Computers & Graphics 38 (2014) 191–198

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on CAD/Graphics 2013

SCAPE-based human performance reconstruction

Jiaxiang Zheng [a], Ming Zeng [b,*], Xuan Cheng [a], Xinguo Liu [a,*]

[a] State Key Lab of CAD&CG Zhejiang University, Hangzhou 310058, China
[b] Software School of Xiamen University, Xiamen, 361005 China

ARTICLE INFO

Article history:
Received 5 August 2013
Received in revised form
6 October 2013
Accepted 19 October 2013
Available online 4 November 2013

Keywords:
Animation reconstruction
Human body shape modeling
Articulated ICP

ABSTRACT

This paper presents an automatic approach to reconstruct human motion using noisy depth data from multiple views. Although multi-view constraints are provided by this setup, it still exhibits great challenges to robustly reconstruct dynamic human performances due to inherent complexity and self-occlusion of human motion. In the insight that the semantics of human motion will supply strong prior in motion reconstruction, we therefore propose a SCAPE-based motion reconstruction algorithm. As the building blocks of this main algorithm, we (1) re-train a SCAPE model based on an expanded human pose database containing human poses collected from different databases to enlarge the tracking space, (2) develop a correspondence estimation method based on articulated ICP to improve the robustness of SCAPE tracking. We conduct experiments to demonstrate the effectiveness of our method, and show that our system is able to capture and reconstruct accurate human motion.

1. Introduction

Animation reconstruction, especially human performance reconstruction, plays an important role in a large number of applications, ranging from robotics to interactive games and movie productions. However, the acquisition of accurate and reliable 3D digitization of the dynamic shapes is not an easy task due to the complexity of human motion, as well as the occlusions of the capturing process. When the acquisition is extended to the entire body performances, the task is even more difficult. Recent advances in depth camera such as Microsoft Kinect have opened new possibilities for the acquisition of dynamic objects. However, these devices pose substantial challenges with noisy data and above problems. Dealing with these problems, we proposed a full human reconstruction system using four Kinects with the help of a statistical model (SCAPE [1]).

SCAPE captures the shape and pose variation of human body, which make the estimation of human motion in a low-dimensional space. Compared with linear blending skinning (LBS) based skeletal model, SCAPE can provide much more semantic constraints on both shape and pose. Besides, SCAPE model is a general template of human objects, which means we do require an explicit template as input.

Along the above insight, we conduct two main works for our reconstruction system. First, to represent a more general shape and pose space, we collect human data from different databases, and then train a SCAPE model on these data. We introduce a transfer based method to setup the correspondences across different databases, which is the main challenge of the database expanding. Second, to incorporate the SCAPE prior, we first track human pose using SCAPE model and then incorporate a detail aggregation step to generate details onto the estimated SCAPE model. In the pose tracking step, we divide the SCAPE model into several parts and make a hierarchical rigid registration for accurate correspondences, then use the SCAPE model to estimate the final pose parameters for each frame. In the detail aggregation step, we use a nonrigid deformation model to reconstruct the details of body motion while preserving the topology and smoothness of the template mesh.

To summarize, this paper presents a framework to reconstruct the motion of human performance. The contributions can be concluded as follows:

- A framework to reconstruct performance from low quality Kinect data with SCAPE model.
- A transfer-based method to enlarge the SCAPE pose training database.
- A hierarchical solution to articulated ICP method to robustly track performer's motion.

2. Related work

Our work is most related to the following research areas: (1) SCAPE-based body estimation, (2) articulated ICP, and (3) animation reconstruction of human body.

* Corresponding author. Tel.: +86 15959236379, +86 571 8820 6681x526, fax: +86 571 8820 6680.
E-mail addresses: jiaxiang.zheng135@gmail.com (J. Zheng), mingzeng85@gmail.com (M. Zeng), xgliu@cad.zju.edu.cn (X. Liu).

0097-8493/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cag.2013.10.023

SCAPE-based body estimation: SCAPE is a morphable model proposed by Anguelov et al. [1], it encodes the variations of both body shape and pose. With this model, a new body can be generated using a set of shape and pose parameters.

Tong et al. [2] use SCAPE to register each frame data and reconstruct a detailed human model. Hasler et al. [3] proposed a method to estimate an accurate body model from dressed 3D scan data. Different from those methods estimating a human model from 3D data, Balan et al. [4,5] use a stochastic search method to estimate 3D body model from 2D images of multiview. Later on, researchers applied SCAPE model with semantic information to estimate the human body shape and pose from a single image for virtual try on [6] and image body reshaping [7] usage. In our system, SCAPE is used to estimate both pose and shape from 3D data for each frame.

Articulated ICP: Original articulated ICP method is used to track a sequence of human motion without markers in motion capture system, its objective is to track an articulated model from a sequence of visual hulls automatically. Mundermann et al. [8] firstly proposed a mathematical formulation of this problem with a global solution, which is not robust because it assumes that each joint contains 6 freedom. To solve this, Corazza et al. [9] use a Levenberg–Marquardt minimization scheme with the consideration of joint freedom. Ganapathi et al. [10] extend the work to real time applications based on MAP inference in a probabilistic temporal model. In our system, we use the articulated ICP to estimate correspondences for pose tracking. We use a similar formulation to Mundermann et al. [8] but with a hierarchical solution.

Animation/performance reconstruction: A lot of research works have been concentrated to reconstructions including static scene [11,12] and motions such as dynamic facial expression [13,14] and human body performance, we only touch the reconstructions on human body performance in this paper. Generally, performance reconstruction can be classified into template based method and template-less method.

Reconstruction with a template: Due to capturing method or single view capture, a lot of occlusions and holes might occur when trying to capture the model of object. Template based method can be viewed as a solution to complete these hole areas and as a prior to reconstruct a fully reconstructed model. Li et al. [15] proposed to drive a deformation graph for large scale deformation to the template model and a later detail aggregations to the deformed template, which result in an accurate fitting to the single view scan data. Our reconstructing procedure is similar to Li et al. [15]. However, our reconstruction is only for human performance and we can choose a more robust pose tracking and transferring method for the large scale pose deformation; and due to the noisy scanning data, we cannot estimate the detail coefficients like Li et al.'s work. As is obvious to see that the single view reconstruction is an under-constrained problem, most of the template based works have focused on multiview data captured from Light Stage System, etc. [16–18] and use a template with fine details to reconstruct motions. Especially, Vlasic et al. [16], Ye et al. [19], Gall et al. [18] and Stoll et al. [20] employ an embedded skeleton with LBS method to drive the template for animation and a deformation for reconstruction refinement. While Aguiar et al. [17] solved the problem using a set of Laplacian deformation with the assistance of silhouettes and SIFT features from multiview. Unlike these methods using an explicit template, our method employs an implicit template generated by SCAPE model.

Reconstruction without a template: Template based method can reconstruct the incomplete data with a template prior, but it is limited to reconstructing the object similar to the template. Some works have been done on the reconstruction work without a template. Vlasic et al. [21] proposed to reconstruct highly detailed performances with Light Stage by using optical flow to correct the motions and employing the Volumetric Range Image Processing (VRIP) method [22] for hole filling and final reconstruction. With a similar hardware setup, Li et al. [23] employ silhouettes of each frame as the prior for shape completion and use a graph based deformation method to piecewise matching the adjacent frame data, and a final complete mesh is generated using Poisson reconstruction [24]. Different from those multiview capturing methods, single view reconstruction without template is also proposed in recent years [25,26]. Zeng et al. [26] take the assumption that the object is quasi-rigid with slight motions while Chang et al. [25] assume that the object is articulated, both of them use a reduced deformation model for an initial coarse registration.

3. Hardware setup and system overview

3.1. Acquisition setup

In order to capture the full body motion while preserving a reasonable accuracy, we use four Kinect cameras with two to capture the front and two to capture the back of the performer. Cameras toward the same orientation are placed to capture the upper and lower body with little overlapping region. The distance of both side cameras to the performer is about 1.5 m. Fig. 1 shows the hardware setup with four Kinect cameras marked in red rectangles and performer marked in green ellipse. The cameras are calibrated with an LED waving technique [27].

3.2. System overview

Fig. 2 illustrates the pipeline of our framework. Our system relies on a parameterized model of body shape and pose. To enhance the representation ability on human pose space, we train a SCAPE model [1] based on an expanded pose training data combined with multiple pose database, which is described in Section 4.2.

We decompose the reconstruction work into two steps: a coarse pose reconstruction step and a detail aggregation step. To initialize the template mesh, a set of corresponding pairs between the first frame scan data and the standard SCAPE model is manually selected, see Section 5.1. After the initialization step, we use a nonlinear SCAPE pose estimating method combined with an articulated ICP tracking step, see Section 5.2. Then the pose of estimated SCAPE mesh is transferred to the template mesh, as well

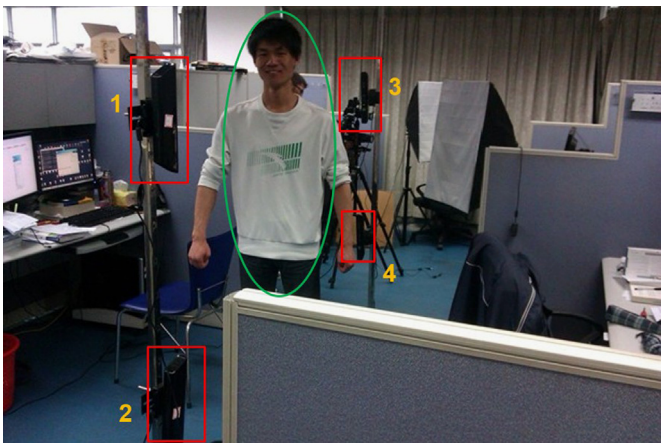

Fig. 1. Hardware setup. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

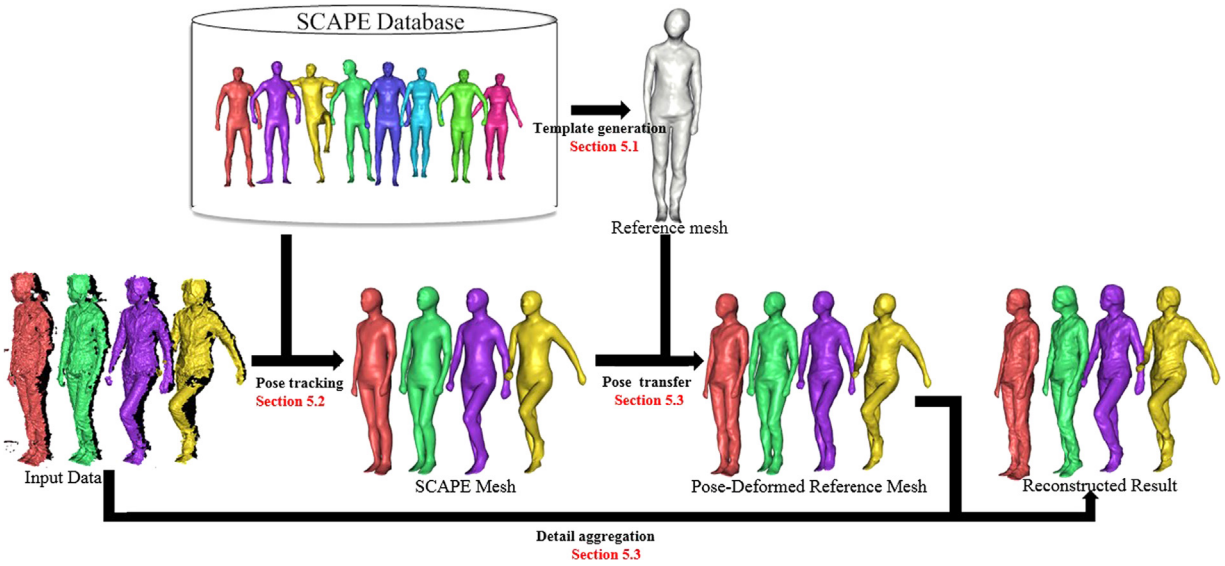

Fig. 2. System pipeline overview.

as a detail aggregation to the pose deformed template mesh, see Section 5.3. To keep the spatial continuity of the reconstructed result, we use the bilateral filter to smoothen the result.

4. SCAPE model review and pose expansion

Since the SCAPE model is the basis of our system, we will take a brief overview to the model and describe how to use a transfer method to expand the pose training data in the following subsections.

4.1. SCAPE model review

SCAPE is a deformable, parametrized model of the human body. It independently learns the shape deformation parameters and pose deformation parameters from a set of shape and pose training data. More specifically, the transformation for each triangle from standard mesh X of SCAPE model to an instance mesh Y is decomposed into three components: the joint rotation R, the muscle deformation Q induced by joint rotation and the body shape deformation S. A collection of these deformations for all triangles of the mesh is represented by $\mathcal{R}, \mathcal{S}, \mathcal{Q}$. After applying a linear regression model to the training pose data, we can learn pose-induced deformations $R(\theta)$ and $Q(\Delta\theta)$ where θ denotes the set of joint rotation angles of the instance mesh, $\Delta\theta$ denotes the relative joint rotation for adjacent joints that are connected by a common body part. For the shape training process, a PCA model is employed for existing shape training data, and the new shape deformations can be represented as $S(U\beta+\mu)$ where β is the shape coefficient, U is the eigenvectors and μ is the mean vector of the PCA result.

After learning all parameters, a new instance mesh Y can be generated with pose coefficients θ and shape coefficients β by minimizing the following energy:

$$E_Y = \| \mathcal{R}(\theta) \circ S(U\beta+\mu) \circ \mathcal{Q}(\Delta\theta) \circ \nabla X - \nabla Y \|^2 \quad (1)$$

where $\circ$ and ∇ operators denote general deformations on gradient space (the triangle's local coordinate system).

A general purpose of the SCAPE model is to estimate the parameters by giving some constraints such that the resulting instance mesh is as close as possible to the constraints. Denote the correspondences C between the input data we are going to estimate and the standard SCAPE mesh X, the corresponding SCAPE mesh Y can be estimated by a modified version of Eq. (1) with constraints. However, we do not know shape coefficients β and pose coefficients θ as a prior in most cases. In fact, we have to solve all these unknown variables at once by minimizing an energy defined below:

$$\arg\min_{Y,\beta,\theta} \sum_{(\mathbf{y},\mathbf{y}') \in C} \| \mathbf{y} - \mathbf{y}' \|^2 + E_Y \quad (2)$$

where E_Y is defined in Eq. (1), $\mathbf{y}$ is the vertex of Y and $\mathbf{y}'$ is the corresponding constraints.

Note Eq. (2) is a non-linear and non-convex problem which is easily converged to a local minima. We adopt the iterative method which is proposed by Anguelov et al. [1] that iteratively fixed two variables and solved the remaining variable.

4.2. Pose training data expansion

As described above, the pose-induced deformations $\mathcal{Q}$ are learned from a linear pose space that approximated to the complex human pose space. This means that if the learned linear pose space is far away from the actual pose space, it is impossible for us to get a reasonable result (see Fig. 3). To make our trained model more robust to pose variations, we choose to enlarge the pose training database. However, the input to pose training process is a parameterized pose database corresponding to the same performer, which means the database is not easy to enlarge. With the insight that different pose databases can be combined together to a much larger database, we devised a deformation method to transfer poses from different databases to a reference database. For simplicity, we discuss how to combine two parameterized pose databases to be a larger database.

Denote two parameterized pose databases as $\mathfrak{D}$ and $\mathfrak{D}'$, associated with the standard mesh X and X' respectively. Suppose X and X' share similar pose, then the correspondences between the two standard meshes can be performed by deforming X to X' to find the nearest neighbour correspondence for each face of X [28].

Given the correspondences C, we can generate the new pose Y for $\mathfrak{D}$ by transferring the poses Y' from $\mathfrak{D}'$ to X. This can be solved by a modified deformation transfer method [28] defined as

$$E_t = \sum \| T_i - T_i' \|^2 + \gamma \sum_i \sum_{j \in N(i)} \| T_i - T_j \|^2 \quad (3)$$

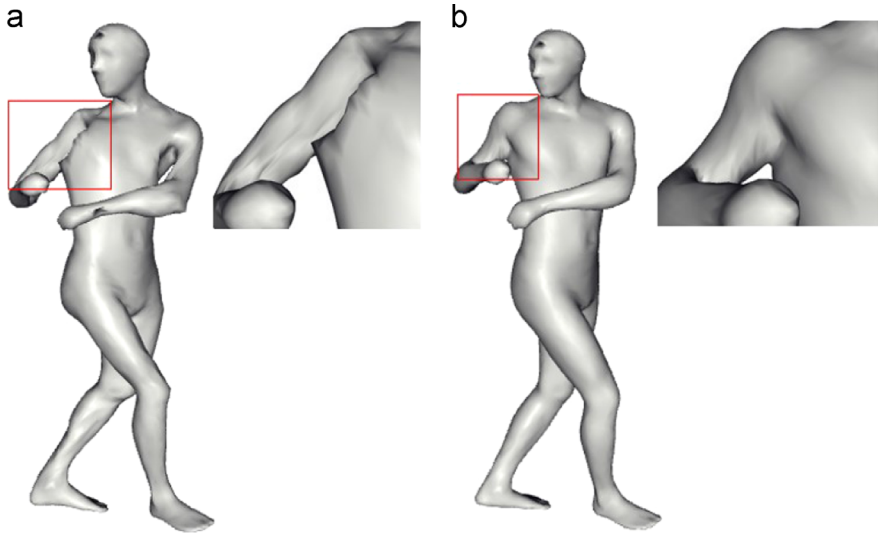

Fig. 3. Comparison of SCAPE mesh learned from original pose training data (a) and the expanding pose training data (b).

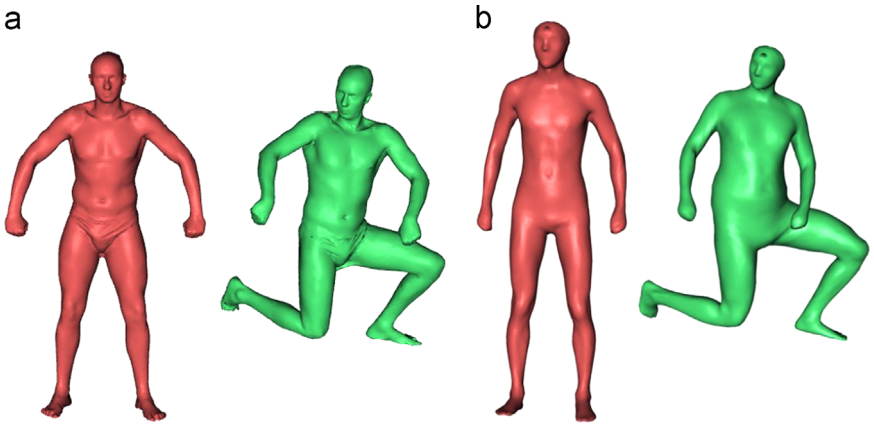

Fig. 4. Illustration of pose transfer result: (a) Reference mesh X' (left) and pose instance mesh Y' (right) from $\mathfrak{D}'$. (b) Reference mesh X (left) and generated instance mesh (right) for $\mathfrak{D}$.

where T denotes the affine transformation from X to Y of each triangle face while T' denotes transformation of the corresponding triangle from X' to Y', $N(i)$ denotes the neighbour faces of i-th triangle of X. γ balances the smoothness and pose similarity.

Note that C is setup with the assumption that X and X' share the similar pose. However, when X and X' exhibit a different pose, we cannot use the formula above directly to transfer pose. A pose-induced deformation should be performed to X so X and X' share the similar pose, and then we can use the strategy above to solve the pose expansion problem.

In our experiment, we use the database from Anguelov et al. [1] and Hasler et al. [29]. The size of pose database is expanded from 34 to 104. Fig. 4 shows one of these poses. Fig. 7 shows the comparison of SCAPE result of different pose databases for real data.

5. Pose tracking and motion reconstruction

In this section, we will describe the pipeline of our system which contains an initial template mesh estimation, a pose tracking and a motion reconstruction stage. For clarity, denote the sequence of character's performance as $\mathcal{D}=\{D^1,\ldots,D^n\}$, where D^t is the point cloud data at frame t. The generated SCAPE mesh corresponding to D^t at frame t is represented as Y^t.

5.1. Initial template estimation

Instead of using an explicit template mesh as many template based reconstruction methods, we use D^1 to generate a coarse template with the assistance of SCAPE model. Given the input D^1, we require a manually selected correspondences $\mathcal{C}$ between D^1 and X, which is always less than 20 in our experiments.

Before estimating the template mesh, we need to estimate a SCAPE mesh firstly, which can be solved with $\mathcal{C}$ by Eq. (2). Here, we set the initial pose coefficients θ and shape coefficients β all to be 0. After a resulting mesh being estimated using $\mathcal{C}$ for the first iteration, we need to refine the result with more dense correspondences. Nearest neighbour correspondences are sufficient as the input since the estimated result is already close to scan, we can iteratively solve Eq. (2) until a reasonable result is estimated.

Note that SCAPE result Y^1 is close to D^1, we can directly use it as our template mesh for later motion reconstruction. However, to reconstruct the details of input scan, the template mesh needs to have a medium scale resolution. In most cases, the subdivided result can be regarded as a template mesh used in the later stage. For the performer with tight clothes, the SCAPE mesh is close to the scan only in general, we have to deform the subdivided result to be as close to the scan as possible. In case of transferring potentially transient features to future scans, we remove all the high-frequency details of the deformed result. We denote this final

result as the template mesh M and use it to reconstruct the later pose and details in Section 5.3.

Since the body shape of a person would not change, after estimating the coefficients of the first frame, we keep β fixed and only estimate θ^t and Y^t at frame t.

5.2. Pose tracking

After the initialization step in Section 5.1, we get the SCAPE mesh Y^1 as well as the template mesh M. In the tracking stage, our goal is to estimate a SCAPE mesh Y^t frame by frame. Suppose we have already estimated a previous SCAPE mesh Y^{t-1} at frame t, we then use Y^{t-1} as input to estimate Y^t. If a set of accurate correspondences between D^t and Y^{t-1} is given, Y^t can be easily estimated similar to the initial SCAPE mesh estimation step. However, the fact is that we have no accurate correspondences for data at frame t.

To solve the correspondence estimation problem, some research works rely on the feature points or optical flow of texture information [30,17], and some rely on the nearest neighbour correspondences [15]. As depicted in Fig. 6, in our experiment, direct nearest neighbour correspondences might result in unnatural deformations because of incorrect correspondences of Eq. (2). Before using a neighbour correspondences directly, we use an articulated ICP [8] strategy to Y^{t-1} for registering current scan part by part to find accurate correspondences coarsely.

Since the SCAPE model already contains a skeletal meaning, we regard the SCAPE mesh as skeletal body parts and associate each part a joint except the root joint. Due to the lack of correspondences for all parts at the beginning of estimation, we choose a hierarchical rigid estimation method rather than a global estimation method [8]. Assume that the body root part's motion is relatively small in adjacent frame, the root part can be aligned well using standard ICP algorithm [31]. For other parts, after a parent part's transformation $\mathbf{T}_p=[\mathbf{R}_p\ \mathbf{t}_p]$ being estimated, we apply $\mathbf{T}_p$ to the part, and recompute the new joint position $\mathbf{c}$ for the part as rotational center. Here, we simply use the center of edge vertices belonging to current part. Since current part's motion is mainly rotation around $\mathbf{c}$, the translation of joint should be relatively small. We can estimate the rigid transformation of current part by minimizing the energy defined below with a similar strategy to classical ICP procedure:

$$\arg\min_{\mathbf{R},\mathbf{t}} \sum_{(\mathbf{v},\mathbf{v}')} \|\mathbf{R}(\mathbf{v}-\mathbf{c})+\mathbf{c}+\mathbf{t}-\mathbf{v}'\|+w\|\mathbf{t}\|^2$$

where $(\mathbf{v},\mathbf{v}')$ is a corresponding pair between current part and D^t by iteratively using nearest neighbour correspondences, $[\mathbf{R}\ \mathbf{t}]$ is the transformation around $\mathbf{c}$ for current part.

Given the transformation $[\mathbf{R}\ \mathbf{t}]$ above, the global rigid transformation for current part can be updated as $\mathbf{T}=[\mathbf{R}\mathbf{R}_p\ \mathbf{R}(\mathbf{t}_p-\mathbf{c})+\mathbf{c}+\mathbf{t}]$. Repeat the process until all joint transformations are estimated to align with scan. We denote the final aligned mesh as Y_A^t.

Although Y_A^t is finely registered to the scan data part by part, in areas nearest the joints, the result always exhibits drastic deformation, so we estimate a nearest correspondences while ignoring the correspondences near joint areas. Using these correspondences combined with Y^{t-1} as an initial input, we can estimate Y^t with an iterative procedure using Eq. (2).

5.3. Motion reconstruction

For each frame, after estimating the SCAPE mesh Y^t for frame t, we need to transfer its motion to the template mesh M. Although the triangle correspondences between Y^1 and M are built already in Section 5.1, we cannot directly transfer the pose using Eq. (3) because the deformation defined in Eq. (3) is performed in gradient space, the deformed template mesh might not align well with D^t because of vertex offsets. Since Y^t is aligned well with D^t in general, we can setup a point to point correspondence between D^t and Y^t, and transfer the correspondence relations to M and D^t, denoting vertex correspondences between M and D^t as C^t. Then M^t can be solved by minimizing the energy with constraints defined as

$$\arg\min_{\mathbf{v}} E_t+\gamma \sum_{(\mathbf{v}_i,\mathbf{v}_i')\in C^t} \|\mathbf{v}_i-\mathbf{v}_i'\|^2 \tag{4}$$

Where γ is used to control the transfer result and vertex constraints after we transfer the estimated SCAPE mesh's pose to the template mesh, we need to add the details of current scan to the deformed result. A similar procedure is proposed in Li et al. [15] which employs a detailed coefficient estimation step. Since Li's method requires an accurate scan data and dense template, we resort to a gradually relaxed optimization strategy for preserving the smoothness of the mesh in general. The deformation model we use here is a Laplacian deformation model:

$$\arg\min_{\mathbf{v}} w_1\|L\mathbf{v}-\delta\|^2+w_2\sum_{(\mathbf{v}_i,\mathbf{v}'_i)\in C}\|\mathbf{v}_i-\mathbf{v}'_i\|^2$$

where L is the cotangent Laplacian matrix, δ are differential coordinates, C denotes the set of correspondence pairs using nearest correspondences.

We solve the above problem by fixing the smooth weight $w_1=1.0$ and iteratively increasing the weight of regular term w_2 with an initial guess $w_2=0.01$.

As a final step, we need to perform a temporal smoothing to keep spatial continuous in case of slight but noticeable flickers between adjacent frames. Denote the i-th vertex of reconstructed detailed mesh at frame t as $\mathbf{v}_i^t$, then the new vertex coordinate is updated as

$$\mathbf{v}_i^t \leftarrow \mathbf{v}_i^t+\left(\frac{\mathbf{v}_i^{t-1}+\mathbf{v}_i^{t+1}-2\mathbf{v}_i^t}{4}\right)\exp^{(-\|\mathbf{v}_i^{t+1}-\mathbf{v}_i^t\|^2-\|\mathbf{v}_i^{t-1}-\mathbf{v}_i^t\|^2)/\sigma^2} \tag{5}$$

where σ denotes the distance variance, here we set $\sigma=0.05$.

6. Experimental results

In this section, we evaluate our method and show results from depth data containing various kinds of motions. The pipeline is implemented with C++, and tested on a desktop PC with Quad 2.8 GHz CPU and 4 GB RAM. In the tracking stage, the articulated ICP tracking step is less than 1 second while the bottle-neck cost is the nonrigid estimation for an accurate SCAPE pose mesh, which is about 20 seconds per frame. The pose transfer stage usually takes a few seconds, and the final detail aggregation step usually takes about 10 seconds per frame.

6.1. Evaluation

With/without the articulated rigid tracking: We show the importance of the articulated rigid tracking before SCAPE pose estimation. As illustrated in Fig. 6, without an articulated ICP estimation, areas of the left knee and feet palm exhibit a flattened deformation due to the wrong correspondences (Fig. 6(c)). Although the SCAPE model implicitly contains a rigid transformation, it still results in an error deformation once most of the vertices have incorrect correspondences. In contrast, with the articulated ICP to rigidly register the body part, most vertices will find a correct correspondences, which result in a later correct SCAPE pose mesh.

With/without SCAPE database expansion: Fig. 7 shows the necessity of the SCAPE database expansion in our system. We compared the SCAPE result of the same input (Fig. 7(a)) with parameters trained from original pose database (Fig. 7(b)) and expanded pose database (Fig. 7(c)). For simplicity, we use the same

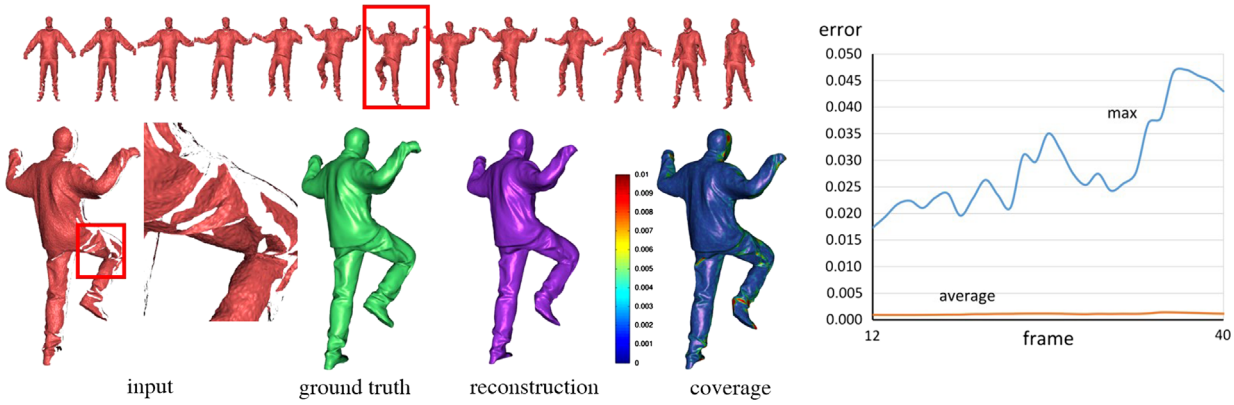

Fig. 5. Comparison to ground truth. The top row shows each frame of input sequence. The color-coded image indicates the accuracy of our result to the ground truth where the range of colorbar is [0, 1] cm. The graph shows the maximum and average error distance of our result to ground truth for each frame. As the pose becomes far way from the initial pose, the maximum error increases, but the average error still keeps steady. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

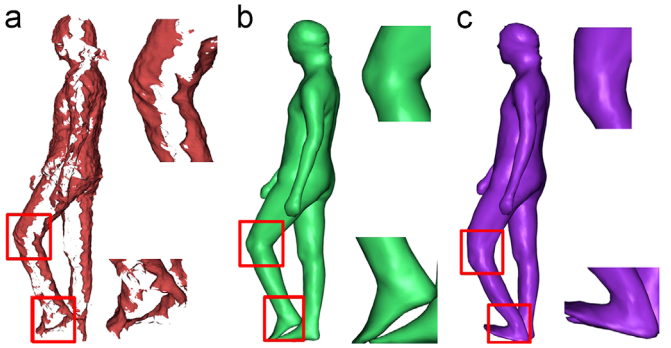

Fig. 6. Comparison of SCAPE estimation between with/without articulated ICP. (a) input data, (b) the SCAPE result with articulated ICP method, and (c) result without articulated ICP.

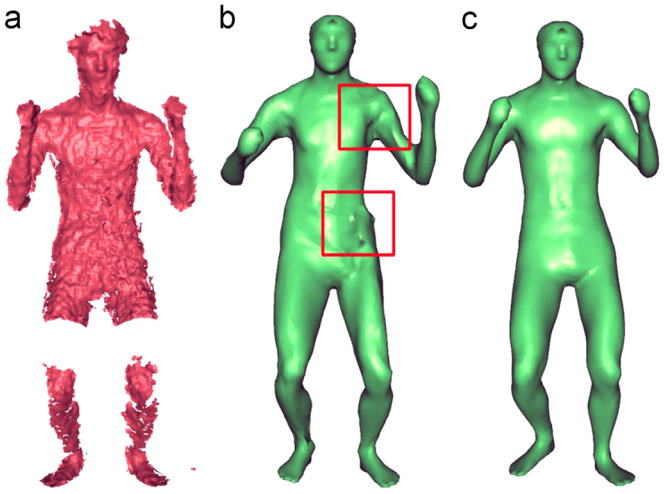

Fig. 7. Comparison of SCAPE result between with/without SCAPE database expansion. (a) input data, (b) SCAPE result without database expansion, and (c) SCAPE result with database expansion.

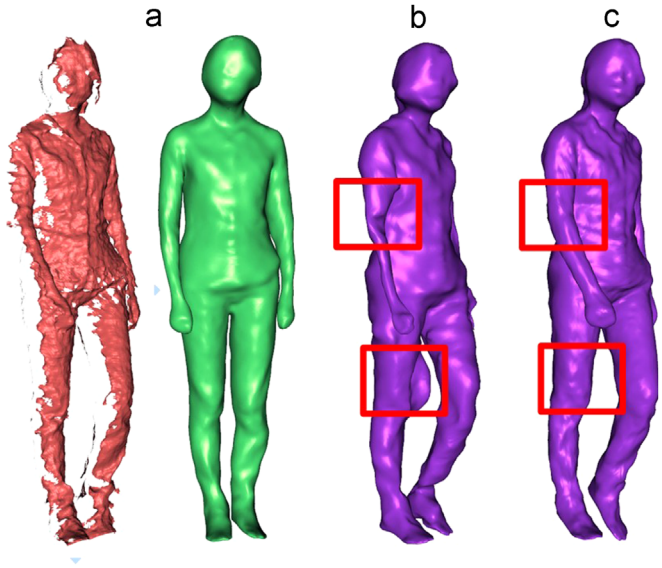

Fig. 8. The comparison between the result of graph-based method (b) and our method (c). The input (a) includes a template and scan data.

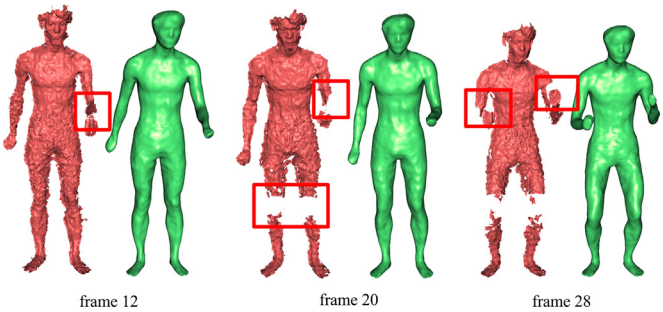

Fig. 9. As shown in the rectangles, the area near elbows disappears and then reappears, our method is robust to handle these issues.

Kinect scan sequences and manual correspondence to estimate the results. As shown in Fig. 7(b), the result using original training data exhibits much unnatural deformations, which is unsuitable for motion reconstruction in our experiment.

Comparison to ground truth: To evaluate the reconstructed result's detail accuracy, we use a performance sequences released by Vlasic et al. [21] as ground truth. In order to simulate a similar environment to our setup, we removed the faces that are opposite to all view directions, as well as those being occluded, and we also add noises to the whole mesh. The result of our method to track and reconstruct these incomplete meshes is shown in Fig. 5. The reconstructed meshes fit ground

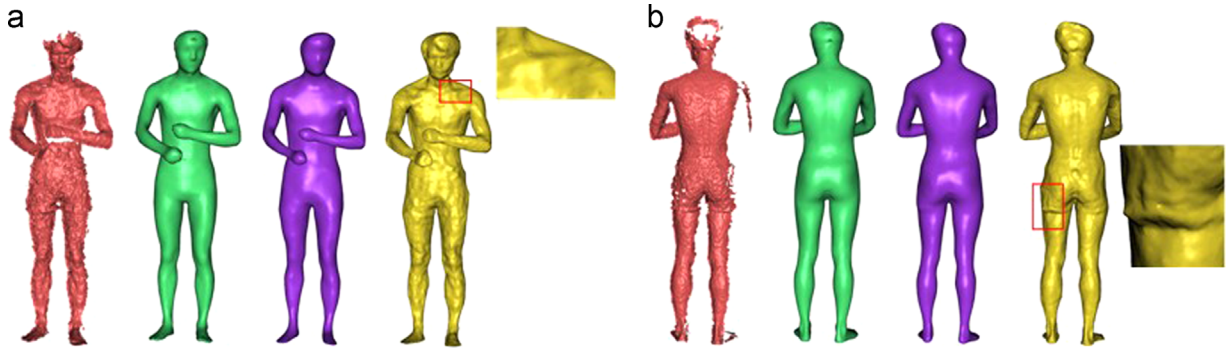

Fig. 10. The reconstructed result of a "rolling-hand" pose. For each subfigure, we show the input data, estimated SCAPE mesh, pose-induced template mesh, final reconstructed mesh. (a) Front view of the "rolling-hand" data and (b) Back view of the "rolling-hand" data.

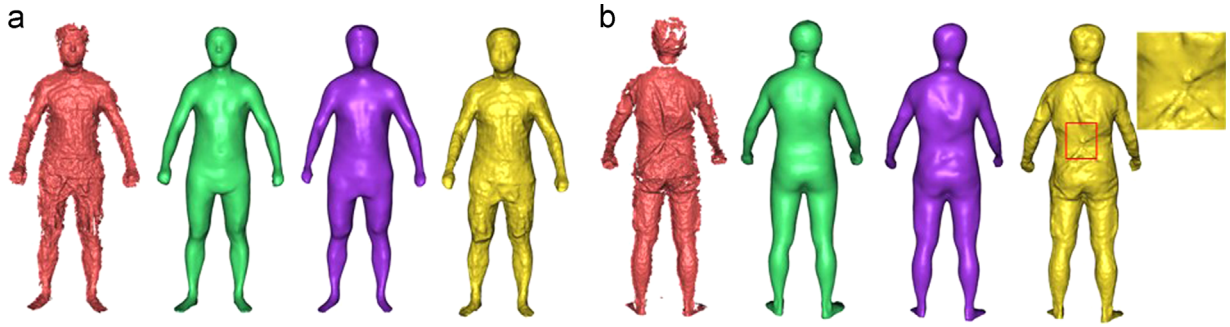

Fig. 11. Reconstruction result of motion with clothes. The meaning of each column is same to Fig. 10. (a) Front view of the "tight-cloth" data and (b) Back view of the "tight-cloth" data.

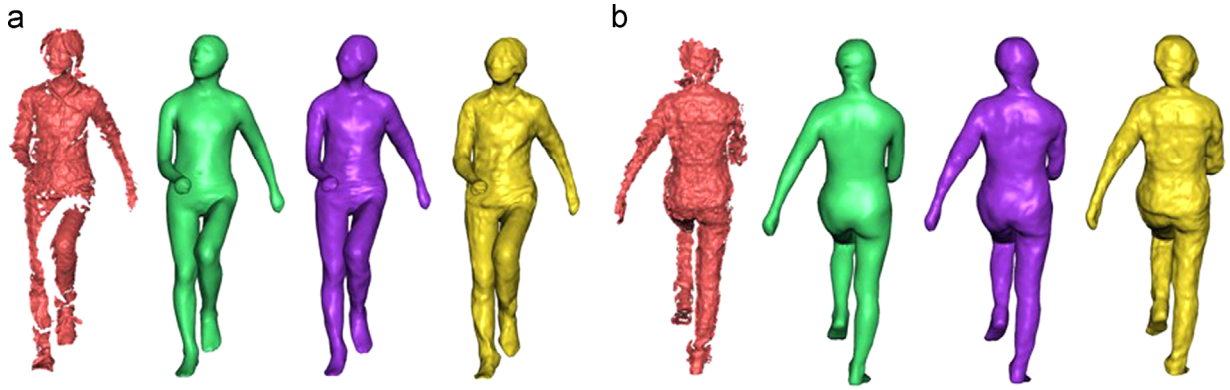

Fig. 12. Reconstruction result of walking pose of another body shape with tight clothes. The meaning of each column is same as Fig. 10. (a) Front view of the "walking" data and (b) Back view of the "walking" data.

truth with an average distance in the hole regions no more than 0.5 cm and an average distance in the regions overlapping with the input no more than 0.2 cm.

Comparison to graph-based deformation: We compared our SCAPE-induced deformation result with the result of graph-based deformation [32,26]. Since the graph method requires a template as input, we use the same result of our pipeline as described in Section 5.1. As shown in Fig. 8, due to the large occlusions and inaccurate correspondences in the knee area, the graph-based deformation method is unable to perform a correct deformation. Furthermore, the parameters balanced the smoothness and constraints of the graph based method is rather difficult to adjust. However, as shown in the figure, with the articulated ICP tracking, each part of the SCAPE mesh is able to align roughly well with the scan, coupled with the SCAPE model, the result seems more natural and contains fewer drastic deformations.

Deal with occlusion: Fig. 9 demonstrates the ability of our system to cope with serious occlusions. The input is a sequence of data where the left elbow firstly disappear and then reappear in the consecutive frames, and our SCAPE-induced deformation combined with articulated ICP tracking is able to complete these gaps, and achieves plausible shape in these regions.

6.2. More results

To show our method on different body shapes at different poses, we conduct an experiment on three performers with various poses. Specifically, it includes a performer in rolling-hand pose (Fig. 10), a performer in a normal pose with tight clothes (Fig. 11), and a performer in walking pose with tight clothes (Fig. 12).

As shown in Fig. 10, the details near the shoulder are reconstructed which are not shown in the template mesh and

the SCAPE mesh. In the area of the left hip, the pant's boundary is clear to be seen. Furthermore, there is an incomplete region in the right arm, but the reconstructed result fills these gaps in a smooth way.

In Fig. 11, we show a result with details of clothes, which is very apparent in the back of the performer. The details is not completely exhibited in the reconstructed result due to balance of smoothness in case of noises.

In Fig. 12, we show the reconstructed result of the walking pose data. With the aid of SCAPE model, the occlusion part of front body still can be reconstructed well.

7. Conclusion

In this paper, we proposed a performance reconstruction system based on SCAPE. It is able to reconstruct a complete mesh from noisy and incomplete point clouds with a plausible detail while preserving the topology of the mesh. With the articulated ICP method, the tracking stage is robust to fast motion. We also proposed a method to expand the pose training database for SCAPE. The work still has lots of limitations due to the naked SCAPE human model and articulated motion assumptions. In fact, the freedom of joint limitation should be considered in our future work, as well as the extension to a performer with loose clothes.

Acknowledgment

Many thanks to the anonymous reviewers for their valuable comments. Also thanks to Yangzi Ding for the scanning data. This work was partially supprted by NSFC (No. 61379068).

Appendix A. Supplementary material

Supplementary data associated with this paper can be found in the online version of http://dx.doi.org/10.1016/j.cag.2013.10.023.

References

[1] Anguelov D, Srinivasan P, Koller D, Thrun S, Rodgers J, Davis J. Scape: shape completion and animation of people. ACM Trans Graph 2005;24(3):408–16.
[2] Tong J, Zhou J, Liu L, Pan Z, Yan H. Scanning 3d full human bodies using kinects. IEEE Trans Vis Comput Graph 2012;18(4):643–50.
[3] Hasler N, Stoll C, Rosenhahn B, Thormählen T, Seidel H-P. Technical section: estimating body shape of dressed humans. Comput Graph 2009;33:211–6.
[4] Balan AO. Detailed human shape and pose from images [Ph.D. thesis], Providence, RI, USA, aAI3430086; 2010.
[5] Bălan AO, Black MJ. The naked truth: estimating body shape under clothing. In: Proceedings of the 10th European conference on computer vision: part II, ECCV '08. Berlin, Heidelberg: Springer-Verlag; 2008. p. 15–29.
[6] Guan P, Weiss A, Balan AO, Black MJ. Estimating human shape and pose from a single image. In: IEEE 12th international conference on Computer vision, 2009. IEEE; 2009. p. 1381–8.
[7] Zhou S, Fu H, Liu L, Cohen-Or D, Han X. Parametric reshaping of human bodies in images. In: ACM SIGGRAPH 2010 papers, SIGGRAPH'10. New York, NY, USA: ACM; 2010. p. 126:1–10.
[8] Mundermann L, Corazza S, Andriacchi TP. Accurately measuring human movement using articulated icp with soft-joint constraints and a repository of articulated models. In: Computer vision and pattern recognition, 2007. CVPR'07. IEEE Conference on, IEEE. 2007; p. 1–6.
[9] Corazza S, Mündermann L, Gambaretto E, Ferrigno G, Andriacchi TP. Markerless motion capture through visual hull, articulated icp and subject specific model generation. Int J Comput Vision 2010;87(1–2):156–69.
[10] Ganapathi V, Plagemann C, Koller D, Thrun S. Real-time human pose tracking from range data. In: Proceedings of the 12th European conference on computer vision—volume part VI, ECCV'12. Berlin, Heidelberg: Springer-Verlag; 2012. p. 738–51.
[11] Izadi S, Kim D, Hilliges O, Molyneaux D, Newcombe R, Kohli P, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In: Proceedings of the 24th annual ACM symposium on user interface software and technology, UIST '11. New York, NY, USA: ACM; 2011. p. 559–68.
[12] Zeng M, Zhao F, Zheng J, Liu X. Octree-based fusion for realtime 3d reconstruction. Graph Models 2013;75(3):126–36.
[13] Weise T, Bouaziz S, Li H, Pauly M. Realtime performance-based facial animation. ACM Trans Graph 2011;30(4):77:1–10.
[14] Li H, Yu J, Ye Y, Bregler C. Realtime facial animation with on-the-fly correctives. ACM Trans Graph 2013;32(4):42:1–10.
[15] Li H, Adams B, Guibas LJ, Pauly M. Robust single-view geometry and motion reconstruction. ACM Trans Graph 2009;28(5):175:1–10.
[16] Vlasic D, Baran I, Matusik W, Popović J. Articulated mesh animation from multi-view silhouettes. ACM Trans Graph 2008;27(3):97:1–9.
[17] de Aguiar E, Stoll C, Theobalt C, Ahmed N, Seidel H-P, Thrun S. Performance capture from sparse multi-view video. ACM Trans Graph 2008;27(3):98:1–10.
[18] Gall J, Stoll C, De Aguiar E, Theobalt C, Rosenhahn B, Seidel H-P. Motion capture using joint skeleton tracking and surface estimation. In: IEEE conference on computer vision and pattern recognition, 2009. CVPR 2009. IEEE; 2009. p. 1746–53.
[19] Ye G, Liu Y, Hasler N, Ji X, Dai Q, Theobalt C. Performance capture of interacting characters with handheld kinects. In: Proceedings of the 12th European conference on computer vision—volume Part II, ECCV'12. Berlin, Heidelberg: Springer-Verlag; 2012. p. 828–41.
[20] Stoll C, Gall J, de Aguiar E, Thrun S, Theobalt C. Video-based reconstruction of animatable human characters. ACM Trans Graph 2010;29(6):139:1–10.
[21] Vlasic D, Peers P, Baran I, Debevec P, Popović J, Rusinkiewicz S, et al. Dynamic shape capture using multi-view photometric stereo. ACM Trans Graph 2009;28(5):174:1–174:11.
[22] Curless B, Levoy M. A volumetric method for building complex models from range images. In: Proceedings of the 23rd annual conference on computer graphics and interactive techniques. ACM; 1996. p. 303–12.
[23] Li H, Luo L, Vlasic D, Peers P, Popović J, Pauly M, et al. Temporally coherent completion of dynamic shapes. ACM Trans Graph 2012;31(1):2:1–2:11.
[24] Kazhdan M, Bolitho M, Hoppe H. Poisson surface reconstruction. In: Proceedings of the 4th eurographics symposium on geometry processing; 2006.
[25] Chang W, Zwicker M. Global registration of dynamic range scans for articulated model reconstruction. ACM Trans Graph 2011;30(3):26:1–15.
[26] Zeng M, Zheng J, Cheng X, Liu X. Templateless quasi-rigid shape modeling with implicit loop-closure. In: Proceedings of the 2013 IEEE conference on computer vision and pattern recognition, CVPR '13. p. 145–52.
[27] Svoboda T, Martinec D, Pajdla T. A convenient multicamera self-calibration for virtual environments. Presence Teleoper Virtual Environ 2005;14(4):407–22.
[28] Sumner RW, Popović J. Deformation transfer for triangle meshes. ACM Trans Graph 2004;23(3):399–405.
[29] Hasler N, Stoll C, Sunkel M, Rosenhahn B, Seidel H-P. A statistical model of human pose and body shape. In: Computer graphics forum, vol. 28. Wiley Online Library; 2009. p. 337–46.
[30] Starck J, Hilton A. Surface capture for performance-based animation. IEEE Comput Graph Appl 2007;27(3):21–31.
[31] Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm. In: Proceedings of 3rd international conference on 3-D digital imaging and modeling, 2001. IEEE; 2001. p. 145–52.
[32] Sumner RW, Schmid J, Pauly M. Embedded deformation for shape manipulation. In: ACM SIGGRAPH 2007 papers. SIGGRAPH '07. New York, NY, USA: ACM; 2007.