Contact Preserving Shape Transfer For Rigging-Free Motion Retargeting

Jean Basset, Stefanie Wuhrer, Edmond Boyer, Franck Multon

HAL Id: hal-02293308
https://hal.science/hal-02293308v3

Submitted on 23 Oct 2019

Contact Preserving Shape Transfer For Rigging-Free Motion Retargeting

Jean Basset
jean.basset@inria.fr
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP*, LJK,
38000 Grenoble, France

Stefanie Wuhrer
stefanie.wuhrer@inria.fr
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP*, LJK,
38000 Grenoble, France

Edmond Boyer
edmond.boyer@inria.fr
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP*, LJK,
38000 Grenoble, France

Franck Multon
fmulton@irisa.fr
Univ Rennes, Inria, CNRS IRISA, M2S
France

ABSTRACT

Retargeting a motion from a source to a target character is an important problem in computer animation, as it allows to reuse existing rigged databases or transfer motion capture to virtual characters. Surface based pose transfer is a promising approach to avoid the trial-and-error process when controlling the joint angles. The main contribution of this paper is to investigate whether shape transfer instead of pose transfer would better preserve the original contextual meaning of the source pose. To this end, we propose an optimization-based method to deform the source shape+pose using three main energy functions: similarity to the target shape, body part volume preservation, and collision management (preserve existing contacts and prevent penetrations). The results show that our method is able to retarget complex poses, including several contacts, to very different morphologies. In particular, we introduce new contacts that are linked to the change in morphology, and which would be difficult to obtain with previous works based on pose transfer that aim at distance preservation between body parts. These preliminary results are encouraging and open several perspectives, such as decreasing computation time, and better understanding how to model pose and shape constraints.

CCS CONCEPTS

• **Computing methodologies → Animation**.

KEYWORDS

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

---

* Institute of Engineering Univ. Grenoble Alpes.
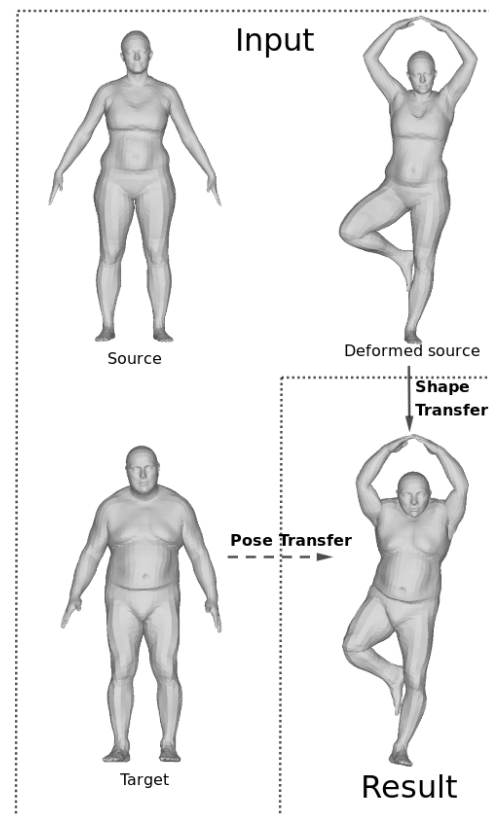
Figure 1: Overview of the deformation transfer algorithm. Classical works [Sumner and Popović 2004] transfer the pose of the source to the target. We propose a novel approach, transferring the shape of the target to the deformed source character.

1 INTRODUCTION

Animation studios have stored terabytes of animation files applied on various 3D characters, with meticulous association between skeletal motion and 3D shape. Traditional computer animation techniques consist in designing the 3D movement of the skeleton and then aligning it to the 3D shape (designed by other artists). This

rigging process also involves meticulously tuning skinning weights to ensure that the character envelope is deformed in accordance with the movement of the skeleton. For a given database of skeletal motion, this is a tedious and manual process that has to be done for any new character. Hence existing databases of rigged animations cannot be directly used, nor automatically recomputed, for new characters or projects, despite recent works aiming at automating part of the process [Avril et al. 2016]. Consequently the so-called retargeting problem, i.e. automatically transferring the motion of a source character (motion capture actor or previously edited motion) to a target one, is a strategic issue in the computer animation community.

Several previous works addressed the problem of skeletal motion retargeting, which consists in adapting the joint angles of the character in order to satisfy kinematic constraints either edited manually [Gleicher 1998; Kulpa et al. 2005] or automatically built based on geometric constraints between body parts [Ho et al. 2010]. However this control on the skeleton does not prevent artifacts, such as collisions, or more generally respect distance constraints of the character. For example, joint angles computed for a skinny character who is touching his belly cannot be directly applied to a fat character as the hands would penetrate the larger belly surface. To address these artifacts, traditional approaches rig the skeleton to the surface, control potential problems, and then apply the appropriate joint angle corrections.

An alternative to this process is to directly deform the surface of the character, without using rigged skeletons or joint angles, by transferring the deformation of one character motion to a new one. With the possibility to capture both shape and motion using multiple RGB or depth cameras, transferring both the shape and the motion from a source to a target character has become a promising approach. Hence, surface based motion transfer [Sumner and Popović 2004] usually transfers the pose of a source character to a target one. In these approaches, the key point is to encode the pose independently from the shape of the characters. Indeed, in the field of statistical shape analysis, the body shape and the pose are usually studied as distinct elements [Anguelov et al. 2005], assuming that changes in pose and in body shape are independent. This is a strong assumption since the morphology, typically the corpulence, can impact the skeletal pose in practice. Skinny people can touch their opposite shoulder with a hand freely, without collisions or contacts between the arm and the chest, whereas characters with large arms and/or chests may need these contacts to achieve the same objective (see Figure 9). In this example, shape and motion are not independent: some of the contacts are linked to the pose (contact between the hand and the opposite shoulder) and others to the shape (surface contacts between the arm and the chest). These two types of surface contacts should be considered differently: the pose-based contacts have to be preserved when transferred to another character, which is not the case for shape-based contacts. All the methods based on building topological constraints between body parts model all these constraints together, which may lead to unrealistic retargeting.

Indeed, character poses are designed for a given shape, and modeling the pose independently of the shape is still an open problem. Contacts and more generally distance between body parts have

been widely explored but other types of constraints may also be important, e.g. orientation of or coordination between body parts.

In this work, to avoid the problem of extracting the relevant constraints associated with a pose, we propose to explore transferring the shape of a target character to the current source character in a given pose (see Figure 1). In other words, we propose to morph the shape of the source character to make it similar to the target character, in its current pose, while maintaining a plausible posture w.r.t. contacts between surfaces. This way, we assume that the resulting target shape+pose would continue to capture the properties of the source shape+pose. Hence, we expect that surface contacts due to shape differences could be handled (appear if necessary because of the specific shape) with this approach. This paper aims at exploring what situation could be handled with shape transfer instead of pose transfer, and what artifacts may still remain with this approach.

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 dynamic data (animations) in Section 6. Finally we present results of our method in Section 7 and conclude the paper with a discussion in Section 8.

2 RELATED WORK

Previous works explore two major approaches to transfer a pose from one source to a target character: tuning the joint angles of a skeleton to satisfy kinematic constraints, or directly modifying the character's surface without using rigged skeletons.

2.1 Skeletal motion retargeting

With the popularization of marker-based motion capture systems in the early 90's, the question of transferring an animation from an actor to a rigged character rapidly arose. Motion retargeting was then considered as solving kinematic constraints on joint positions and ensuring continuity using displacement maps [Choi and Ko 2000; Gleicher 1998; Lee and Shin 1999]. Another approach consists in defining a morphology-independent representation [Hecker et al. 2008; Kulpa et al. 2005] with efficient constraint solvers [Kulpa and Multon 2005], or in using an intermediate skeleton [Monzani et al. 2000].

All these methods generally use predefined kinematic constraints that should be manually tuned. Automatic kinematic constraint detection in the source motion has been proposed [Le Callennec and Boulic 2006] to automate the constraint editing problem. Most of these constraints consist in spatial relationship between body segments, which can be modeled as distance constraints [Al-Asqhar et al. 2013] or more generalized spatial relationship between joints [Baciu and Iu 2006]. 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 [Ho et al. 2010, 2014]. Recent works introduced the concept of egocentric plane to ensure that the instantaneous separating plane between each pair of body parts is transferred between the source

and target motion [Molla et al. 2018]. This leads to 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 surfaces instead of simplified body segments. Moreover, they consider that all the relative positions between body segment have to be preserved from source to target character, while this may not correspond to reality for constraints mostly linked to shape differences.

2.2 Surface mesh retargeting

As skeleton-based retargeting cannot accurately deal with the relationship between the surfaces of a character, another approach consists of directly acting on the mesh instead of controlling joint angles of a skeleton. Some of the approaches directly transfer the surface's pose to a target character, by displacing vertices of the mesh, without using a skeleton. Hybrid approaches propose to jointly control skeleton and mesh parameters when transferring the pose from a source to a target character.

2.2.1 Surface mesh Pose Transfer. Deformation transfer based on surface mesh aims at avoiding the rigging and skinning processes by directly modifying the vertices of the mesh instead of controlling a skeleton. These methods mostly encode the pose of the source character as the deformation of the surface mesh, and transfer it to the target character [Sumner and Popović 2004; Zhou et al. 2010]. Some works have explored semantic deformation transfer between surfaces enabling to adapt very different shapes in accordance with the pose encoded in the source surface mesh. These methods use a set of corresponding poses [Baran et al. 2009] or animations [Boukhayma et al. 2017] of the source and target characters to define semantically equivalent base motions, and to compute for a new source pose the corresponding target pose by combining these base motions. A recent work has studied the use of deep neural networks for deformation transfer [Gao et al. 2018], where training is performed on various poses of the source and target that do not need to be in correspondence. It is also possible to learn mappings between semantically different poses of humans and animals to interactively control animation generation [Rhodin et al. 2015]. These methods neither require skeletons not point-to-point correspondences between source and target. However, heavy pre-processing needs to be preformed for every pair of source and target characters. These methods also do not explicitly account for surface interactions.

Recent works have added constraints to encode the pose as spatial relationships between the surfaces of the body. Liu et al. [Liu et al. 2018] introduced the idea of context graph, an extension of interaction mesh [Ho et al. 2010] to deal with body surface instead of joint centers. In context graph, nodes are points placed on the body surface and edges model distance constraints in-between these points. Hence, this paper assumes that transferring the contextual meaning of a source pose to a target character consists in satisfying the distance constraints encoded in this graph, in the target surface mesh. Alternatively, Jin et al. [Jin et al. 2018] defined 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 mesh both define an arbitrary distance under which spatial relationships should be preserved, assuming that they encode the contextual meaning of the pose. However, distant interactions between body surface, beyond this distance threshold, could also embed contextual information. Moreover, as stated in the introduction section, some close interaction may not be related to the contextual meaning of the pose, but may be the consequence of the shape. Hence, for our Figure 9 example, these approaches would try to preserve the distance between the arm and the chest, whereas for a bigger character it would be natural for this distance to significantly shrink and for new contacts to appear. Both previous works also consider simplified external mesh to model contacts and consequently may miss the fine contacts between the detailed surface mesh of body segments.

2.2.2 Hybrid approaches mixing skeleton and mesh. Surface deformation of human body is strongly linked to the motion of the skeleton, and some authors proposed to use both to transfer a pose from a source to a target character. Molla et al. [Molla et al. 2018] used the skeleton to model the pose with joint angles, but added constraints (egocentric mapping to preserve the topology between body parts) applied on body surface approximation. Other methods used a complete surface mesh together with the skeleton [Huang et al. 2013; Le Naour et al. 2019] to control the surface mesh deformation while preserving the coherence with the skeleton topology. By respecting both skeletal and surface constraints, they are able to generate natural animations and poses. While deformation transfer is not the main contribution of these works, their methods can be adapted for this purpose to obtain convincing results [Le Naour et al. 2019]. However these works have not explored surface interactions, and important collisions can appear for deformation transfer to significantly different shapes.

3 METHOD OVERVIEW

Our goal is to make a target character mimic the pose of a source character. Hence the inputs of our problem are: the source character in a *standard pose* (e.g. A-pose), the same source character in the *deformed pose* we wish to mimic, and the target character in the standard pose (see Figure 1). These inputs are 3D surface meshes with point-to-point correspondence, along with a segmentation into body parts (see Section 4.2.1). Note that our method does not require a rigged skeleton. Starting from the source character in the deformed pose, our approach consists in morphing its surface to make its shape become the one of the target character, while preserving contacts of the source deformed pose (see Figure 1). This way, we aim at transferring the shape of the target character to the given deformed pose of the source, in contrast to previous works that transfer the pose of the source character to the target one.

Our method includes both a deformation of shape and a deformation of pose. We follow the common hypothesis that pose deformations of the human body can be modeled as near-isometric [Cosmo et al. 2019]. 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.

Our input meshes are defined as $\mathcal{V} = (\mathbf{V}, E)$, where E is the set of edges of the mesh and $\mathbf{V} = v_1, \ldots, v_n$ are the coordinates of the mesh's vertices. We define the rigid body part transformations $\Theta = \{R_P\}_{P \in BP}$, where R_P is the rotation associated with body part P. To perform the shape transfer, we search for the vertices' positions $\mathbf{V}$ and the rigid transformations Θ that minimize:

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

The terms E_{Shape} and E_{Vol} penalize the difference 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 deformations of pose, and therefore is 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. The method to iteratively minimize energy 1 is presented in Section 5.

4 ENERGY TERMS

To perform the shape transfer from a source to a target character, we minimize a set of energy functions: shape energy term transferring non-isometric deformations, volume energy term to ensure preserving the volume of each body part of the target character, and contact energy term to preserve relevant pose contacts, and avoid penetration between the body parts. These energy terms are detailed in the following.

4.1 Shape

A common hypothesis is that the deformation of pose of the human body is near-isometric [Cosmo et al. 2019]. If we define the shape of a character as a property that does not change with pose deformations, this means that two meshes in different poses but with the same shape should be near-isometric. This intuition is coherent with previous works that use isometries to define shape preserving deformations [Lipman et al. 2007]. To deform the shape of a source mesh to match that of a target, we thus want to make the source as close as possible to an isometry of the target. The hypothesis of an isometric shape is a simplification; indeed two poses of the same character are not perfectly isometric, due to anatomical properties such as the elasticity of the skin. We assume that this is not an important problem and that using this hypothesis as a soft constraint leads to realistic results.

To encode the shape of the target in an isometry-invariant way, we use the following shape representation for our computation [Wuhrer et al. 2012]. The key idea of this representation is to encode for each vertex offsets w.r.t its neighboring vertices in a local coordinate frame.

The representation first computes the uniform Laplacian matrix L of the template. Since the input meshes all follow this template, i.e. have the same number of vertices and connectivity, this matrix is the same for all our inputs. This matrix is used to compute for each vertex v_i the Laplacian offset Δ_{v_i} as

$$\begin{pmatrix} \Delta_{v_1} \\ \ldots \\ \Delta_{v_n} \end{pmatrix} = L \begin{pmatrix} v_1 \\ \ldots \\ v_n \end{pmatrix} = \begin{pmatrix} \sum_{v_j \in N_1(v_1)} \frac{1}{deg(v_1)} v_j - v_1 \\ \ldots \\ \sum_{v_j \in N_1(v_n)} \frac{1}{deg(v_n)} v_j - v_n \end{pmatrix}.$$
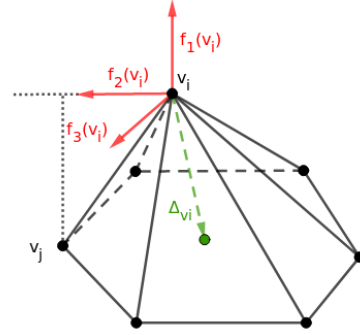

Figure 2: Local frame at v_i (red) and of Δ_{v_i} (green)

Finally, Δ_{v_i} is 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. It is invariant to translation and rotation of the neighbourhood of v_i. Thus, expressing Δ_{v_i} in this coordinate system makes the shape invariant to pose. The offset Δ_{v_i} and the local coordinate system are represented in Figure 2.

In a preliminary step of the algorithm, we compute the target shape $\Omega^T = \{\omega_{v_i}^1, \omega_{v_i}^2, \omega_{v_i}^3\}_{v_i \in V}$ at each vertex of the target character in the standard pose, where ω_i^j are the coordinates of Δ_{v_i} expressed in the local coordinate system of the target.

We then deform the source character in the deformed pose to respect these local offsets, thereby changing its body shape. That is, for each vertex v_i we want $\Delta_{v_i} = \Delta_{v_i}^T$. The desired position for each vertex is then:

$$d(v_i) = \sum_{v_j \in N_1(v_i)} \frac{1}{deg(v_i)} v_j - \\ (\omega_{v_i}^1 f_1(v_i) + \omega_{v_i}^2 f_2(v_i) + \omega_{v_i}^3 f_3(v_i)). \quad (2)$$

This allows to define the shape energy as in [Wuhrer et al. 2012]:

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

4.2 Volume

Encoding the shape as described in the previous section is purely based on the assumption of isometric deformations. However, isometries do not encode the volume of the shape [Lipman et al. 2007], and thus two isometric shapes can have drastically different volumes [Cosmo et al. 2019]. To remedy this, we add a volume term to

allow for an accurate transfer of both the volume and local shape of the target character.

Instead of preserving the total body volume we propose to preserve the volume of each body part independently. The reason is that the volume of the human body is not equally distributed among its body parts. Simply transferring the global volume of the target will lead to an under-constrained problem and result in unnatural distribution of the volume, e.g. inflated arms and face. Transferring the volume at a local level will constrain our method to rigid deformations, however shape deformation of the human body is inherently non-rigid. In our method we hence encode the volume at the body part level. Assuming that the volume of each body part is constant is an approximation because the volume of a body part can change during deformation, e.g. because of muscle deformation or breathing. We assume that this variation is negligible, and in our method we try to make the volume of a body part as close as possible to the volume of the same body part in the target character. This way we obtain a coherent volume distribution while allowing non-rigid deformation of the mesh's surface.

4.2.1 Body Parts Segmentation. We segment our template in a number of body parts (see Figure 3a). The segmentation is made by hand, and aims to segment the body parts that should deform rigidly.
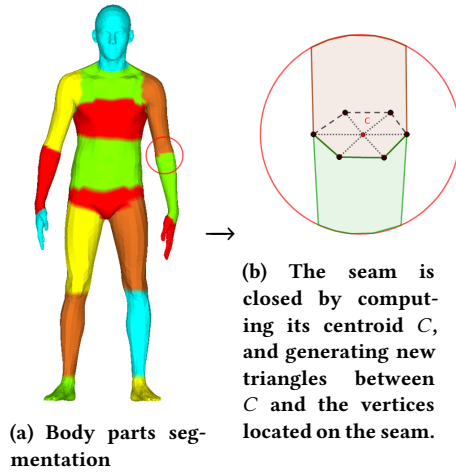

(a) Body parts segmentation

(b) The seam is closed by computing its centroid C, and generating new triangles between C and the vertices located on the seam.

Figure 3: Segmentation of body parts and close up of a seam.

We can then compute the volume of each body part as the sum of the signed volumes of the tetrahedrons formed by the body part's triangles and the origin O [Zhang and Chen 2001]. With $\{v_i, v_j, v_k\}$ a triangle and O the origin, the signed volume of tetrahedron $\{O, v_i, v_j, v_k\}$ is:

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

where (x_i, y_i, z_i) are the coordinates of point v_i.

The volume of a body part P is then $V_P = \sum_{T \in P} V_{O,T}$. This computation of volume needs a closed object. We close the segmented body parts by computing the centroid of the seam between two

body parts, and by generating triangles between this centroid and the vertices on the seam (see Figure 3b).

4.2.2 Volume Energy. Given the body parts segmentation of the template, we compute the volumes of the target character's body parts $V^T = \{V_P^T\}_{P \in BP^T}$ with BP^T the body parts of the target. The volume energy term measures the difference to the volumes of the target body parts V_P^T:

$$E_{Vol}(\mathbf{V}) = \sum_{P \in BP} (V_P - V_P^T)^2. \tag{4}$$

4.3 Contact

The shape and volume terms accurately morph the source character in the deformed pose to the morphology and local shape of the target. However, the morphology of the target can make it impossible to perform the exact same pose as the source. For example, if the source is a thin character with arms close to the body, a big target might be unable to reproduce the pose because his arms would enter his body. Conversely, if a big source touches his belly with his hand, a thin target might have his hand floating in front of the belly, therefore losing this contextually significant contact. Thus, changing the morphology of a character while keeping the same pose might result in impossible and/or semantically different poses.

Previous works [Ho et al. 2010; Jin et al. 2018; Liu et al. 2018] have solved this problem by deforming the pose to preserve the distances between interacting surfaces or joints. We argue that distances between surfaces are dependent of the character's morphology. Thus, two characters of different shapes performing an equivalent pose will have different interaction distances between surfaces. In our method, we avoid constraining distances between surfaces, while simply preserving existing contacts in the source pose and avoiding inter-penetrations. We assume that this way the target pose will be equivalent to the source's, with distances between surfaces coherent with the target's morphology.

For this purpose, our method includes a contact energy term. This term aims to maintain a coherent posture of the subject w.r.t. the source pose. In particular, we aim to maintain all contacts present in the source pose, while not introducing inter-penetrations. Our contact term is based on the contact loss presented in [Hasson et al. 2019]. This energy is composed of a repulsion term, which increases when surface inter-penetrations occur, and an attraction term, which increases when a contextual contact is lost:

$$E_C(\mathbf{V}) = \gamma_r E_r(\mathbf{V}) + \gamma_a E_a(\mathbf{V}), \tag{5}$$

where E_r and E_a are the repulsion and attraction terms, with associated weights γ_r and γ_a. The contact energy term corrects the pose of the result w.r.t. self interactions of the surface of the mesh (and the 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.

Jean Basset, Stefanie Wuhrer, Edmond Boyer, and Franck Multon
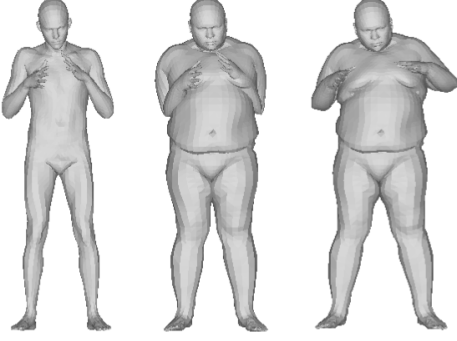
(a) Left to right: source pose, transfer result without repulsion term, and with repulsion term (target from Figure 5c). Notice the arms do not penetrate the torso anymore, and the wider gap between legs to avoid thigh colliding.

(b) Left to right: source pose, transfer result without attraction term, and with attraction term (target from Figure 6b). The right foot does not penetrate the leg thanks to the repulsion term, but needs the attraction term to keep the contact present in the source pose.

Figure 4: Effect of contact energy terms.

The repulsion term is defined as

$$E_r(\mathbf{V}) = \gamma_r \sum_{P \in BP} \sum_{v \in Int(P), v \notin P} d(v, P)^2 + \gamma_{rg} \sum_{v \in Int(G)} d(v, G)^2, \quad (6)$$

where BP is the set of the segmented body parts, 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; $d(v, X) = \inf_{w \in X} \|v - w\|_2$. In our experiments we set $\gamma_r = 1$ and $\gamma_{rg} = 0.1$. The effect of the repulsion term is demonstrated in Figure 4a.

We use raycasting to test whether a vertex is in the interior of a body part. If a ray from a vertex crosses the body part surface an odd number of times, the vertex is inside the body part [Jiménez et al. 2001]. Raycasting is a computationally costly method. To limit the number of tests required, we first cull the possibly colliding couples of vertex/body part using bounding boxes. For each body part, we create an axis-aligned and an oriented bounding box, and select the box with the tighter fit of the body part. We then test for collision of bounding boxes [Eberly 1999], and for each couple of colliding bounding boxes we select the vertices inside both for raycasting tests.

4.3.2 Attraction Term. Our contact energy term should also preserve contacts between surfaces present in the source's deformed pose. These contacts give 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:

$$\begin{aligned} E_a(\mathbf{V}) = &\gamma_a \sum_{(v_i, v_j) \in C} \max\big((d(v_i, v_j) - t), 0\big)^2 \\ &+ \gamma_{ag} \sum_{v \in C_G} \max\big((d(v, G) - t), 0\big)^2, \end{aligned} \quad (7)$$

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. In our experiments we set $\gamma_a = 1$ and $\gamma_{ag} = 0.1$ The effect of the attraction term is demonstrated in Figure 4b.

4.3.3 Rigid formulation. The contact energy (Equation 5) aims to preserve a coherent pose of the subject. As such, it should be minimized w.r.t. the pose deformation of the mesh.

To do so, we use the body part segmentation described in Section 4.2.1. We order the body parts 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:

$$E_C(\Theta) = \gamma_r E_r(\Theta) + \gamma_a E_a(\Theta), \quad (8)$$

using the mesh's vertex positions $\mathbf{V}$ after applying rotations Θ.

5 ITERATIVE SOLVING

Energy function 1 is complex to solve: this is mostly due to the shape term 3 being highly non-linear. Indeed, the differential coordinates that encode the shape are expressed in a local coordinate system, which depends on the position of the vertices of the mesh; moving a vertex also transforms its local frame. We therefore minimize Energy 1 iteratively.

The first step of our iterative framework is to move the vertices of the mesh to match the target's morphology. This step aims to minimize Equations 3 and 4 of our energy. We define the moving direction as the weighted sum of a shape and volume directions as:

$$v_i' = v_i + \epsilon(\gamma_s d_s(v_i) + \gamma_v d_v(v_i))$$

with v_i' the new position of vertex v_i, γ_S and γ_V the weights associated to the directions d_s and d_v, and ϵ the displacement offset.

The shape direction is simply the desired position w.r.t. shape defined in Equation 2.

We choose the volume direction for a vertex as the direction of the normal of the surface at the vertex. The moving distance is the difference between the target and current volume of the vertex's body part. This gives us the volume direction $d_v(v_i) = n(V_P^T - V_P)$. Extending or shrinking a shape in the direction of its surface's normal is an intuitive way to deform its volume. This is justified by Stokes' theorem; if we consider our moving direction as a vector field over the surface, we are looking for the vector field u that leads to the correct change in volume. By Stokes' theorem and its resulting divergence theorem, instead of integrating u over the interior of our shape, we can integrate on the surface using only the normal components of u along the surface.

The second step of our iterative framework aims to minimize the contact energy 8. This step corrects inter-penetrations and loss of contextual contacts that can be induced by moving the vertices. We use auto-differentiation to obtain the gradient of the contact energy 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.

After this second step, we compute the new local frames at each vertex of the mesh, and go back to step one. We iterate between these two steps until convergence. To determine whether convergence has been reached, we consider the evolution of the sum of energy terms (equation 1) compared to the previous iteration.

We experimentally find that $\epsilon = 0.3$, $\gamma_{Shape} = \gamma_{Vol} = \gamma_r = \gamma_a = 1$ give good quality results.

6 ADAPTATION TO DYNAMIC DATA

Our method computes shape from standard poses of the source and target characters, and transfers the shape to the deformed pose of the source. As such, it is designed for static data and we have no guarantee to obtain smooth animations when applied to dynamic data. However, our method is very robust; for a minor deformation between two frames of an animation, we generate close results, with the same general orientation of each body part. For movements with few close surface interactions, when retargeting between characters with relatively similar morphology, the resulting animation is smooth and our method is applicable as is.

For movements containing close surface interactions and contacts, or when retargeting between characters with significantly different morphology, jitter can appear in the resulting animation. This is due to the pose constraints; the deformed pose of the target is the pose of the source adapted to the target character's morphology. Without temporal data in the method, this adaptation can be slightly different between two consecutive frames. This difference is however relatively small, and we provide a simple adaptation of the method to dynamic data. In particular, we remove jitter with a simple low-pass filter of the sequence in a post processing step.

Another problem caused by the correction of pose is the ground contact. For example, when transferring a motion from a skinny character to a big one, the gap between the legs is widened to avoid collisions between the thighs. This results in slightly different positions of the feet in consecutive frames depending on the correction of the pose. This is an important contextual artifact in an animation.

To remove this artifact, we modify the ground contact part of the attraction term (Equation 7) of our method. For static data, the ground contact is forced by constraining the concerned vertices' height to be at the ground level, but not at a fixed point. This allows for correction of the pose while keeping the ground contact. For dynamic data, we log the ground contacts of the source at each frame. If for a new frame, we detect ground contacts that were already present at the previous frame, we consider that this ground contact should be preserved at the same position. Therefore for the corresponding vertices v, the ground contact term in the attraction energy now minimizes $max((d(v^t, v^{t-1}) - t), 0)^2$. Once this artifact is removed, the resulting animation has coherent ground contacts.

7 RESULTS

7.1 Implementation Details

For our experiments, to have point to point correspondence, all our meshes follow the SMPL [Loper et al. 2015] template (6890 vertices and 13776 faces), segmented in 17 body parts (see Figure 3a). We use meshes from SMPL, Faust [Bogo et al. 2014], Dyna [Pons-Moll et al. 2015], and from Liu et al. [Liu et al. 2018]. All experiments were run on a PC with an Intel Xeon CPU and 16GB of RAM. Computation time is highly dependent on the surface interactions present in the transfer (contacts and possibly colliding surfaces). For surface interaction free poses (see Dyna running sequence in the video), the method takes around 7 minutes to perform the complete transfer. For the transfer of figure 5 where no contextual contacts are present but corrections are needed for surface collisions, the methods takes around 25 minutes. For the opposite situation, i.e. preservation of contacts but no important surface collision, with the transfer of figure 9b, the method takes around 30 minutes. Finally, when the both contextual contacts and surface collisions are present in the transfer like in figure 9c, the method takes around 45 minutes.

7.2 Qualitative Results

In this section we show general results of our method on different characters and poses.

Figure 5 illustrates the iterative process of our method. We show the state of the result at several points in the iterations (5d). We see that the shape and volume evolve quickly to match the target, while the contact term avoids interpenetrations, here by widening the gap between the legs and raising the arms. Figure 5b shows the evolution of each energy term during this transfer. Note that the shape term (Eq. 3) and volume term (Eq. 4) decrease rapidly in the first iterations. The initial spike of the repulsion term (Eq. 6) is due to interpenetrations appearing following the change in morphology. The correction of interpenetrations causes loss of contacts around the armpits, explaining the slight increase in the attraction term (Eq. 7). Our iterative process efficiently minimizes the shape and volume energies, while maintaining the contact energy at a reasonable level.

Figure 6 shows results of shape transfer from the source of figure 7 to characters with varying morphology. Notice the evolution of the space between arms and body depending on the target's shape.

7.3 Comparisons

In this section, we compare our results with related works. We show that we are able to generate similar results for similar inputs, and illustrate the contribution of our method.
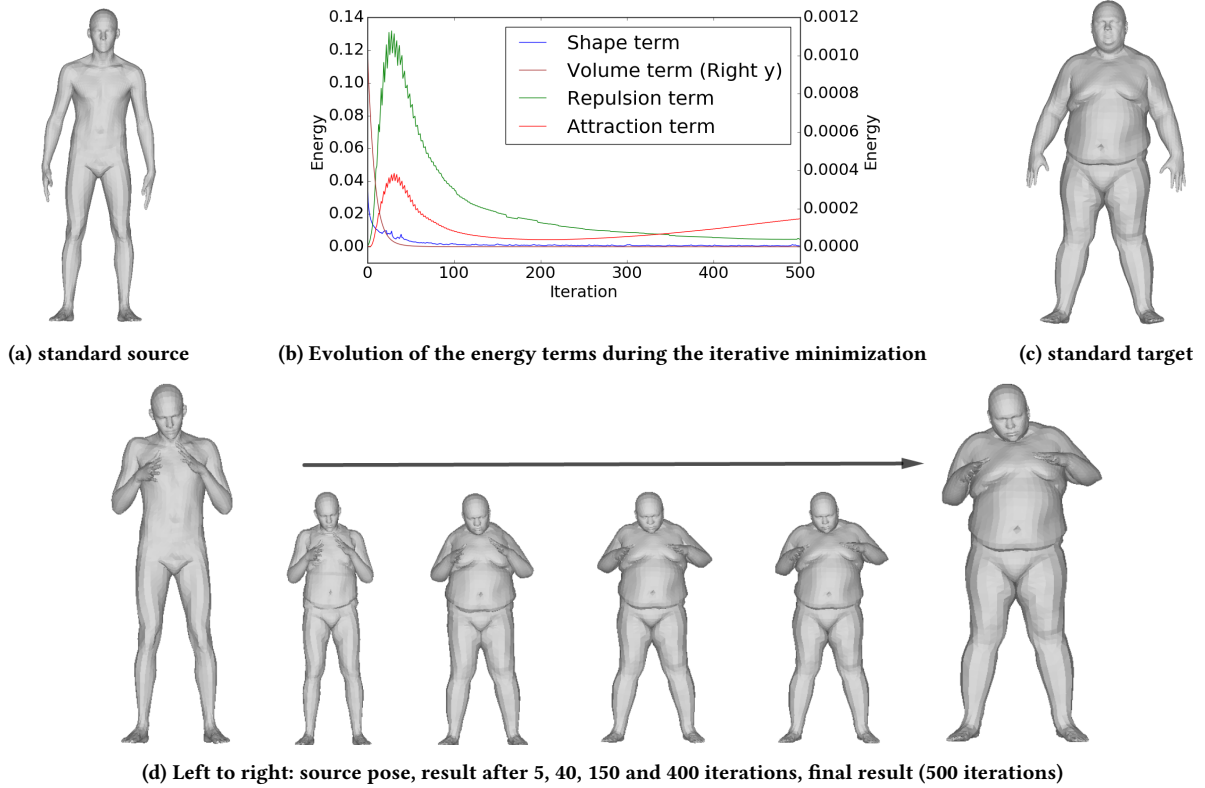
(a) standard source

(b) Evolution of the energy terms during the iterative minimization

(c) standard target

(d) Left to right: source pose, result after 5, 40, 150 and 400 iterations, final result (500 iterations)

Figure 5: Evolution of the shape transfer from a thin to a big character through the iterations.

Figure 7 illustrates the motion retargeting problematic. The source pose (left) is generated by hand-tuning SMPL pose and shape parameters. We then apply the same pose parameters to a character with different shape parameters (center). We see that this straightforward approach does not give an acceptable result; the left hand enters the belly, while the contact between the right hand and the hip is incorrect. We then show the result of our method with these models (right). Our method successfully avoids inter-penetrations and preserves the hand/hip contact. Moreover, notice that the space between the arms and the body shrinks. This distance is linked to the morphology of the character and not the pose.

Figure 8 shows a result of our method on the models from Liu et al. [Liu et al. 2018, Figure 6]. We show that even with an extreme change in morphology, our result is relatively close to the solution proposed by an artist. In particular, when viewed from above, we see that our result and the artist's both create new contacts between the arms and the body. This confirms our hypothesis that some distances (here between the arms and the body) are a result of the target's shape, and are not fixed by the pose. Moreover, it shows that our method is able to handle some of these cases.

Figure 9 shows results of our method on a shoulder rubbing pose. We see that our method preserves the hand/shoulder contact, even with radical changes of morphology. Notice that for a close morphology (Figure 9b), the distance between the elbow and the torso does not change much in the result, but that for significantly bigger target character (Figures 9c and 9d) this distance shrinks or

even disappears to form new contacts. This result further illustrates the correlation between the pose and the shape of a character. Jin et al. [Jin et al. 2018, Figure 8] explore a similar example. However, their method preserves exactly the distances between surfaces. We assume that applying their method to important morphology differences, like Figures 9c and 9d, would not create these new morphology dependent contacts, and would result in an unnatural position of the arm to preserve the distances present in the source.

We argue that our method proposes an alternative to such distance preserving methods [Ho et al. 2010; Jin et al. 2018; Liu et al. 2018], and is a first step in taking into account the correlation between shape and pose in motion retargeting.

7.4 Limitations

Our work suffers from several technical drawbacks. First, our method is slow compared to recent methods that are close to real-time. Implementation improvements would be possible to accelerate the method. For example, since most of our computations are per-vertex, re-implementing a highly parallelized version of the method on the GPU would represent an important gain of time.

Another drawback is that our method requires complete correspondence of input meshes (number of vertices and connectivity). In the context of animation retargeting, we often want to transfer poses between meshes coming from different databases, so this condition is rarely met. In our experiments we worked around
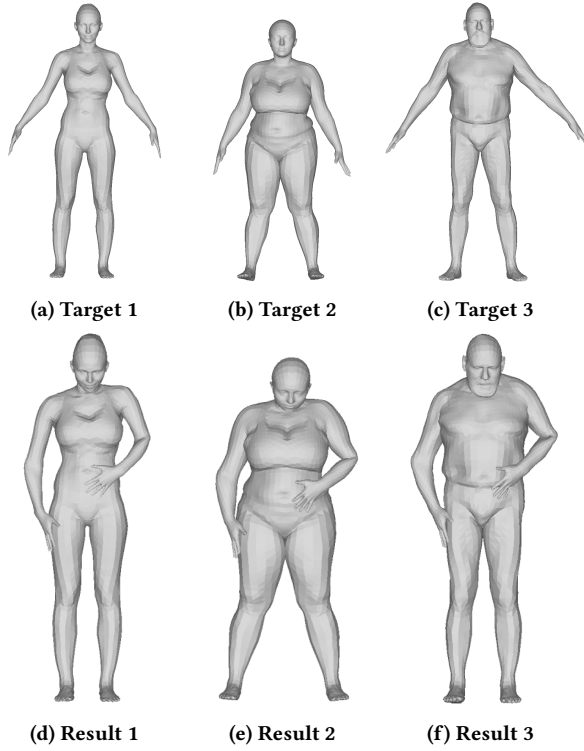
(a) Target 1   (b) Target 2   (c) Target 3

(d) Result 1   (e) Result 2   (f) Result 3

Figure 6: Shape transfer results on several characters from a SMPL generated pose with contacts (Figure 7, left)
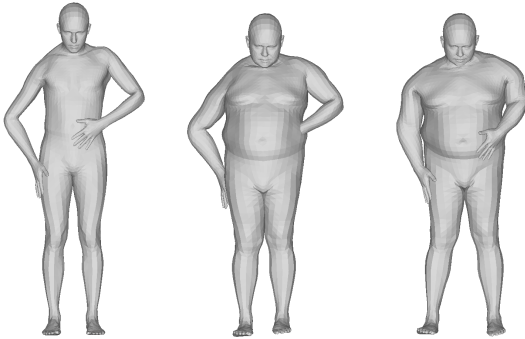

Figure 7: Left: Source pose generated by manually tuning SMPL shape and pose parameters. Center: same SMPL pose parameters applied to new shape parameters. Right: result of our method from the left pose to the center shape.

this problem by fitting our inputs to the SMPL template. This pre-processing is relatively light and only needs to be applied once per mesh, and is thus preferable to the rigging of a skeleton.

While our method is skeleton free, we still require a body part segmentation. Several works have explored automatic body part segmentation [Benhabiles 2011; Wang et al. 2016], their solutions could be used in our system. However, if we consider fitting the
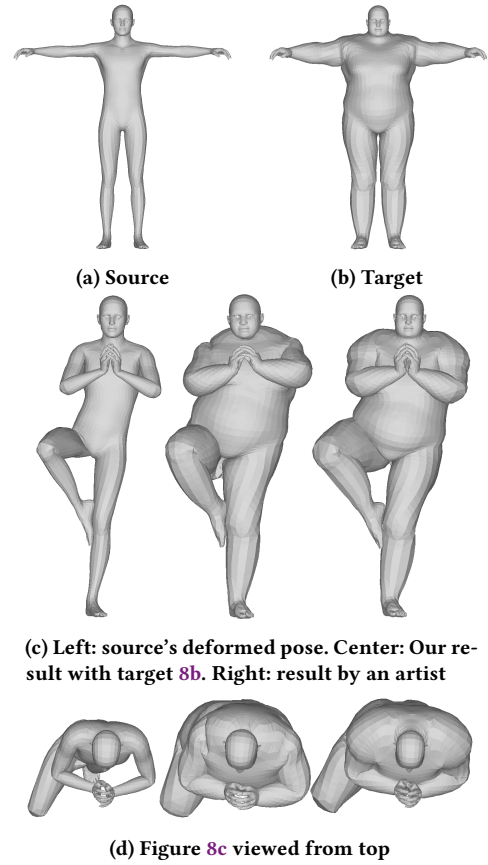

(a) Source   (b) Target

(c) Left: source's deformed pose. Center: Our result with target 8b. Right: result by an artist

(d) Figure 8c viewed from top

Figure 8: Results of our method on models from [Liu et al. 2018] (used with authors' agreement)

(a) Source pose   (b) Target in figure 6a

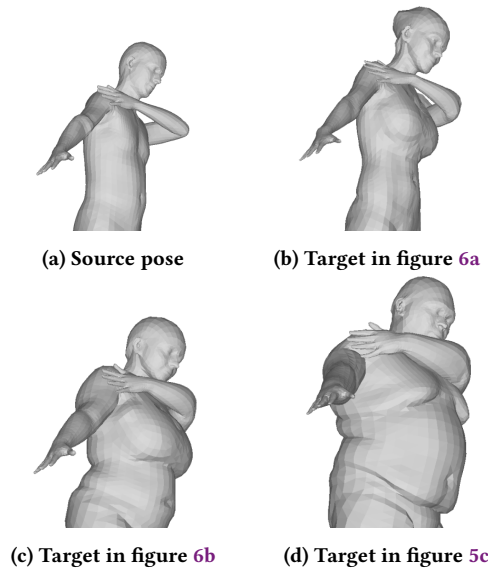(c) Target in figure 6b   (d) Target in figure 5c

Figure 9: Shape transfer results on several characters from a shoulder rubbing pose

inputs to the SMPL template part of the solution, this drawback disappears since the segmentation only needs to be computed once.

When retargeting from a thin to a fat character, our method can result in new morphology dependent contacts (Figure 9d), which we argue is one of the main contributions of our method. However, the other way around, when retargeting from a fat to a thin character, our method could preserve contacts from the source pose that are due to morphology and not pose. Since our method does not work at interactive times, a solution would be to have input from the animator, who could specify which contacts to preserve. This semi-automatic solution would be lighter than the usual rigging of a skeleton while giving artistic freedom to the animator.

8 CONCLUSION AND FUTURE WORK

In this paper we investigated a novel approach to motion retargeting without requiring meticulous rigging of the skeleton. The idea is to avoid trial-and-error editing by directly controlling the surface mesh instead of the internal skeleton. Previous deformation transfer approaches consider that the pose is modelled thanks to topological relationships (mainly distances or relative positions) between body parts, whereas it could be more complex, including orientations and coordination between body parts.

The main contribution of the paper is to explore transferring the shape instead of the pose, assuming that it would better preserve the contextual meaning of the source shape+pose. Whereas other methods [Jin et al. 2018; Liu et al. 2018] preserve the distances between surfaces, we allow surface interactions (e.g. distances) to change during the transfer because the shape is changing. Our results show that the resulting target shape+pose is more plausible than when preserving the distances between the body parts.

Several improvements could significantly decrease computation time, such as better controlling the number of iterations, optimizing the code, and possibly using parallel computing for some parts. Moreover, we could also use the principle of displacement maps [Gleicher 1998] to solve the shape transfer when contacts occur only, and interpolate or filter solutions in-between.

A promising future direction is to use a database of characters with varying morphologies performing equivalent poses, in order to learn the relevant surface interactions for shape transfer. This would allow to disambiguate between pose and shape-related contacts based on examples.

ACKNOWLEDGMENTS

REFERENCES

Rami Ali Al-Asqhar, Taku Komura, and Myung Geol Choi. 2013. Relationship Descriptors for Interactive Motion Adaptation. In *ACM SIGGRAPH/SCA (SCA '13)*. 45–53.

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.

Quentin Avril, Donya Ghafourzadeh, Srinivasan Ramachandran, Sahel Fallahdoust, Sarah Ribet, Olivier Dionne, Martin de Lasa, and Eric Paquette. 2016. Animation Setup Transfer for 3D Characters. *Computer Graphics Forum* 35, 2 (2016), 115–126.

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.

Ilya Baran, Daniel Vlasic, Eitan Grinspun, and Jovan Popović. 2009. Semantic deformation transfer. In *ACM TOG*, Vol. 28. 36.

Halim Benhabiles. 2011. *3D-mesh segmentation: automatic evaluation and a new learning-based method*. Ph.D. Dissertation.

Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. 2014. FAUST: Dataset and evaluation for 3D mesh registration. In *CVPR*. IEEE.

Adnane Boukhayma, Jean-Sébastien Franco, and Edmond Boyer. 2017. Surface motion capture transfer with gaussian process regression. In *CVPR*. 9.

Kwang-Jin Choi and Hyeong-Seok Ko. 2000. Online motion retargetting. *The Journal Of Visualisation and Computer Animation 2000* 11, 5 (2000), 223–235.

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).

David Eberly. 1999. Dynamic collision detection using oriented bounding boxes.

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.

Michael Gleicher. 1998. Retargetting Motion to New Characters. In *Siggraph (SIGGRAPH '98)*. ACM, 33–42.

Yana Hasson, Gül Varol, Dimitrios Tzionas, Igor Kalevatykh, Michael J Black, Ivan Laptev, and Cordelia Schmid. 2019. Learning joint reconstruction of hands and manipulated objects. *CVPR* (2019).

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.

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.

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.

Chun-Hao Huang, Edmond Boyer, and Slobodan Ilic. 2013. Robust human body shape and pose tracking. In *3DV*. IEEE, 287–294.

Pablo Jiménez, Federico Thomas, and Carme Torras. 2001. 3D collision detection: a survey. *Computers & Graphics* 25, 2 (2001), 269–285.

Taeil 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.

Richard Kulpa and Franck Multon. 2005. Fast inverse kinematics and kinetics solver for human-like figures. In *IEEE-RAS Humanoids*. 38–43.

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.

Benoît Le Callennec and Ronan Boulic. 2006. Robust Kinematic Constraint Detection for Motion Data. In *ACM SIGGRAPH/SCA (SCA '06)*. 281–290.

Thibaut Le Naour, Nicolas Courty, and Sylvie Gibet. 2019. Skeletal mesh animation driven by few positional constraints. *Computer Animation and Virtual Worlds* (2019), e1900.

Jehee Lee and Sung Yong Shin. 1999. A hierarchical approach to interactive motion editing for human-like figures, In Siggraph. *39–48*.

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.

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.

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.

Eray Molla, Henrique Galvan Debarba, and Ronan Boulic. 2018. Egocentric Mapping of Body Surface Constraints. *IEEE TVCG* 24, 7 (2018), 2089–2102.

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.

Gerard Pons-Moll, Javier Romero, Naureen Mahmood, and Michael J. Black. 2015. Dyna: A Model of Dynamic Human Shape in Motion. *ACM TOG, (Proc. SIGGRAPH)* 34, 4 (2015), 120:1–120:14.

Helge Rhodin, James Tompkin, Kwang In Kim, Edilson De Aguiar, Hanspeter Pfister, Hans-Peter Seidel, and Christian Theobalt. 2015. Generalizing wave gestures from sparse examples for real-time character control. *ACM TOG* 34, 6 (2015), 181.

Robert W Sumner and Jovan Popović. 2004. Deformation transfer for triangle meshes. In *ACM TOG*, Vol. 23. 399–405.

Hanqing Wang, Changyang Li, Zikai Gao, and Wei Liang. 2016. Joint labelling and segmentation for 3D scanned human body. In *SIGGRAPH ASIA*. ACM, 13.

Stefanie Wuhrer, Chang Shu, and Pengcheng Xi. 2012. Posture-invariant statistical shape analysis using Laplace operator. *Computers & Graphics* 36, 5 (2012), 410–416.

Cha Zhang and Tsuhan Chen. 2001. Efficient feature extraction for 2D/3D objects in mesh representation. In *ICIP*, Vol. 3. IEEE, 935–938.

Kun Zhou, Weiwei Xu, Yiying Tong, and Mathieu Desbrun. 2010. Deformation transfer to multi-component objects. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 319–325.