of a surface, we encode the contact as the couple of the vertex and its closest vertex on the surface. The attraction term also forces the vertices at ground level to stay at ground level in the result.

The attraction term increases when the distance between vertices in contact in the source exceeds the fixed contact threshold as:

$$E_a(\mathbf{V}) = \gamma_a \sum_{(v_i, v_j) \in C} \max[(d(v_i, v_j) - T), 0]^2 + \gamma_{ag} \sum_{v_i \in C_G} \max[(d(v_i, G) - T), 0]^2, \quad (9)$$

where C is the set of pairs of vertices in contact, C_G are the vertices in contact with the ground, and T is the contact threshold. The effect of the attraction term is illustrated in Figure 4b.

4.3.3. Rigid formulation

The contact energy (Equation 7) aims to preserve a coherent pose of the subject, i.e. a pose of the source character that preserves contacts. As such, the term should be minimized by modifying the source pose parameters.

To do so, we use the body part segmentation described in Section 4.2.1. Body parts are ordered in a tree hierarchy, with the crotch as the root. We then define a rotation for each body part $\Theta = \{R_P\}_{P \in BP}$. These rotations are applied to a body part and its children, around a "joint" defined as the centroid of the

seam between the body part and its parent (see Figure 3b). The root body part rotates around its centroid.

By minimizing the contact energy w.r.t. these rotations, each body part deforms rigidly. The contact energy becomes:

$$E_C(\mathbf{V}(\Theta)) = \gamma_r E_r(\mathbf{V}(\Theta)) + \gamma_a E_a(\mathbf{V}(\Theta)), \quad (10)$$

using the mesh vertex positions \mathbf{V} as functions of the rotations Θ .

5. Iterative Solving

Optimizing the full sum of energies in expression 1 appears difficult in practice since the shape term $E_{Shape}(\mathbf{V})$ (see expression 3) is non-linear. This results from the fact that the differential coordinates ω_i that encode the shape are expressed in a local coordinate system, which depends on the position of the vertices of the mesh. Hence moving a vertex also transforms its local frame. We therefore minimize expression 1 iteratively. In a first step vertices \mathbf{V} are moved with respect to the target shape information, then the pose Θ is optimized in order to satisfy the contact constraints and finally local frames are re-estimated. This is iterated until the absolute difference in the sum of energies between two successive iterations is below a threshold. The first two steps are detailed below.

The first step aims to optimize the shape fidelity term $E_{Shape}(\mathbf{V})$ and the volume preservation term $E_{Vol}(\mathbf{V})$, expressions 3 and 6 respectively. To this purpose vertices are moved in a direction that accounts for both terms:

$$v'_i = v_i + \epsilon(\gamma_s d_s(v_i) + \gamma_v d_v(v_i)), \quad (11)$$

with v'_i the new position of v_i , γ_s and γ_v the weights associated to the directions d_s and d_v , respectively, and ϵ a displacement offset function. The shape direction d_s is the direction towards the optimal position as defined in Equation 4. The volume direction d_v is computed based on the Stokes' Theorem and its resulting divergence theorem. That is, d_v is the direction of the normal n_i of the surface at vertex v_i , and the offset by which we move v_i is the difference in volume of the body part containing v_i between the target and the current shape. This leads to $d_v(v_i) = (V_p^T - V_p) n_i$.

The second step of the iterative framework aims to minimize the contact energy defined in expression 10. Auto-differentiation is used to obtain the gradient of the contact energy E_C w.r.t. rotations of the body parts Θ . We then apply a gradient descent iteration to the rotations. Since the deformation at each iteration is relatively small, this slight correction is enough.

6. Adaptation to continuous motion sequences

The approach presented in the previous sections enables to transfer the shape of a target character to a source character in a given static pose. When considering motion sequences, a per frame strategy can be applied with however no guarantee of temporal continuity in the resulting animations. While our experiments demonstrate that such a strategy provides already

good results when retargeting motions between shapes that do not differ significantly (see for instance the running sequence in the accompanying video), discontinuities together with inconsistent ground contacts can appear with shapes that require more important pose corrections. To tackle this issue we introduce a post-processing step that enforces temporal consistency in the retargeted animations. Details are given below and results presented in section 7.6.

6.1. Temporally Consistent Ground Contact

The pose correction applied to shapes can give rise to inconsistent ground contacts. For example, when transferring motion from a skinny character to a corpulent one, the gap between the legs can be widened to avoid potential collisions between the thighs. When such a corrections occurs, the feet positions can deviate in consecutive frames, leading to the so-called "foot-skating" artifacts in the resulting animation.

To fix this problem, we modify the attraction term in Equation 9. For static data, ground contacts are enforced by constraining the concerned vertex heights to be at the ground level. Although this avoids collision or loss of contact between the foot and the ground, it does not guarantee the foot to remain at a fixed position, hence yielding foot-skating artifacts. For a continuous sequence of poses, we store the ground contact position of the source pose at each frame and compare it with the previous frame. When detecting a ground contact that was already a ground contact in the previous frame, we consider that the associated vertex should remain at the same position. Therefore, the ground contact term for such a vertex v_i becomes: $\max[(d(v_i^t, v_i^{t-1}) - T), 0]^2$ where t denotes to the current frame time within the motion sequence, and v_i^t (respectively v_i^{t-1}) corresponds to the vertex v_i at the frame time t (respectively $t-1$).

6.2. Animation Smoothing

The pose correction can also induce jitter in the resulting animation. Under the continuity assumption we experimented, in a previous version of this work [8], a 5-frame rolling average that reduces jitter, yet retaining residual oscillations. In this work, we explore and evaluate two other post-processing filtering approaches: a global approach that treats the full sequence jointly, and a local approach that operates on the motion trajectory of each vertex independently.

6.2.1. Discrete Cosine Transform

The first approach smooths globally the retargeted animation by filtering high frequencies in the temporal domain. It builds on the work of Akhter et al. [1] that shows that the PCA basis learned from human motion sequences converges to the basis of the Discrete Cosine Transform (DCT). This property was used to combine a DCT basis for temporal data with a spatial shape basis computed using PCA to create a model encoding spatio-temporal data [2]. This model allows to smooth motion sequences by removing first the high frequencies of the temporal DCT basis, and second the basis vectors of the PCA shape space corresponding to small eigenvalues. In our experiments, we apply this approach to retargeted animation sequences. Since the sequences we consider are relatively short, we only project over the DCT basis and remove high frequencies.

6.2.2. De Boor Spline Approximation

The second approach smoothes locally the retargeted animation by post-processing the trajectory of each mesh vertex independently. To this end, each trajectory is approximated by a spline [15], that have shown effective in trajectory smoothing [17]. We use the original de Boor algorithm [16] to approximate vertex trajectories with a regularization term based on the curve second derivatives.

7. Evaluation and Discussion

In this section, we present results of our method and discuss its strengths and limitations. We introduce first the data used in our evaluations and give the implementation details. We further show the evaluations for minimally dressed humans, casually dressed humans, wild animals, and animations, and we compared our method to state-of-the-art works. Finally, we discuss the method's limitations.

Quantitatively evaluating the results of motion retargeting remains an open problem and is especially challenging when handling generic characters. For this reason, all the evaluations and comparisons provided in what follows are qualitative. For better illustrations, especially for results on motion sequence retargeting, the reader is invited to refer to the supplemental video.

7.1. Data

To demonstrate the generality of our approach, we evaluate our method on two different shape classes. First, human characters, both in a minimally dressed scenario and in a casually dressed one. This class is the most commonly considered in retargeting applications, and the minimally dressed scenario allows in particular comparisons to the state of the art. The second class of shapes we consider are wild animals. Wild animal shapes are interesting as they can exhibit very different morphologies while still respecting our assumption of near-isometrically deformations during motion.

As input, we require source and target character meshes in correspondence. For humans, the correspondence is established using the SMPL template (6890 vertices and 13776 faces) [37]. This template is segmented into the 17 body parts shown in Figure 3a. For animal models, the correspondence is established using the SMAL template (3889 vertices and 7774 faces) [53]. We segment this template into 24 body parts as shown in Figure 8.

For minimally dressed human characters, we use the example animations provided with SMPL, Faust [9], Dyna [41], and models from Liu et al. [36] fitted to the SMPL template. For dressed humans we use meshes from 3D Poses in the Wild [48] that are already fitted to the SMPL template. For animal models, we create different poses and shapes using the statistical model SMAL.

7.2. Implementation Details

We implemented our algorithm in Julia, and use a python implementation [42] of de Boor's smoothing algorithm to post-process the trajectory of each mesh vertex.

1 7.2.1. Parameter Settings

2 Our method has a number of weight parameters that need to
 3 be adjusted. In this work, we set them empirically based on a
 4 few examples of minimally dressed human bodies. Applying
 5 these fixed parameters to all human and animal examples we
 6 have tested leads to visually pleasing results.

7 In practice, the parameters are set as follows. The parame-
 8 ters weighing the different energy terms in Equation 1 are set
 9 to $\gamma_{Shape} = \gamma_{Vol} = \gamma_C = 1$, the parameters weighing the influ-
 10 ence of the contact and repulsion terms in the contact energy
 11 of Equation 7 are set to $\gamma_r = \gamma_a = 1$, and the weights han-
 12 dling ground contact in Equations 8 and 9, respectively, are set
 13 to $\gamma_{rg} = \gamma_{ag} = 0.1$. The offset weight ϵ in Equation 11 is set to
 14 0.3, and the parameter p employed during spline smoothing to
 15 0.1.

16 7.2.2. Computation Times

17 All experiments were run on a PC with an Intel Xeon E5-
 18 2623 v3s and 32GB of RAM. The computation time is highly
 19 dependent on the surface interactions present in the pose trans-
 20 fer, i.e. contacts and possibly colliding surfaces. The new con-
 21 tact detection method improves the computation time. We now
 22 discuss the computation time needed for a single frame in differ-
 23 ent scenarios. When the deformed pose is free of any body-to-
 24 body interactions, the method requires around 5 minutes. In the
 25 example shown in Figure 5, some corrections are needed due to
 26 body-to-body surface collisions, and our method takes around
 27 15 minutes with the new collision detection, compared to 20
 28 minutes using the previous raycasting approach. In the example
 29 show in Figure 14b, contacts in the deformed pose of the source
 30 need to be maintained, and our method takes around 20 minutes
 31 with the new collision detection, compared to 30 minutes with
 32 the previous approach. In the example shown in Figure 14c, the
 33 transfer needs to both maintain contacts in the deformed pose
 34 and avoid surface collisions, and our method takes around 24
 35 minutes with the new collision detection, compared to 44 min-
 36 utes with the previous approach. All experiments applied on
 37 static animals models based on SMAL take less than 5 minutes
 38 for the transfer.

39 7.3. Minimally Dressed Humans

40 This section discusses the convergence behaviour of our
 41 method, and shows an example of transferring a pose with con-
 42 tact to different morphologies.

43 Figure 5 illustrates the iterative process of our method and
 44 shows intermediate results (5d). The shape and volume evolve
 45 quickly to match the target, while the contact term avoids inter-
 46 penetrations here by widening the gap between the legs and rais-
 47 ing the arms. Figure 5b shows the evolution of each energy term
 48 during this transfer. Note that the shape fidelity (Eq. 3) and vol-
 49 ume preservation (Eq. 6) terms decrease rapidly in the first iter-
 50 ations. The initial spike of the repulsion term (Eq. 8) is due to
 51 interpenetrations that appear as the morphology changes. The
 52 correction of interpenetrations causes loss of contacts around
 53 the armpits, explaining the slight increase in the attraction term
 54 (Eq. 9). Our iterative process efficiently minimizes the shape

and volume energies, while maintaining the contact energy at a
 55 reasonable level.

56 Figure 6 shows results of shape transfer from the source of
 57 Figure 12 (left) to characters with varying morphology. Notice
 58 the evolution of the space between the upper arms and the torso
 59 depending on the morphology of the target; while skinny char-
 60 acters have a large distance, this distance disappears for larger
 61 bodies.

62 7.4. Casually Dressed Humans

63 This section illustrates results of transferring poses from min-
 64 imally dressed humans to humans wearing casual clothing. Fig-
 65 ure 7 shows three frames obtained when retargeting poses from
 66 the sample animations provided with SMPL to clothed char-
 67 acters from 3DPW. In all three results, the cloth details, in-
 68 cluding wrinkles present in the standard pose of the target, are
 69 transferred to the deformed pose. Furthermore, the method can
 70 transfer hair, shown in Figures 7g and 7h, and even accessories
 71 such as the backpack and baseball cap in Figure 7i.

7.5. Animals

72 This section illustrates results that retarget poses between dif-
 73 ferent wild animals. Figure 9 shows results that retarget a pose
 74 of a fox to a lion and a hippopotamus. Despite important differ-
 75 ences in the morphology and volume distributions among body
 76 parts for the different animals, the resulting retargeted poses
 77 are plausible overall. Note that the characteristics of the heads,
 78 trunk and legs are maintained in the retargeted pose for the lion
 79 and the hippopotamus. However, some artifacts occur for body
 80 parts with smaller volume, such as the tail of the hippopotamus,
 81 which is elongated after the transfer. The reason for this is that
 82 we match the volume of the body parts to the target, but not
 83 their lengths: our method allows body parts to get elongated as
 84 long as their volume is correct.

85 Figure 10 shows results of transferring a pose with contact
 86 from a feline to a lion, a fox and a hippopotamus. The results
 87 show that the method is able to preserve contact constraints
 88 even for animals with significantly different morphologies.

89 7.6. Animations

90 This section compares the different strategies to adapt our
 91 method to continuous motion sequences, as introduced in Sec-
 92 tion 6, and shows some qualitative results. Table 1 provides
 93 quantitative measurements over the motion sequence for the
 94 static motion retargeting applied frame by frame, the DCT
 95 smoothing method with different percentages of low frequen-
 96 cies retained, and the spline-based smoothing method. The
 97 quantitative measures are (1) the displacement of a vertex lo-
 98 cated in the middle of the forehead between two consecutive
 99 frames (mean and standard deviation), and (2) the volume of the
 100 right forearm during the animation (mean and standard devia-
 101 tion). These measures are evaluated for the punching sequence
 102 shown in Figure 11. For a correct retargeting, we expect that the
 103 mean and standard deviation of the displacement of the vertex
 104 on the forehead are similar to the ones in the source animation,
 105 and that there is a low standard deviation of the measured vol-
 106 ume with its mean close to the volume measured on the target

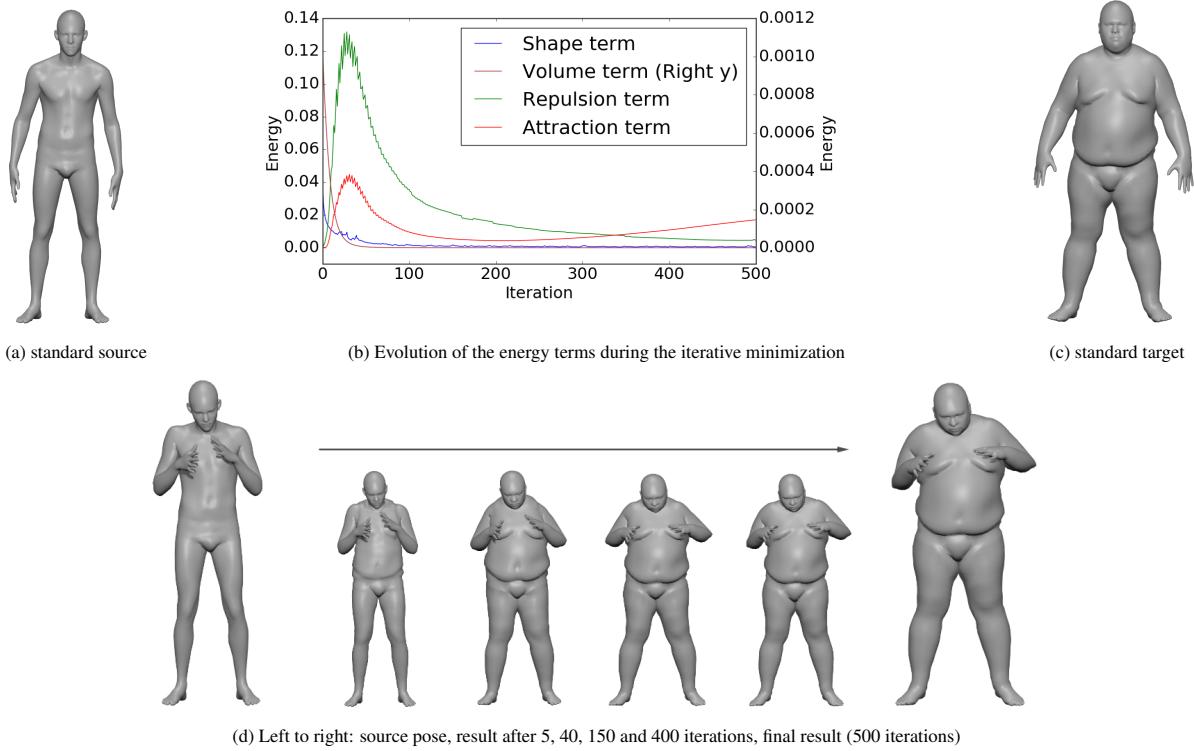


Fig. 5: Evolution of the shape transfer from a thin to a larger character through the iterations.

character. Without post-processing, the vertex on the forehead performs large motions that lead to jittering artifacts (visible in the supplemental video). When smoothing with DCT, keeping a high percentage of components does not remove the jitter on the head. When keeping a low percentage of components, however, artifacts such as volume shrinking on the arm appear, as can be seen by the increasing standard deviation of the arm volume. In contrast, the spline-based smoothing leads to a result without apparent jitter while preserving the arm volume. Figure 11 shows the corresponding motion retargeting result. Animation results shown in the supplementary video are thus smoothed using the splined-based method.

13

7.7. Comparisons

In this section, we compare results of our method with previous works. First, our results are compared to a skeleton-based approach where joint angles are directly applied to a new character. Second, we applied our method to character meshes used in previous surface mesh retargeting methods, namely context graphs [36] and AuraMesh [27], and compare our results to those obtained in these previous works.

Figure 12 compares our method to a skeleton retargeting baseline. The source pose (left) was generated by hand-tuning SMPL pose and shape parameters. Applying the same pose parameters to a character with different morphology leads the result of the baseline shown in the center. This straightforward approach leads to artifacts: the left hand enters the belly, and

	head displacement (cm)		arm volume (dm^3)	
	mean	std.	mean	std.
source animation	0.345	0.220	/	/
target character	/	/	2.315	/
no post-process	1.094	0.900	2.221	0.101
50% DCT	0.835	0.605	2.220	0.104
25% DCT	0.718	0.458	2.215	0.123
10% DCT	0.410	0.260	2.147	0.440
5% DCT	0.237	0.148	1.913	0.551
spline smoothing	0.553	0.343	2.175	0.176

Table 1: Comparisons between different smoothing approaches. Mean and standard deviation of the displacement of a vertex on the middle of the forehead between two consecutive frames, and of the volume of the right forearm, evaluated for the motion sequence in Figure 11.

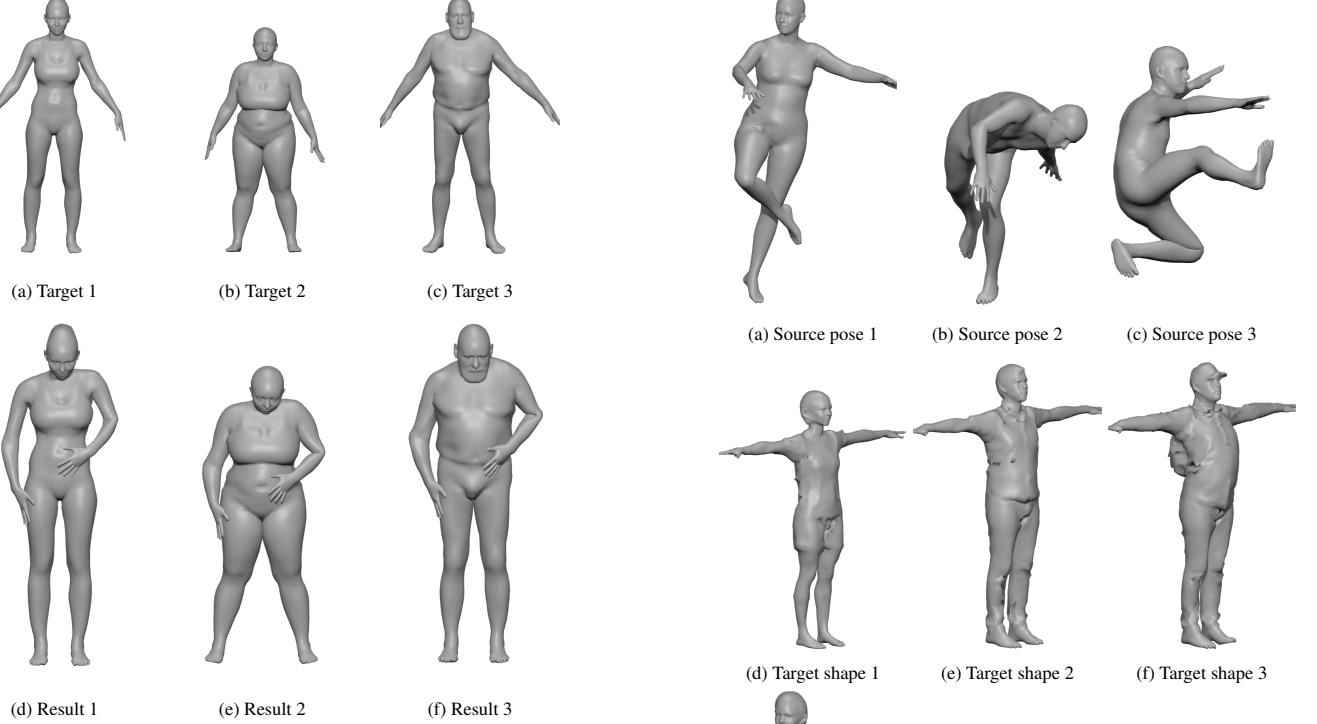


Fig. 6: Shape transfer results on several characters of the deformed source pose shown in Figure 12 (left).

the contact between the right hand and the hip is incorrect. The result of our method is presented in the right of the figure. The artifacts reported with the baseline do not occur. Moreover, notice that the space between the arms and the body shrinks during the transfer. This demonstrates that the method was able to find a solution without artificially spreading the arms far from the torso to preserve the distances associated with the thin source character.

Figure 13 depicts results obtained with our method when applied to 3D models used in [36]. Our results are compared to those obtained by an artist (artist performance initially reported in [36, Figure 6]). Note that even with a relatively large change in morphology, our result is close to the solution proposed by an artist. In particular, when viewed from above, one can see that the artist created new contacts between the arms and the body. These additional contacts did not change the contextual meaning of the pose, but have been introduced to adapt to the morphology of the target character. These additional contacts have also been mostly recovered by our method, compared to the context graph method, which aims to preserve distances observed with the source character.

Figure 14 applies our method on a shoulder rubbing pose that is similar to the one used in AuraMesh [27, Figure 8]. We see that our method preserves the hand/shoulder contact, even with important changes of morphology. Notice that for a close morphology (Figure 14b), the distance between the elbow and the torso does not significantly change in the result. However, for drastically larger target characters (Figures 14c and 14d) this distance shrinks or even disappears to create new contacts. These pose changes do not alter the contextual meaning but are

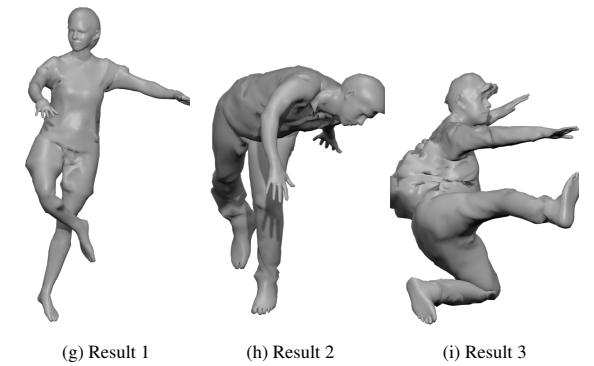


Fig. 7: Results of the method from sample poses of SMPL to clothed characters of 3DPW.

required to keep the morphology consistent. For the same kind of example, AuraMesh aims at preserving the initial distances observed with the source character, which is unlikely to adapt to a larger target character and contrary to our results.

7.8. Discussion

In this paper we explore the strategy of shape transfer, as an alternative to the widely adopted pose transfer strategy, to address retargetting problems. While our results validate this approach, it suffers anyway from some limitations. First, the method may appear slow compared to recent methods that can be close to real-time. This results from the efficient but computationally expensive optimization framework that is used and does not invalidate the shape transfer principle. Directions for improvements on this aspect include parallelized implementations as well as alternatives to optimization with, for instance, learning methods.

Another potential limitation is that our method requires correspondences between the input meshes along with a body part

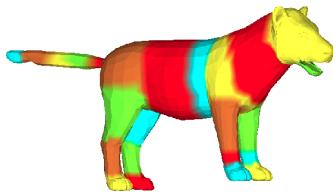


Fig. 8: Body part segmentation of the SMAL template

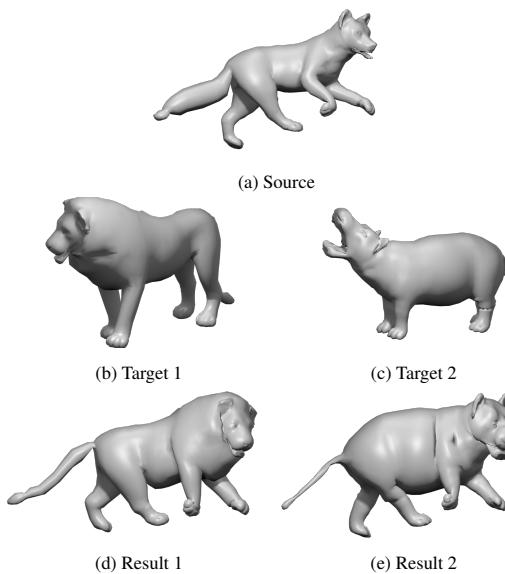
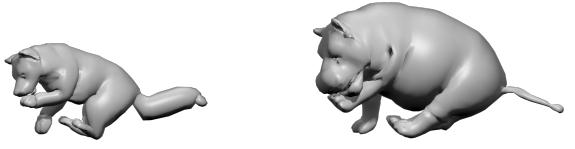


Fig. 9: Shape transfer results on animal models taken from SMAL.



(b) Result for the target shown in Figure 9b



(c) Result for the target shown in Figure 9a (d) Result for the target shown in Figure 9c

Fig. 10: Shape transfer results on several animals from a paw licking pose of a feline.

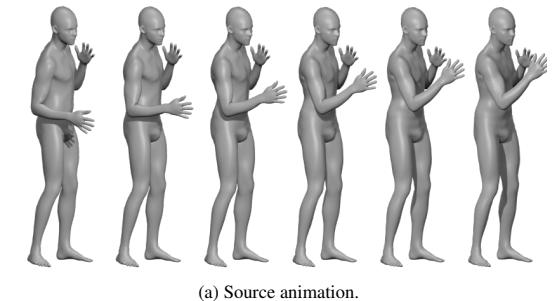


Fig. 11: Result of transferring a punching animation to the target in Figure 5c, using the spline smoothing in post-processing.

1 segmentation. In our experiments, we tackled this problem by
2 fitting a template equipped with a pre-defined segmentation to
3 the input meshes. This pre-processing is light and only needs
4 to be applied once per mesh. Moreover it could be made automatic,
5 using some template fitting method such as [46]. This
6 is hence simpler and faster than standard rigging processes tradi-
7 tionally used in computer animation. The body part segmen-
8 tation only needs to be created once per template, making its
9 pre-processing complexity negligible in the long term.

10 The contextual meaning of the human posture introduced in
11 this work raises interesting questions. By contextual meaning
12 we assume additional posture elements which have priority over
13 traditional pose parameters represented by the body part rigid
14 transformations. In this work we focus on preserving contacts
15 in the source pose as well as preventing inter-penetrations, this
16 by allowing pose parameters to evolve. Our results show that
17 this strategy genuinely preserves part of the posture context. In
18 particular, it enables to introduce new contacts that are induced
19 by the target shape, especially when retargeting from a skinny
20 to a large character. However, when transferring from a large to
21 a skinny character, contacts should not necessarily be preserved
22 as they may not be natural. This raises the question of which
23 contacts should be preserved in a transfer between characters
24 and more generally on how to model the contextual meaning of
25 a human posture. Apart from user inputs on a per case basis, an
interesting direction is to explore here data-driven approaches

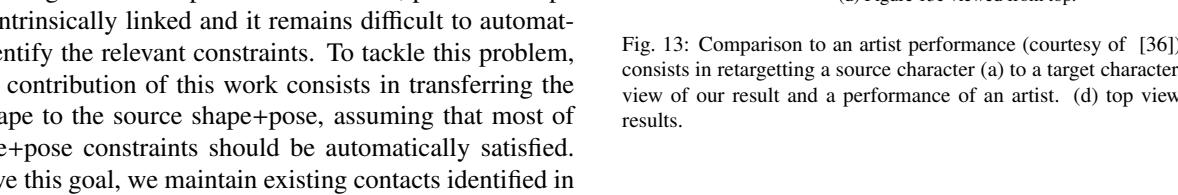
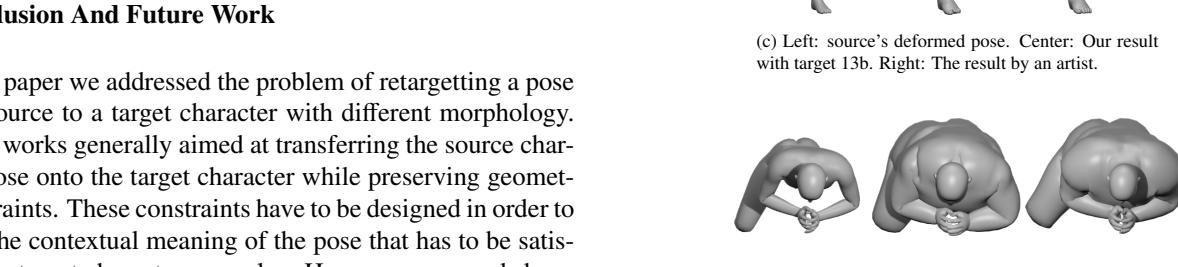
26 aiming at learning such a contextual meaning.

Finally, the method described in this paper is considering a pose independently from the previous and next frames. The advantage of this strategy is to avoid drift in the estimation and hence to gain robustness. The drawback is the potential inconsistency over time. To remedy this we additionally filter the results in a post-processing step, with clear benefits as shown in the accompanying video. However, such filtering is based on a continuous assumption and is therefore not able to handle potential dynamic movements of the surface, such as clothes deformation during motion. This dynamic behavior of the surface has not been addressed in this paper. Future works could explore how to simulate this dynamic property of the surface accordingly to the body motion.

27
28
29
30
31
32
33
34
35
36
37
38
39
40



Fig. 12: Comparison to a skeleton retargeting baseline. Left: The source deformed pose generated by manually tuning SMPL shape and pose parameters. Center: The same SMPL pose parameters applied to new shape parameters. Right: The result with our method.



8. Conclusion And Future Work

In this paper we addressed the problem of retargetting a pose from a source to a target character with different morphology. Previous works generally aimed at transferring the source character's pose onto the target character while preserving geometric constraints. These constraints have to be designed in order to capture the contextual meaning of the pose that has to be satisfied in the target character pose also. However, pose and shape may be intrinsically linked and it remains difficult to automatically identify the relevant constraints. To tackle this problem, the main contribution of this work consists in transferring the target shape to the source shape+pose, assuming that most of the shape+pose constraints should be automatically satisfied. To achieve this goal, we maintain existing contacts identified in the source pose while avoiding inter-penetrations during shape transfer. Our results (e.g. figure 13) show that this method enables us to find new contacts linked to the target shape, which is generally impossible to address when preserving distances between joints, as usually proposed in previous works.

Whereas the core of the method has been designed to consider static poses [8], another contribution of this paper is to add post-processing to deal with continuous motion sequences. Future work could include temporal information directly in the core of the method instead of introducing postprocessing step.

Our method showed satisfying results when transferring isolated poses and motion from a source character to another one with different morphology (e.g. from a thin to a big character or from a tall man to a small woman). We also extended previous experiments [8] to apply the method to casually dressed humans and wild animals. For all types of shapes, the same parameter settings were used without adjustment. These new experiments demonstrate the robustness and generality of our shape transfer approach. To handle new creatures or more accurate models (including fingers for example), the user simply has to segment the 3D mesh into the new required body parts. Moreover, to address meshes with different topology, preprocessing is necessary to calibrate a template model into both the source and target models, before applying the method.

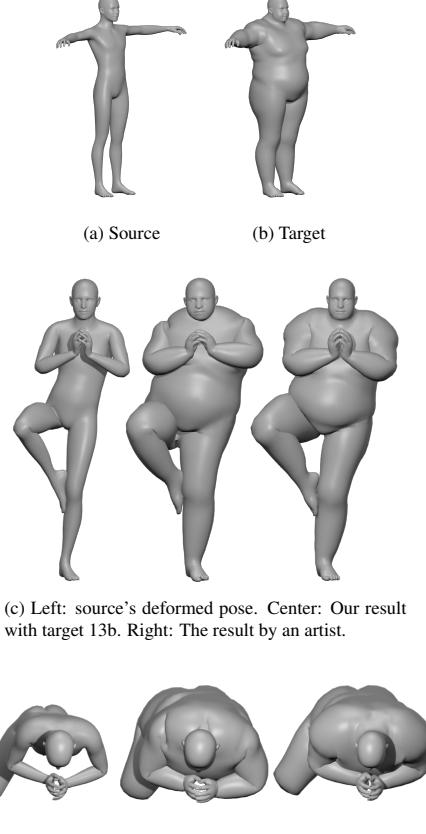


Fig. 13: Comparison to an artist performance (courtesy of [36]). The results consists in retargetting a source character (a) to a target character (b). (c) front view of our result and a performance of an artist. (d) top view of the same results.

From the fundamental point of view, this work contributes to better automatically retarget motion from one source to a target character, without requiring tedious rigging manual processing. However, there remains a series of challenges that have not been addressed in this approach. Indeed, although this approach enabled us to better handle contacts between body parts, future works are necessary to capture more contextual meaning information, such as preserving body segment orientation or distances. Even for contact constraints, there are some open questions, such as the release of contact constraints when transferring a pose from a fat character to a thin one. The development of new machine learning approaches opens new directions to better capture and transfer the contextual meaning information from one shape to another. Hence, a promising future direction would consist in using a large database of characters with different morphology performing equivalent poses to train a machine learning retargetting system.

Acknowledgments

This project was supported by the Inria IPL AVATAR project. We also thank Ludovic Hoyet and Sergi Pujades for their help in gathering and fitting the models from [36] to SMPL.

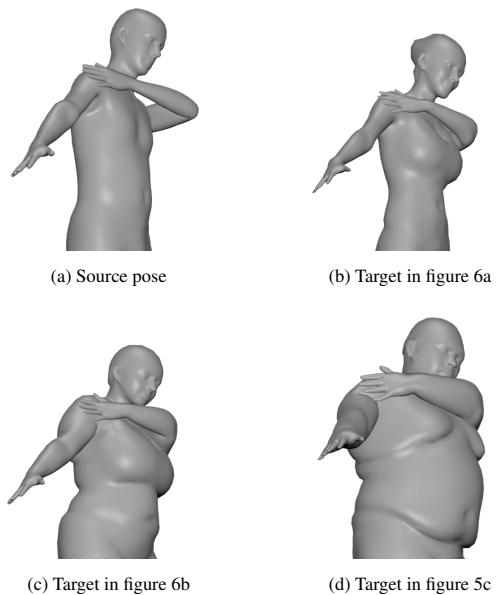


Fig. 14: Comparison to the AuraMesh method [27] on a shoulder rubbing pose.

References

- [1] Ijaz Akhter, Yaser Sheikh, Sohaib Khan, and Takeo Kanade. 2010. Trajectory space: A dual representation for nonrigid structure from motion. *PAMI* 33, 7 (2010), 1442–1456.
- [2] Ijaz Akhter, Tomas Simon, Sohaib Khan, Iain Matthews, and Yaser Sheikh. 2012. Bilinear spatiotemporal basis models. *ToG* 31, 2 (2012), 1–12.
- [3] Rami Ali Al-Asghar, Taku Komura, and Myung Geol Choi. 2013. Relationship Descriptors for Interactive Motion Adaptation. In *ACM SIGGRAPH/SCA (SCA '13)*. 45–53.
- [4] Mazen Al Borno, Ludovic Righetti, Michael J Black, Scott L Delp, Eugene Fiume, and Javier Romero. 2018. Robust Physics-based Motion Retargeting with Realistic Body Shapes. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 81–92.
- [5] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. 2005. SCAPE: shape completion and animation of people. In *ACM TOG*, Vol. 24. 408–416.
- [6] George Baciu and Bartholomew K. C. Iu. 2006. Motion retargeting in the presence of topological variations. *Computer Animation and Virtual Worlds* 17, 1 (2006), 41–57.
- [7] Ilya Baran, Daniel Vlasic, Eitan Grinspun, and Jovan Popović. 2009. Semantic deformation transfer. In *ACM TOG*, Vol. 28. 36.
- [8] Jean Basset, Stefanie Wuhrer, Edmond Boyer, and Franck Multon. 2019. Contact Preserving Shape Transfer For Rigging-Free Motion Retargeting. In *MIG*. ACM, 1–10.
- [9] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. 2014. FAUST: Dataset and evaluation for 3D mesh registration. In *CVPR*. IEEE.
- [10] Sofien Bouaziz, Yangang Wang, and Mark Pauly. 2013. Online modeling for realtime facial animation. *ToG* 32, 4 (2013), 1–10.
- [11] Adnane Boukhayma, Jean-Sébastien Franco, and Edmond Boyer. 2017. Surface motion capture transfer with gaussian process regression. In *CVPR*. 9.
- [12] Caroline Chan, Shiry Ginosar, Tinghui Zhou, and Alexei A Efros. 2019. Everybody dance now. In *ICCV*. IEEE, 5933–5942.
- [13] Kwang-Jin Choi and Hyeong-Seok Ko. 2000. Online motion retargetting. *The Journal Of Visualisation and Computer Animation* 2000 11, 5 (2000), 223–235.
- [14] Luca Cosmo, Mikhail Panine, Arianna Rampini, Maks Ovsjanikov, Michael M Bronstein, and Emanuele Rodolà. 2019. Isospectralization, or how to hear shape, style, and correspondence. *CVPR* (2019).
- [15] Peter Craven and Grace Wahba. 1978. Smoothing noisy data with spline functions. *Numerische mathematik* 31, 4 (1978), 377–403.
- [16] Carl De Boor. 1978. *A practical guide to splines*. Vol. 27. Springer-Verlag New York.
- [17] Magnus Egerstedt and Clyde F Martin. 2001. Optimal trajectory planning and smoothing splines. *Automatica* 37, 7 (2001), 1057–1064.
- [18] Patrick Esser, Ekaterina Sutter, and Björn Ommer. 2018. A variational u-net for conditional appearance and shape generation. In *CVPR*. IEEE, 8857–8866.
- [19] Lin Gao, Jie Yang, Yi-Ling Qiao, Yu-Kun Lai, Paul L Rosin, Weiwei Xu, and Shihong Xia. 2018. Automatic unpaired shape deformation transfer. In *SIGGRAPH Asia 2018*. ACM, 237.
- [20] Michael Gleicher. 1998. Retargetting Motion to New Characters. In *Siggraph (SIGGRAPH '98)*. ACM, 33–42.
- [21] Yana Hasson, Güray Varol, Dimitrios Tzionas, Igor Kalevatykh, Michael J Black, Ivan Laptev, and Cordelia Schmid. 2019. Learning joint reconstruction of hands and manipulated objects. *CVPR* (2019).
- [22] Chris Hecker, Bernd Raabe, Ryan W. Enslow, John DeWeese, Jordan Maynard, and Kees van Prooijen. 2008. Real-time Motion Retargeting to Highly Varied User-created Morphologies. *ACM TOG* 27, 3 (Aug. 2008), 27:1–27:11.
- [23] Edmond S. L. Ho, Taku Komura, and Chiew-Lan Tai. 2010. Spatial Relationship Preserving Character Motion Adaptation. *ACM TOG* 29, 4 (July 2010), 33:1–33:8.
- [24] Edmond S. L. Ho, He Wang, and Taku Komura. 2014. A Multi-resolution Approach for Adapting Close Character Interaction. In *ACM VRST (VRST '14)*. 97–106.
- [25] Chun-Hao Huang, Edmond Boyer, and Slobodan Ilic. 2013. Robust human body shape and pose tracking. In *3DV*. IEEE, 287–294.
- [26] Boyi Jiang, Juyong Zhang, Jianfei Cai, and Jianmin Zheng. 2019. Learning 3D Human Body Embedding. *ArXiv* abs/1905.05622 (2019).
- [27] Taekil Jin, Meekyoung Kim, and Sung-Hee Lee. 2018. Aura Mesh: Motion Retargeting to Preserve the Spatial Relationships between Skinned Characters. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 311–320.
- [28] Martin Komaritzan and Mario Botsch. 2019. Fast Projective Skinning. In *MIG*. ACM, 1–10.
- [29] Richard Kulpa and Franck Multon. 2005. Fast inverse kinematics and kinetics solver for human-like figures. In *IEEE-RAS Humanoids*. 38–43.
- [30] Richard Kulpa, Franck Multon, and Bruno Arnaldi. 2005. Morphology-independent representation of motions for interactive human-like animation. *Computer Graphics Forum* 24, 3 (2005), 343–351.
- [31] Benoît Le Calennec and Ronan Boulic. 2006. Robust Kinematic Constraint Detection for Motion Data. In *ACM SIGGRAPH/SCA (SCA '06)*. 281–290.
- [32] Thibaut Le Naour, Nicolas Courty, and Sylvie Gibet. 2019. Skeletal mesh animation driven by few positional constraints. *Computer Animation and Virtual Worlds* (2019), e1900.
- [33] Jehee Lee and Sung Yong Shin. 1999. A hierarchical approach to interactive motion editing for human-like figures. In *Siggraph*. 39–48.
- [34] Yaron Lipman, Daniel Cohen-Or, Ran Gal, and David Levin. 2007. Volume and shape preservation via moving frame manipulation. *ACM TOG* 26, 1 (2007), 5.
- [35] Wen Liu, Zhixin Piao, Jie Min, Wenhan Luo, Lin Ma, and Shenghua Gao. 2019. Liquid warping GAN: A unified framework for human motion imitation, appearance transfer and novel view synthesis. In *ICCV*. IEEE, 5904–5913.
- [36] Zhiguang Liu, Antonio Mucherino, Ludovic Hoyet, and Franck Multon. 2018. Surface Based Motion Retargeting by Preserving Spatial Relationship. In *MIG (MIG '18)*. ACM, Article 7, 7:1–7:11 pages.
- [37] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. 2015. SMPL: A skinned multi-person linear model. *ACM TOG* 34, 6 (2015), 248.
- [38] Etienne Lyard and Nadia Magnenat-Thalmann. 2008. Motion adaptation based on character shape. *Computer Animation and Virtual Worlds* 19, 3–4 (2008), 189–198.
- [39] Eray Molla, Henrique Galvan Debarba, and Ronan Boulic. 2018. Ego-centric Mapping of Body Surface Constraints. *IEEE TVCG* 24, 7 (2018), 2089–2102.
- [40] Jean-Sébastien Monzani, Paolo Baerlocher, Ronan Boulic, and Daniel Thalmann. 2000. Using an Intermediate Skeleton and Inverse Kinematics for Motion Retargeting. *Computer Graphics Forum* 19, 3 (2000), 11–19.
- [41] Gerard Pons-Moll, Javier Romero, Naureen Mahmood, and Michael J. Black. 2015. Dyna: A Model of Dynamic Human Shape in Motion. *ACM*

- 1 *TOG, (Proc. SIGGRAPH) 34, 4 (2015), 120:1–120:14.*
- 2 [42] Eugene Prilepin. 2017-2020. csaps Cubic Spline Approximation. <https://csaps.readthedocs.io/en/latest/index.html>.
- 3 [43] Helge Rhodin, James Tompkin, Kwang In Kim, Edilson De Aguiar,
4 Hanspeter Pfister, Hans-Peter Seidel, and Christian Theobalt. 2015. Gen-
5 eralizing wave gestures from sparse examples for real-time character con-
6 trol. *ACM TOG* 34, 6 (2015), 181.
- 7 [44] Robert W Sumner and Jovan Popović. 2004. Deformation transfer for
8 triangle meshes. In *ACM TOG*, Vol. 23. 399–405.
- 9 [45] Matthias Teschner, Bruno Heidelberger, Matthias Müller, Danat Pomer-
10 antes, and Markus H Gross. 2003. Optimized Spatial Hashing for Colli-
11 sion Detection of Deformable Objects.. In *Vmv*, Vol. 3. 47–54.
- 12 [46] Gul Varol, Duygu Ceylan, Bryan Russell, Jimei Yang, Ersin Yumer, Ivan
13 Laptev, and Cordelia Schmid. 2018. Bodynet: Volumetric inference of
14 3d human body shapes. In *Proceedings of the European Conference on
15 Computer Vision (ECCV)*. 20–36.
- 16 [47] Ruben Villegas, Jimei Yang, Duygu Ceylan, and Honglak Lee. 2018.
17 Neural kinematic networks for unsupervised motion retargetting. In *Pro-
18 ceedings of the IEEE Conference on Computer Vision and Pattern Recog-
19 nition*. 8639–8648.
- 20 [48] Timo von Marcard, Roberto Henschel, Michael J Black, Bodo Rosen-
21 hahn, and Gerard Pons-Moll. 2018. Recovering accurate 3d human pose
22 in the wild using imus and a moving camera. In *ECCV*. 601–617.
- 23 [49] Jungdam Won and Jehee Lee. 2019. Learning body shape variation in
24 physics-based characters. *ACM Transactions on Graphics (TOG)* 38, 6
25 (2019), 1–12.
- 26 [50] Stefanie Wuhrer, Chang Shu, and Pengcheng Xi. 2012. Posture-invariant
27 statistical shape analysis using Laplace operator. *Computers & Graphics*
28 36, 5 (2012), 410–416.
- 29 [51] Cha Zhang and Tsuhan Chen. 2001. Efficient feature extraction for 2D/3D
30 objects in mesh representation. In *ICIP*, Vol. 3. IEEE, 935–938.
- 31 [52] Kun Zhou, Weiwei Xu, Yiyi Tong, and Mathieu Desbrun. 2010. Defor-
32 mation transfer to multi-component objects. In *Computer Graphics Fo-
33 rum*, Vol. 29. Wiley Online Library, 319–325.
- 34 [53] Silvia Zuffi, Angjoo Kanazawa, David Jacobs, and Michael J. Black.
35 2017. 3D Menagerie: Modeling the 3D Shape and Pose of Animals. In
36 *CVPR*. IEEE.
- 37