Computers & Graphics 129 (2025) 104229

Contents lists available at ScienceDirect

Computers & Graphics

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

Technical Section

Parametric 3D human modeling with biharmonic SMPL☆

Yin Chen [a,*], Yuping Ye [b], Weiwei Xu [c], Qiliang Yang [a], Qizhen Zhou [a]

[a] School of Defense Engineering, Army Engineering University of PLA, Nanjing, 210007, China
[b] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[c] State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China

ARTICLE INFO

Keywords:
3D human body parametric models
Biharmonic
Data driven
Model training

ABSTRACT

3D human body parametric models are high-level semantic information that can provide effective prior knowledge for 3D human body reconstruction. The existing parametric models all use manual rigid part segmentation, which is difficult to ensure rationality. Meanwhile, without considering the local similarity of parameters, it is difficult to ensure the generation of a smooth model. We propose parametric 3D human modeling with biharmonic SMPL to address the above issues. Firstly, based on the SMPL model, biharmonic constraints are introduced into the vertex skinning deformation and non-rigid deformation processes to ensure that the parameters satisfy local similarity. Then rigid parts are segmented automatically based on the optimized skinning weights. Finally, in order to increase the solution space and obtain more reasonable model parameters, we propose a coregistration framework from the original point cloud to accomplish parametric model training and mesh registration. During the coregistration process, biharmonic constraints are introduced, and the process is transformed into several optimization problems. Experiments conducted on the large dataset consisting of SCAPE, CAESAR and FAUST demonstrate the superiority of our model. Compared with manual segmentation methods, the skinning weight distribution and rigid part segmentation of our model are more reasonable. Compared with the SMPL model, the area with fitting errors exceeding 5 mm in our model decreases by 17.6% after skinning weight optimization. And the area with smoothing errors exceeding 5 mm decreases by 83.2% after optimization of non-rigid deformation parameters.

1. Introduction

3D human body parametric models [1] represent the deformation of the human body shape as a subspace spanned by attributes such as body shape and pose, offering a simpler and more efficient modeling approach. They are widely used in areas such as virtual fitting, virtual training, and film production. The pioneering work, SCAPE (Shape Completion and Animation of People) [2], applies human body deformation to triangular faces. Each face undergoes non-rigid deformation, shape deformation, and rigid deformation successively, ultimately obtaining consistent vertex positions by solving the least squares equation. Subsequent work focus on optimizing the performance of parametric models, such as BlendSCAPE [3], SMPL (Skinned Multi-Person Linear Model) [4], SymmetricSCAPE [5], and using parametric models as high-level prior knowledge for reconstruction and recognition problems, such as dynamic reconstruction [6,7] and image-based reconstruction [8,9].

The most commonly used parametric model is SMPL whose workflow is similar to the Linear Blending Skinning(LBS). In the SMPL model, the human body is divided into several rigid bones, and the displacement of each vertex includes the local offset caused by the relative transformation of adjacent rigid bones and the skinning deformation caused by the absolute transformation of rigid bones. The pros and cons of SMPL depend on the rigid part segmentation of human bones. Reasonable segmentation helps to generate more realistic results in joint areas. However, the current human body parametric models use manual segmentation, which cannot guarantee optimality. The pros and cons of SMPL also depend on the choice of skinning weights. Bounded biharmonic weights minimize the Laplacian energy of the weight distribution under bounded constraints to generate smooth and intuitive deformations with handles or bones. Jacobson et al. [10] firstly proposed bounded biharmonic skinning deformation weights, which they used for real-time deformation, achieving smooth transformations of the model. The general principle of 3D human parametric models is to categorize the factors that induce human body shape deformation, such as SCAPE [2] and BlendSCAPE [3], which divide deformation

☆ This article was recommended for publication Marcel Campen.

* Corresponding author.
E-mail addresses: chen.yin@aeu.edu.cn (Y. Chen), yp.ye@siat.ac.cn (Y. Ye), xww@cad.zju.edu.cn (W. Xu), yql@893.com.cn (Q. Yang), zhouqz@aeu.edu.cn (Q. Zhou).

https://doi.org/10.1016/j.cag.2025.104229
Received 27 September 2024; Received in revised form 13 April 2025; Accepted 14 April 2025
Available online 30 April 2025

into rigid and non-rigid deformation induced by pose and non-rigid deformation induced by shape. Deformations are obtained through the linear blending of pose and shape variables. The blending weights are trained in a data-driven manner, and their distribution affects the quality of model deformation. To achieve smooth deformation, it is necessary to ensure that neighboring areas have similar weights, meaning the weights conform to biharmonic constraints. However, in the early definition and training process of parametric models, this constraint was not considered, making it difficult to ensure the smooth deformation of the parametric models. Meanwhile, the commonly used vertex offset model SMPL is based on the registered mesh model obtained by the SCAPE or BlendSCAPE method during training, which results in the mesh being fixed and unable to adaptively adjust during the model training process, actually limiting the solution space of the model parameters.

To address the above issues, we incorporate biharmonic constraints into the 3D human body parametric models and propose parametric modeling with biharmonic SMPL. The biharmonic constraint is applicable to all 3D human body parametric models. Based on the SMPL [4] model, we consider this constraint in the vertex skinning deformation and non-rigid deformation processes induced by pose. Therefore, local similarity of the parameters can be ensured. We also segment the human body automatically based on optimized skinning weights to get more reasonable rigid parts. Meanwhile, in order to increase the solution space and obtain more reasonable model parameters, we propose a coregistration framework from the original point cloud to accomplish parametric model training and mesh registration. During the coregistration process, biharmonic constraints are introduced, and the training process is transformed into solving problems of several minimum energy functions. Experiments on the large dataset consisting of SCAPE, CAESAR and FAUST show that compared to manual segmentation methods, the skinning weight distribution and rigid part segmentation in this paper are more reasonable. Compared with the SMPL model, our model improves the fitting and smoothness characteristics significantly after parameter optimization.

The main contributions of this paper include:

1. We propose parametric 3D human body modeling with biharmonic SMPL, which achieves smooth deformation of the human body mesh;
2. We propose an automatic segmentation method for human rigid parts, achieving more reasonable segmentation;
3. We propose a new coregistration framework for our biharmonic SMPL parametric model, which accomplishes parametric model training and mesh registration simultaneously.

2. Related work

This section provides a brief overview of works in the domain of 3D human body parametric models, applications of these models, and works related to biharmonic constraints.

2.1. 3D human parametric models

The groundbreaking SCAPE [2] mathematical model has limited representational capabilities and slow solving speeds, requiring pre-established landmarks to process individual instances. These limitations were gradually overcome in subsequent works. For instance, statistical SCAPE [11] utilizes the invariance mechanism of local rotational transformations to jointly encode 3D human body shape and pose deformations, speeding up the solving process. More importantly, it proves that the reconstruction of 3D human parametric models could be achieved with a limited number of landmarks. The BlendSCAPE [3] model simplifies the rigid pose deformations of the 3D human body into linear blend skinning deformations, reducing distortion at joints and combining the establishment of a 3D human database with the training of the mathematical model into one process. The SCAPE model separately models pose and shape deformations, assuming the same pose deformations for different individuals in the same pose. However, human pose deformations vary from person to person. Chen et al. [12] addressed this problem by jointly modeling pose and shape deformations using tensor representation, enhancing the model's representational power, especially in joint deformations. Realtime SCAPE [13] achieves real-time reproduction of human motion through pre-computation of skinning weights. The symmetricSCAPE [5] model extracts the inherent symmetry of body shape and pose from human spatial data, significantly enhancing the descriptive power of the parametric models. Bogo et al. [14] released the textured European FAUST human body database. Based on FAUST, Tsoli et al. [15] proposed an improved SCAPE model that accounted for detailed variations induced by breathing and other factors. The Delta model [6] achieves dynamic reconstruction under monocular depth cameras, preserving geometric details and texture features. Bogo et al. [16] extends the popular FAUST dataset to dynamic 4D data and releases Dynamic FAUST dataset. The SMPL mathematical model [4] adopts a vertex-based linear blend skinning mechanism, linearly expressing 3D human body shape and pose deformations, and ensuring compatibility with industrial modeling software like Maya. Li et al. [17] parameterized facial deformations, distinguishing between deformations induced by facial shape and expressions, thus expanding the scope of human parametric models. Romero et al. [18] built upon the human limb foundation to parameterize hand gesture deformations, further enhancing the expressiveness of human parametric models. Joo et al. [19] proposed a unified parametric model representing human limbs, gestures, and facial expressions. Pavlakos et al. [8] introduced the more powerful unified parametric model SMPL-X, which significantly improved visual presentation results. The STAR model [20] extends the body space of SMPL with more data and associates non-rigid vertex displacements caused by pose with body mass index (BMI) by concatenating pose and body variable parameters, while introducing sparsity to improve generalization. Osman et al. train an expressive human body model called SUPR [21], where each joint strictly influences a sparse set of model vertices. Keller et al. develop SKEL [22], which re-rigs the SMPL body model with a biomechanics skeleton and has more accurate joint locations than SMPL.

BlendSCAPE [3] optimizes both the data-to-model correspondences and the human model parameters (pose and shape), but needs a reliable initial solution. Bhatnagar et al. [23] proposed LoopReg: an end-to-end learning framework to register a corpus of scans to a common 3D human model. LoopReg includes a backward map and a backward map. The former predicts the correspondence from every scan point to the surface of the human model while the latter transforms the corresponding points back to the scan based on the model parameters.

Neural network approaches aim to enhance the expressive power by replacing or augmenting parts of the original model with neural networks. NASA [24] represents the 3D human model as a combination of independent parts, each represented by an occupancy network. Rigid transformations of these parts, based on input skeletal transformations, resulted in deformed shapes. While this formulation preserves the global structure post-joint connection, it disrupts the continuity of surface deformations, leading to artifacts at the intersections of body parts. Moreover, NASA requires real surface skinning weights to learn correct part allocations. SNARF [25] learns forward skinning weights without such supervision and captured pose-related deformations, addressing the aforementioned issues. Learned implicit function cannot retain the detail in the input data. To solve this problem, Chibane et al. [26] proposed Implicit Feature Networks (IF-Nets). IF-Nets extracts a learnable 3-dimensional multi-scale tensor of deep feature and classifies deep features extracted from the tensor at a continuous query point, thus forces the model to make decisions based on global and local shape structure. The fundamental principle of neural network methods is similar to vertex displacement, hence they also do not resolve the smooth

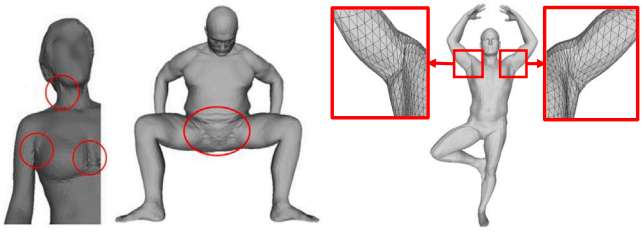

Fig. 1. Reconstruction results of different parametric models: (a) Original BlendSCAPE model [3]; (b) BlendSCAPE [3]; (c) SMPL [4].

deformation problem. X-Avatar [27] further expands upon these by incorporating innovative sampling techniques and part-aware skinning modules within the SMPL-X framework for more detailed animations and textures. Despite these strides, challenges remain, particularly in ensuring smooth deformations and addressing deformation similarity discrepancies, as depicted in Fig. 1.

2.2. Applications of 3D human parametric models

3D human body parametric models have driven innovations across several applications [28]. Guan et al. [9] utilized the SCAPE human body parametric model, employing constraints from 2D image contours and human parameters such as height and weight, to instantiate and solve for the corresponding 3D human mesh within a 2D image. Weiss et al. [29] enhanced the accuracy of the SCAPE model's solutions by using multiple depth images captured by a single camera as constraints, reducing the ambiguity of 2D image-constrained solutions. Zheng et al. [30] deployed four cameras around the user to simultaneously capture their movements, decreasing the incompleteness of the point cloud established from depth images and improving the quality of the reconstructed 3D mesh. Leveraging the human body parametric models, Bălan et al. [31] attempted to reconstruct the naked model of a body under normal clothing conditions. The algorithm took photos of the same person in multiple poses as input and output the naked model of the body. Zhang et al. [32] reconstructed the naked model of a body from a sequence of 3D scans taken under normal clothing conditions, with experiments showing high accuracy of the reconstructed naked models. Other researchers [7,13,33,34] attempted to simultaneously reconstruct the shape (both shape and pose) from dynamic sequences captured by a single depth camera. In [35], the authors employ the Pseudo-Linear Inverse Kinematic Solver (PLIKS) for 3D human body reconstruction from 2D images, leveraging a linearized SMPL model to enable precise analytical reconstruction and the flexible integration of camera calibration data. 3DPMesh [36] presents a training-free methodology for 3D human mesh reconstruction from 2D images, which amalgamates 3D Pose estimation, enhancement, and Mesh articulation, thereby economizing computational resources and abbreviating inference time.

2.3. Biharmonic constraints

Jacobson et al. [10] revolutionized linear blend skinning (LBS) with the introduction of bounded biharmonic weights, a method that automates weight computation for smoother deformations. Building on this, Yuan et al. [37] proposed a data-driven approach that incorporates biharmonic, rigidity, and regularization constraints into a coherent least squares problem. Their findings underscore the potential of these constraints to significantly enhance deformation realism, especially in localized model regions, paving the way for more natural and believable animations.

3. Biharmonic 3D human parametric model

Analytical forms of 3D body parametric models can be classified into two main types: triangular deformation and vertex displacement. Vertex displacement models utilize a linear blend skinning deformation mechanism to describe body shape and pose deformations quasi-linearly. It enhances compatibility with industrial modeling software such as Maya, leading to widespread academic application. This paper elaborates on a biharmonic 3D body parametric model using vertex displacement as an example.

3D body parametric models represent the human body $T(\theta,\beta)$ as subspaces influenced by shape variables β and pose variables θ. Specifically, for vertex displacement-based models like SMPL and SMPL-X, each vertex undergoes shape-guided non-rigid displacement $D_V(\beta)$, pose-guided non-rigid displacement $B_V(\theta)$, and skinning deformation $W_V(\theta)$ sequentially to obtain the final vertex coordinates. To compute joint translations in skinning deformations, the position of each joint in a standard pose must be calculated. It is correlated with the body shape and is represented by a function $J(\beta)$.

The proposed model follows similar principles but incorporates biharmonic constraints on deformation parameters during the training process. When in a standard pose, a 3D human template $\bar{T}$ consists of vertices $V=\{v_0,\dots,v_i\cdots\}$, edges $E=\{e_0,\dots,e_i,\dots\}$, and faces $F=\{f_0,\dots,f_i,\dots\}$ ($|V|=15\,132$, $|E|=45\,390$, $|F|=30\,260$). The deformed 3D body can be described by an articulated model with $K(K=19)$ rigid parts forming a tree structure. The 0th rigid part is the root node, and the rotation of the root node relative to the template $\bar{T}$ is represented by the Rodrigues rotation vector θ_0. Each other rigid part has a unique parent node, and its rotation relative to the parent node is represented by Rodrigues rotation vector θ_i, forming the pose $\theta=\{\theta_i,0\le i<K\}$.

Vertex skinning deformation $W_V(\theta;\mathcal{W}_V)$ is achieved through weighted contributions from rigid transformations of all rigid parts. Common methods include linear blend skinning (LBS) and dual quaternion skinning (DQS). The set of weights contributed by rigid part to vertex v is denoted as $w_v=\{w_{v,i},0\le i<K\}$, while the collection of all vertex weights is $\mathcal{W}_V=\{w_v,v\in V\}$.

Since using identical skinning deformation for all human bodies can result in substantial local distortions for certain poses, pose-guided vertex non-rigid displacements $B_V(\theta)$ are introduced to compensate local deformations caused by different poses. This is represented by a linear regression equation $B_V(\theta;\mathcal{B}_V)$ for pose variables, where the non-rigid displacement for vertex v is given by $B_v(\theta;b_v)$. Here, $b_v=\{b_{v,i},0<i<K\}$ represents the regression coefficients for the rotation vector on vertex displacement and $\mathcal{B}_V=\{b_v,v\in V\}$ embodies the set of all such coefficients. The parameters for pose deformation need pre-training and include both $\mathcal{W}_V$ and $\mathcal{B}_V$.

Shape-guided non-rigid displacement $D_V(\beta;d_v)$ is defined as the regression of shape variables on vertex displacement. For vertex v, the displacement is represented by $D_v(\beta;d_v)$, where $d_v=\{d_{v,i},0<i<|\beta|\}$ denotes the regression coefficients for shape variables and $D_V=\{d_v,v\in V\}$ includes the set of all such coefficients.

The analytical expression for vertex displacement-based human parametric models can be described as: $W_V(\bar{T}+D_V(\beta;D_V)+B_V(\theta;\mathcal{B}_V),J(\beta;\mathcal{J}),\theta;\mathcal{W}_V)$. The complete model parameters to be trained are $\{\mathcal{W}_V,\mathcal{B}_V,D_V,\mathcal{J}\}$. During pose deformation, adjacent vertices should exhibit similar deformations. To achieve this, biharmonic constraints are introduced into the training of parameters $\mathcal{W}_V$ and $\mathcal{B}_V$.

3.1. Biharmonic constraints on vertex skinning deformation parameters

In previous work on parametric modeling, rigid body parts were manually segmented, and skinning weights for vertices were specified manually rather than learned from data. It has two limitations. Firstly, unreasonable segmentation of rigid parts makes it difficult to effectively represent the rigid segmentation of the human body. Secondly, the

significant difference in skinning weights between neighboring vertices leads to obvious local deformation distortion during pose deformation.

To address above issues, this paper employs data-driven training to derive biharmonic skinning weights $\mathcal{W}_V$ and automatically segment rigid parts based on these weights. Skinning deformation expresses vertex transformations as a weighted sum of rigid deformations, formulated to satisfy the following equations:

$$\begin{cases} p_v = \sum\limits_{i \in N(I(v))} w_{v,i}(R_i \tilde{p}_v + t_i) \\ w_{v,i} \geq 0 \\ \sum\limits_{i \in N(I(v))} w_{v,i} = 1 \end{cases} \tag{1}$$

where p_v represents the transformed vertex position, $w_{v,i}$ denotes the skinning weight of vertex v with respect to the ith rigid part. Let $I(v)$ represent the rigid part to which a vertex v belongs. $N(I(v))$ denotes the set of neighboring parts (including $I(v)$ itself). The rotation of the ith rigid part is denoted as R_i, and the translation is t_i.

We propose to define the weights $w_{*,i}$ as minimizers of a higher-order shape-aware smoothness functional, namely, the Laplacian energy, subject to constraints that enforce interpolation of the rigid parts:

$$\underset{w_{*,i}, i=0,\ldots,K-1}{\arg\min} \frac{1}{2} \int_{\Omega} \left\| \Delta w_{*,i} \right\|^2 dV \tag{2}$$

where $w_{*,i}$ stands for the weight distribution of ith rigid part. The above Laplacian energy is discretized using the standard linear FEM Laplacian $M^{-1}L$ where M is the lumped mass matrix (with Voronoi area M_i of vertex v_i on each diagonal entry i) and L is the symmetric stiffness matrix. After discretizing the continuous integral term, we have

$$\sum_{i=0}^{K-1} \frac{1}{2} \int_{\Omega} \left\| \Delta w_{*,i} \right\|^2 dV = \frac{1}{2} \sum_{i=0}^{K-1} {w_{*,i}}^T (LM^{-1}L) w_{*,i} \tag{3}$$

which is equivalent to

$$\frac{1}{2} \sum_{i=0}^{K-1} M^{-1} \left\| L w_{*,i} \right\|^2 \tag{4}$$

For more details of derivation, please refer to Jacobson et al. [10].

Accordingly, the Laplacian energy caused by biharmonic constraint on vertex skinning weights is defined as follows:

$$\frac{1}{a_v} \sum_{i=0}^{K-1} \left\| \sum_{u \in N(v)} \varphi_{u,v}(w_{v,i} - w_{u,i}) \right\|^2 \tag{5}$$

Here, $N(v)$ represents the one-ring neighborhood of vertex v. a_v is the Voronoi area of vertex v. The Laplacian operator $\varphi_{u,v} = \delta / \|u - v\|$ follows the definition method proposed by Taubin [38] and is normalized accordingly. δ denotes the average edge length in $\bar{T}$, and $\varphi_{u,v}$ represents the harmonic weight between neighboring vertices u and v which are determined by the structure of the mesh and remain fixed during training.

By imposing these biharmonic constraints, the model ensures smooth transitions of skinning weights across neighboring vertices, thereby reducing potential deformation artifacts and enhancing the realism of the pose deformation.

3.2. Biharmonic constraints on vertex non-rigid deformation parameters

Non-rigid vertex displacements are defined by a linear regression on pose variables $\theta = \{\theta_i, 0 < i < K\}$ as follows: $B_v(\theta; b_v) = \sum_{i=1}^{K-1} b_{v,i}\theta_i$. Here, $b_{v,i}$ denotes the linear coefficient matrix (of size 3×3) for the ith Rodrigues rotation vector θ_i applied to vertex v. The collection of regression coefficients for all vertices is denoted by $\mathcal{B}_V = \{b_v, v \in V\}$.

Methods such as SMPL and SMPL-X impose regularization constraints on the parameters to prevent overfitting during training. A weaker constraint allows the model to fit the training data more closely but may result in less smooth transitions in neighboring regions. Conversely, a stronger constraint ensures smoother transitions between neighboring regions but reduces the fitting of the model to the training data. To address the conflict between model fitting and smoothness, we introduce biharmonic constraints and define the Laplacian energy as:

$$\frac{1}{a_v} \sum_{i=1}^{K-1} \left\| \sum_{u \in N(v)} \mu_{u,v}(b_{v,i} - b_{u,i}) \right\|^2 \tag{6}$$

where $\mu_{u,v}$ denotes the weights between neighboring vertices u and v, utilizing Cotangent Laplacian weights. These weights are determined by the mesh structure $\bar{T}$ and remain fixed during training.

Considering the coupling among non-rigid deformation parameters $\mathcal{B}_V$ of all vertices, the number of unknowns is large and challenging to optimize. Therefore, symmetry constraints [5] are applied to reduce the number of unknowns by half. Furthermore, to ensure the sparsity of $b_{v,i}$, only the rotation vectors adjacent to vertex v are considered. For the ith rigid part, with its parent part denoted as $p(i)$ and Rodrigues rotation vector denoted as $\theta_i (0 < i < K)$, θ_i contributes to vertex offset B_v only when the weight $w_{v,i}$ or $w_{v,p(i)}$ are non-negative.

After the above simplifications, the total number of parameters in the optimization process is approximately 120,000, which can be solved directly using matrix decomposition methods.

3.3. Deformation constraints on body shape

The parameters to be optimized for shape deformation $D_V(\beta; D_V)$ are denoted as D_V. To train D_V, it is necessary to obtain the shape of the human body in a standard A-pose, denoted as $\bar{T} + D_V(\beta; D_V)$, and subsequently derive D_V through principal component analysis (PCA). Following the approach of SCAPE, different human bodies in a standard A-pose can be considered as local non-rigid deformations S of a template human body, where each element of S is represented by a 3×3 matrix. For adjacent triangular faces f and g, their non-rigid deformations S_f and S_g should be similar.

Non-rigid deformation S is a crucial component in the parametric models of triangular deformation. For a specific human body with shape variables β, the non-rigid deformation of a triangular face f is represented as $S_f(\beta)$. For the human body in pose θ, the relative positions of adjacent vertices u and v in f satisfy the following equation:

$$p_u - p_v = R_f(\theta) \cdot S_f(\beta) \cdot Q_f(\theta) \cdot (\tilde{p}_u - \tilde{p}_v) \tag{7}$$

The rigid transformation $R_f(\theta)$ of a triangular face f is defined as the weighted sum of the rotation transformations of neighboring rigid parts [3], which is $R_f(\theta) = \sum_{i=0}^{K-1} w_{f,i} R_i(\theta)$. The skinning weight of the ith rigid part for a triangular face f is defined as the arithmetic mean of the skinning weights of the three vertices of f.

Non-rigid transformation $Q_f(\theta; q_f)$ is defined by a linear regression on pose variables $\theta = \{\theta_i, 0 < i < K\}$ as follows:

$$Q_f(\theta) = q_{f,0} + \sum_{i=1}^{K-1} q_{f,i}\theta_i \tag{8}$$

Here, $q_{f,0}$ represents the deformation matrix when there is no relative deformation among rigid parts, given in the form of a vector derived from a 3×3 identity matrix. $q_{f,i}$ is the linear coefficient matrix (of size 9×3) for the ith Rodrigues rotation vector θ_i applied to the triangular face f. The collection of parameters to be trained for all triangular faces is denoted as $\mathcal{Q}_F = \{q_f, f \in F\}$.

4. Model training

The original SMPL model was trained using pre-registered mesh models. These pre-registered models were obtained using methods like SCAPE or BlendSCAPE which limits the flexibility of the SMPL model parameters. In this paper, we propose a coregistration framework to

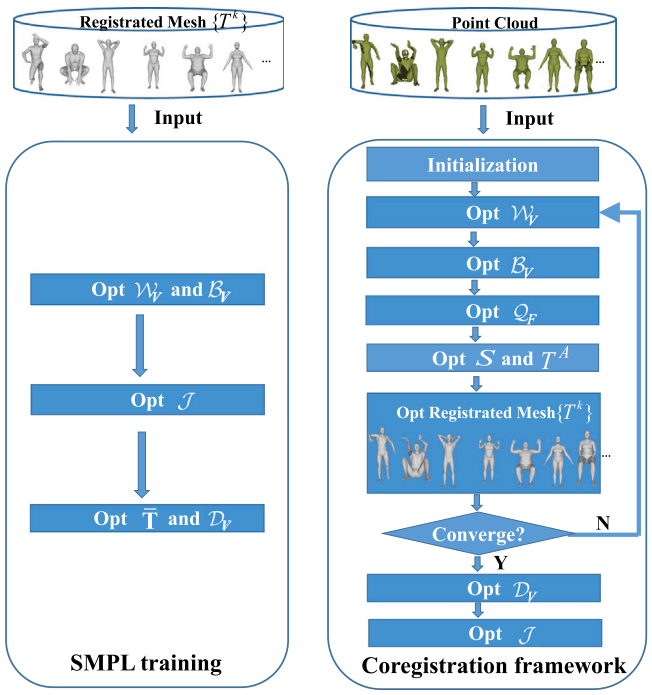

Fig. 2. Comparison of training process between SMPL method and our method: (a) SMPL training process; (b) Our coregistration framework.

accomplish biharmonic SMPL model training and mesh registration from raw point cloud data.

The coregistration process is shown in Fig. 2(b), and the SMPL model training process is shown in Fig. 2(a). For comparison, the training dataset in this paper is point cloud data, and the training results are the model parameters and the registered meshes.

4.1. Data preparation for training

The input training dataset comprises three parts: 1. The public SCAPE database, which includes 3D point cloud of Anguelov in 71 poses. 2. The CAESAR database, which includes point cloud data of 997 males and 996 females in 3 poses. 3. The FAUST database, which includes point cloud data of 5 males and 5 females in 30 poses. The training dataset is represented as $\{P^k\}$. Each point cloud data contains 73 markers.

The output is the reconstructed 3D body mesh $\{T^k\}$, the 3D body mesh in an A-pose $\{T^A\}$, and the corresponding non-rigid shape deformation $\{S\}$ and model parameter set $\Phi = \{\mathcal{W}_V, \mathcal{B}_V, \mathcal{D}_V, \mathcal{Q}_F, \mathcal{J}\}$. The kth point cloud is denoted as P^k, and the reconstructed 3D mesh is denoted as T^k.

4.2. Training of skinning weight $\mathcal{W}_V$ and automatic segmentation of rigid parts

In the previous vertex displacement based human body parametric model, the training of skinning weight $\mathcal{W}_V$ does not consider spatial continuity, which could not ensure smooth skinning deformation. Meanwhile, rigid parts are manually segmented, which could not guarantee optimality. We introduce biharmonic constraints during the training of $\mathcal{W}_V$ and automatically segment rigid parts based on the optimized $\mathcal{W}_V$.

With known 3D mesh T^k and T^A, consider the biharmonic constraints, classical LBS (Linear Blend Skinning) deformation constraints, and symmetry constraints [5], the formula for training $\mathcal{W}_V$ is:

$$\begin{aligned}\mathcal{W}_V{}^* = \arg\min_{\mathcal{W}_V}\ & \lambda_a \cdot \sum_k \sum_{v\in V} a_v \left\| \sum_{i\in N(I(v))} w_{v,i}(R_i^k p_v^{A,h(k)} + t_i^k) - p_v^k \right\|^2 \\ & + \lambda_b \cdot \left|\{T^A\}\right| \cdot \sum_{v\in V} \frac{1}{a_v} \sum_{i=0}^{K-1} \left\| \sum_{u\in N(v)} \varphi_{u,v}(w_{v,i} - w_{u,i}) \right\|^2 \end{aligned} \tag{9}$$

s.t.

$$\begin{cases} \sum_{i\in N(I(v))} w_{v,i} = 1 \\ w_{v,i} \geq 0 \\ w_{v,i} = w_{s(v),s(i)} \end{cases} \tag{10}$$

Among them, a_v represents the area of the vertex v in $\bar{T}$, $h(k)$ represents the human body index of the kth model, and $p_v^{A,h(k)}$ represents the position of the vertices of the kth mesh model under the A-pose. p_v^k represents the position of vertex v in the kth data T^k. $s(v)$ represents the symmetric vertex of vertex v, and $s(i)$ represents the symmetric part of rigid part i.

The number of training parameters is:

$$|V| \times K = 15132 \times 19 = 287508 \tag{11}$$

Without considering inequality constraint, this formulation can be directly solved. We propose a particular strategy to solve the inequality constraint problem. Firstly, solve for skinning weights that satisfy the equality constraint. Then check if the skinning weights satisfy the inequality constraint. For vertex v, if it contains skinning weights that do not satisfy inequality constraints, selects the smallest $w_{v,i}$ and sets its value to zero in the subsequent optimization process. Repeat above process until skinning weights of all vertices satisfy the inequality constraints.

Subsequently, the rigid parts are automatically segmented based on the skinning weight. For vertex v, if the rigid part that contributes the most to its skin weight is j, it is classified as the jth rigid part.

4.3. Training of the non-rigid deformation parameter $\mathcal{B}_V$

With known skinning weights $\mathcal{W}_V$ and the 3D mesh model T^k, the reconstructed mesh should satisfy the parametric SMPL model, expressed as:

$$E_1 = \sum_k \sum_{v\in V} a_v \left\| \sum_{i\in N(I(v))} w_{v,i}(R_i^k(p_v^{A,h(k)} + B_v(\theta^k; b_v)) + t_i^k) - p_v^k \right\|^2 \tag{12}$$

where θ^k represents the pose variables of the kth point cloud. Biharmonic constraints are applied to $\mathcal{B}_V$:

$$E_2 = \left|\{T^A\}\right| \cdot \sum_{v\in V} \frac{1}{a_v} \sum_{i=1}^{K-1} \left\| \sum_{u\in N(v)} \mu_{u,v}(b_{v,i} - b_{u,i}) \right\|^2 \tag{13}$$

and symmetry constraints [5] are also used:

$$E_3 = \sum_{v\in V} \sum_{i=1}^{K-1} \left\| b_{v,i} - P_{3\times3} \otimes b_{s(v),s(i)} \right\|^2 \tag{14}$$

Among them, $P_{3\times3} = \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$. $\otimes$ represents the multiplication of elements at the corresponding positions of $P_{3\times3}$ and $b_{s(v),s(i)}$. The final training of $\mathcal{B}_V$ is equivalent to solving the minimum energy function:

$$\mathcal{B}_V{}^* = \arg\min_{\mathcal{B}_V} \lambda_1 E_1 + \lambda_2 E_2 + \lambda_3 E_3 \tag{15}$$

The parameters of all vertices are coupled and require simultaneous solving, with approximately 120,000 unknowns, which can be solved globally.

4.4. Training of the non-rigid deformation parameter $\mathcal{Q}_F$ for triangular face

The SMPL-alike parametric models use the pipeline of skin deformation. During the training process, the translation $\{t\}$ of rigid parts depends on the mesh model T^A of a specific human body in the standard pose. However, the point cloud data P^A of the human body in the standard pose is often difficult to obtain accurately. Only the point cloud data of the specific human body in other poses can be obtained. The calculation of T^A also depends on $\{t\}$, and the two are coupled. In order to solve this problem, we introduce the BlendSCAPE parametric model in the training process. BlendSCAPE parametric models do not introduce $\{t\}$ and the calculation of T^A can be finished independently. So we also need to train the non-rigid deformation parameter $\mathcal{Q}_F$ of BlendSCAPE model.

With known skinning parameters $\mathcal{W}_V$, shape deformation parameters S, and the mesh model T^k, the triangular faces in the reconstructed mesh should satisfy the parametric BlendSCAPE model, expressed as:

$$E_4 = \sum_k \sum_{f\in F} a_f \sum_{i=0}^{2} \left\| R_f(\theta^k) S_f^{h(k)}(\beta) Q_f(\theta^k; q_f) \tilde{e}_{f,i} - e_{f,i}^k \right\|^2 \tag{16}$$

Among them, a_f represents the area of triangular face f in $\bar{T}$. θ^k represents the pose variable of the kth point cloud in the training set. $\tilde{e}_{f,i}$ represents the ith edge of the $\bar{T}$'s triangle f. $e_{f,i}^k$ represents the ith edge of the triangle f of the kth human body in the training set.

Meanwhile, apply L2 regularization constraint to $\mathcal{Q}_F$:

$$E_5 = \sum_{f\in F} a_f \sum_{i=1}^{K-1} \left\| q_{f,i} \right\|^2 \tag{17}$$

and apply symmetry constraints [5] to $\mathcal{Q}_F$:

$$E_6 = \sum_{f\in F} \sum_{i=1}^{K-1} \left\| q_{f,i} - P_{9\times 3} \otimes q_{s(f),s(i)} \right\|^2 \tag{18}$$

Among them $P_{9\times 3} = [1,-1,-1,-1,1,1,-1,1,1]^{\mathrm{T}} \cdot [1,-1,-1]$. $\otimes$ represents the multiplication of elements at the corresponding positions of $P_{9\times 3}$ and $q_{s(f),s(i)}$. The training of $\mathcal{Q}_F$ equates to minimizing the energy function:

$$\mathcal{Q}_F{}^* = \arg\min_{\mathcal{Q}_F} \lambda_4 E_4 + \lambda_5 E_5 + \lambda_6 E_6 \tag{19}$$

which can be independently solved for each face.

4.5. Training of body shape deformation S and T^A

When the skin deformation parameters $\mathcal{W}_V$, vertex non-rigid deformation parameters $\mathcal{B}_V$, triangular face non-rigid deformation parameters $\mathcal{Q}_F$, and 3D mesh models T^k are known, the shape deformation S must satisfy similarity constraints, expressed as:

$$E_7 = \sum_k \sum_{(f,g)\in\Theta} a_{f,g} \left\| S_f^k - S_g^k \right\|^2 \tag{20}$$

where $a_{f,g}$ is the weight of adjacent triangular faces determined by $\bar{T}$:

$$a_{f,g} = (a_f + a_g)/3 \cdot \left(\bar{d}_\Theta^2 / d(f,g)^2 \right) \tag{21}$$

Among them, a_f and a_g denotes the area of triangular face f and g respectively. Θ represents the set of all adjacent face pairs. $\bar{d}_\Theta$ represents the average geodesic distance between adjacent triangular faces, and $d(f,g)$ represents the geodesic distance between the center points of adjacent triangular face f and g.

The shape deformation S and the mesh model T^A must satisfy the following constraint:

$$E_8 = \sum_k \sum_{f\in F} a_f \sum_{\substack{i,j=0 \\ i\neq j}}^{2} \left\| S_f^k \cdot (\tilde{p}_{f,j} - \tilde{p}_{f,i}) - (p_{f,j}^{A,h(k)} - p_{f,i}^{A,h(k)}) \right\|^2 \tag{22}$$

Selecting one vertex p_{arc}^A as an anchor, the training of S and T^A is equivalent to solving the following minimal energy function:

$$S^*, T^{A*} = \arg\min_{S,T^A} \lambda_1 E_1 + \lambda_4 E_4 + \lambda_7 E_7 + \lambda_8 E_8 + \lambda_9 \left\| p_{arc} - p_{\mathrm{arc}}^A \right\|^2 \tag{23}$$

For each individual, the dimension of shape deformation S is $|F|\times 9$, the dimension of T^A is $|V|\times 3$, and the total number of unknowns is $|F|\times 9+|V|\times 3$. Considering coordinate independence, it can be transformed into three sub problems, each with an unknown variable size of $|F|\times 3+|V|$. Considering the similarity constraint, the above unknown variables need to be solved simultaneously, which is time-consuming.

4.6. Parametric model reconstruction

The parametric reconstructed 3D mesh model should fit the corresponding 3D point cloud. Considering that the template $\bar{T}$ is adaptively subdivided, with finer meshes in areas like the face, hands, and armpits, and coarser meshes in the torso and thighs, Poisson disk sampling is applied to the template $\bar{T}$. Each sampling point is represented as the weighted barycentric coordinates of the three vertices of its associated triangular face, forming the set:

$$F\left(\bar{T}\right) = \left\{ \sum_{j=0}^{2} w_{i,v_j^i} \tilde{p}_{v_j^i} \right\} \tag{24}$$

where $\left|F\left(\bar{T}\right)\right| = 50091$, $v_j^i (0 \le j \le 2)$ represents the index of the jth weighted vertex of the ith sampling point, $p_{v_j^i}$ represents the v_j^ith vertex in T. A correspondence is established between sampling points s from $F\left(T^k\right)$ and 3D points c from P^k if and only if s and c are mutually nearest points, forming the set Ω^k of all correspondences. Mathematically, the constraint for the correspondences set is defined as:

$$E_9 = \sum_k \sum_{(s,c)\in\Omega_k} \bar{a}_{F(\bar{T})} \left\| s - c \right\|^2 \tag{25}$$

where $\bar{a}_{F(\bar{T})} = \sum_{f\in F} a_f / \left|F(\bar{T})\right|$ denotes the area of each sampled point in $F\left(\bar{T}\right)$. The set of markers is M, with the constraint defined as:

$$E_{10} = \sum_k \sum_{i\in M} \bar{a}_{F(\bar{T})} \left\| p_{v(i)}^k - m_i^k \right\|^2 \tag{26}$$

where $v(i)$ is the vertex index on template for the ith marker, and m_i^k is the ith marker position of the kth data. Simultaneously, the reconstructed 3D human mesh model should comply with the parametric deformation equation. Consequently, the final reconstruction problem is equivalent to solving the following minimal energy function:

$$\arg\min_{T^k} E(T^k) = \lambda_1 E_1 + \lambda_4 E_4 + \lambda_9 E_9 + \lambda_{10} E_{10} \tag{27}$$

We can registrate each point cloud data separately. Taking into account coordinate independence, we need to solve three sub optimization problems for each point cloud data, and the number of unknowns for each sub problem is $|V|$.

4.7. Training of body shape deformation parameters $\mathcal{D}_V$ and joint deformation parameters $\mathcal{J}$

After obtaining the mesh model T^A for each individual in the A-pose, the parameters D_V and the shape variables β of each individual are determined through principal component analysis on vertex coordinates. Furthermore, for the mesh model of each individual in different poses, joint positions are calculated. The joint deformation parameters $\mathcal{J}$ are then trained using the joint positions and the shape variables β of each individual.

4.8. The implementation of the proposed algorithm

During the entire training process, the steps in Section 4.2 to Section 4.6 are interleaved iteratively, completing the optimization of each variable until convergence. The steps of the algorithm are listed in Algorithm 1 (see [39–41]).

Algorithm 1 Coregistration framework of parametric 3D human modeling with biharmonic SMPL

Initialization:
1.1 Initialize the vertex skinning weights $\mathcal{W}_V$ using the method by Baran et al.[39], and divide the template $\bar{T}$ into 19 rigid parts based on the skinning weights.
1.2 Initialize the matrix S_f^k as a 3×3 identity matrix, and set $b_{v,i}(0 < i < K)$ to zero for all elements.
1.3 For $\mathcal{Q}_F$, $q_{f,0}$ is initialized as a vector form of a third-order identity matrix, and $q_{f,i}(0 < i < K)$ is initialized as zero vector.
1.4 T^k was registered through $\bar{T}$ using Li et al.[40] method. Initialize T^A to $\bar{T}$.
for iteration: 0 to 50 **do**
  Fix other variables and optimize $\mathcal{W}_V$:
  3.1 Calculate the rotation transformation $\{R_i^k, 0 \le i < K\}$ relative to $\bar{T}$ for each rigid part using the method by Horn et al. [41] .
  3.2 Replace the rigid parts' original vertices with Poisson-sampled points.
  3.3 Calculate the translation variables $\{t_i^k, 0 \le i < K\}$ of T^k relative to $T^{A,h(k)}$ in the database, and then optimize $\mathcal{W}_V$.
  3.4 Update the rigid parts' segmentation, assigning each vertex and face to the rigid part with the largest skinning weight.
  Fix the remaining variables, calculate the pose variable $\{\theta^k\}$ based on $\{R^k\}$ in Step 2, and optimize $\mathcal{B}_V$.
  Optimize $\mathcal{Q}_F$ while fixing other variables, based on pose variables $\{\theta^k\}$.
  Optimize $\{S\}$ and $\{T^A\}$ while fixing other variables.
  Optimize $\{T^k\}$ while fixing other variables.
end for
Optimize $\mathcal{D}_V$ using PCA based on the mesh model $\{T^A\}$.
Calculate joint positions for each individual and optimize $\mathcal{J}$.

Due to the large number of data in the training set, directly applying the above steps to all data is time-consuming. Therefore, to reduce training time, the training steps are optimized. Initially, sample the training set data, selecting 3D point cloud of Anguelov in 71 poses, point cloud data of 40 males and 40 females in 3 poses for CAESAR database and point cloud data of 2 males and 2 females in 30 poses for FAUST database. Then train to obtain pose parameters $\mathcal{W}_V$, $\mathcal{B}_V$ and $\mathcal{Q}_F$. Subsequently, fix $\mathcal{W}_V$, $\mathcal{B}_V$, $\mathcal{Q}_F$ and optimize $\{T^k\}$, $\{T^A\}$ and $\{S\}$. Finally, train the remaining parameter $\mathcal{D}_V$ and $\mathcal{J}$.

During training, the number of iterations is set to 50, with weight settings as follows:

- The weight λ_a is 0.25
- The weight λ_b is 0.5
- The weight λ_1 of E_1 item is 0.25
- The weight λ_2 of E_2 item is 1.5
- The weight λ_3 of E_3 item is 100
- The weight λ_4 of item E_4 is 1.0, which is similar to the meaning of λ_D in BlendSCAPE
- The weight λ_5 of item E_5 is $2.0e-5$, which has a similar meaning to λ_Q in BlendSCAPE
- The weight λ_6 of E_6 item is 100
- The weight λ_7 of E_7 item is $1.0e-4 \cdot |\{T^k\}|$
- The weight λ_8 of E_8 item is 100
- The weight λ_9 of E_9 item is 1

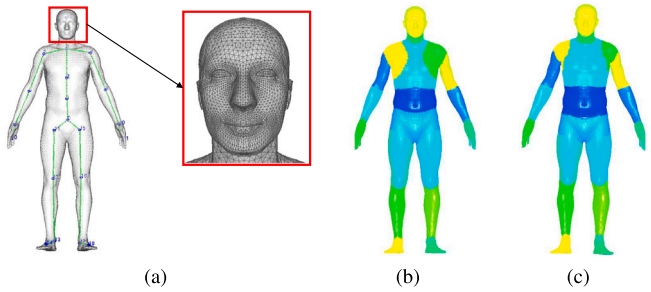

Fig. 3. Segmentation results for rigid parts: (a) Pose template and skeleton; (b) Initial segmentation; (c) Automatic segmentation after training.

- The weight λ_{10} of E_{10} item is 100

After 25 iterations, λ_{10} is set to zero to eliminate local distortions caused by markers.

5. Experimental results and analysis

5.1. Optimization of parameter $\mathcal{W}_V$

As shown in Fig. 3, rigid parts automatically segmented through optimized skinning weights appear more natural, especially at joints such as the neck and armpits.

Fig. 4 shows that the automatic segmentation method proposed in this article has good consistency in the optimized segmentation results for different initial values. In the experiment, we calculate five different skin weights based on five different skeletons, and obtain five different initial segmentation results. For vertex v on the template, the segmentation difference value of v is defined as the proportion of pairs with different segmentation values to all pairs, belonging to the range of 0 to 1. (a) and (b) demonstrates that the automatic segmentation method proposed in this paper can achieve almost consistent segmentation results after optimization for the five different initial segmentation methods mentioned above. For any two segmentations, the segmentation difference ratio is defined as the proportion of vertices with different partition values to all vertices, belonging to the range of 0 to 1. The segmentation difference ratios between two different segmentation tasks before and after optimization are displayed in (c), indicating that after optimization, certain different initial values are able to obtain consistent segmentation results. The above experimental results indicate that the automatic segmentation method proposed in this paper avoids the influence of different initial values generated by manual segmentation on the final optimization results.

Fig. 5 compares the results with and without biharmonic constraints to skinning weights. The former achieves a more uniform distribution of skinning weights. Fig. 6 compares the area of regions with fitting errors exceeding 9 mm with and without the optimization of skinning weights. As seen in Fig. 6, the area with errors exceeding 5 mm decreased by 17.6% with optimization.

Furthermore, Table 1 compares the area of regions within the 0–1 cm error range with and without the optimization of skinning weights. It is observed that, the area of regions with errors exceeding 4 mm was consistently larger without optimization. Conversely, the area of regions with errors less than 4 mm is consistently smaller without optimization, indicating that the optimization of skinning weights significantly reduces larger fitting errors.

5.2. Optimization of non-rigid deformation parameter $\mathcal{B}_V$

In this subsection, we compare the reconstruction results of parameter $\mathcal{B}_V$ when subjected to regularization constraints [3] and biharmonic constraints. When considering only regularization constraints, larger

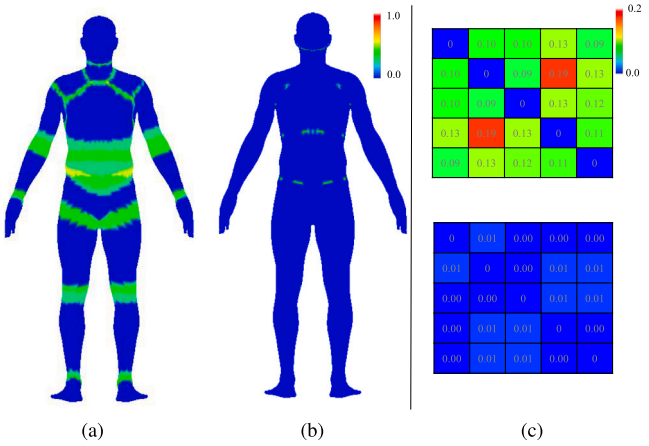

Fig. 4. Consistence of automatic segmentation for different initialization. (a) The average segmentation difference between five different initial segmentations; (b) The average segmentation difference between above five optimized segmentations; (c) The segmentation difference ratio of different segmentations between each other before and after optimization.

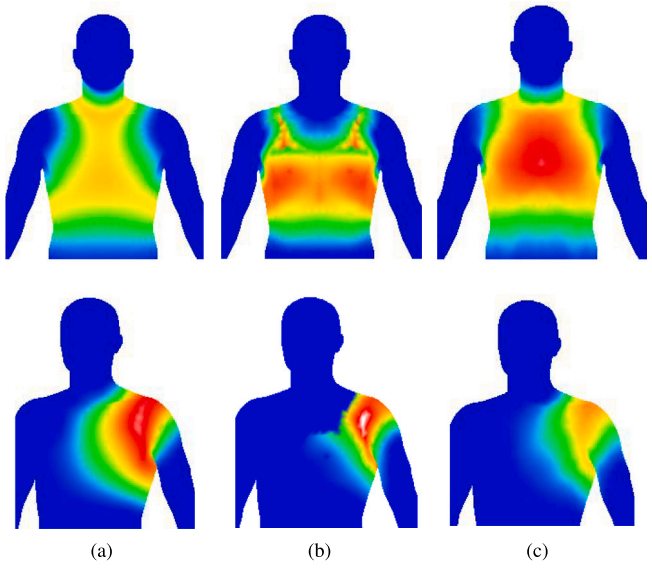

Fig. 5. Results of skinning weights: (a) Initial distribution of skinning weights; (b) Weight distribution without biharmonic constrains; (c) Weight distribution with biharmonic constraints.

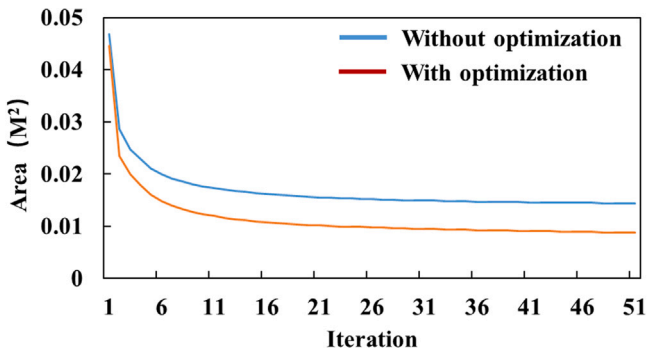

Fig. 6. Area of regions with fitting errors exceeding 9 mm with and without skinning weight optimization.

Table 1
Comparison of area within different error ranges without and with skinning weight optimization (m^2).

	Unoptimized	Optimized	Unoptimized-optimized
0–1 mm	8.88e−2	9.13e−2	−2.57e−3
1–2 mm	5.02e−1	5.15e−1	−1.31e−2
2–3 mm	5.24e−1	5.36e−1	−1.19e−2
3–4 mm	2.44e−1	2.45e−1	−7.01e−4
4–5 mm	1.07e−1	1.03e−1	4.75e−3
5–6 mm	5.55e−2	5.00e−2	5.58e−3
6–7 mm	3.04e−2	2.58e−2	4.56e−3
7–8 mm	1.71e−2	1.35e−2	3.66e−3
8–9 mm	9.83e−3	6.96e−3	2.87e−3
9–10 mm	1.44e−2	8.74e−3	5.63e−3

Table 2
Comparison of local smoothness of the registered mesh model in the training dataset before and after optimization of $\mathcal{B}_V$, under conditions without and with skinning weight $\mathcal{W}_V$ optimization (m^2).

$\mathcal{B}_V$ / $\mathcal{W}_V$	Unoptimized / Unoptimized	Optimized / Unoptimized	Unoptimized / Optimized	Optimized / Optimized
0–1 mm	1.99	1.97	2.02	2.02
1–2 mm	4.90e−2	8.16e−2	3.83e−2	4.36e−2
2–3 mm	1.60e−2	1.27e−2	7.06e−3	3.31e−3
3–4 mm	7.23e−3	3.38e−3	1.61e−3	4.68e−4
4–5 mm	3.68e−3	1.25e−3	4.76e−4	7.57e−5
5–6 mm	1.88e−3	5.74e−4	1.34e−4	6.68e−6
6–7 mm	1.05e−3	3.00e−4	3.76e−5	3.51e−6
7–8 mm	6.06e−4	1.54e−4	1.88e−6	0
8–9 mm	3.56e−4	7.80e−5	4.58e−7	5.85e−7
9–10 mm	4.86e−4	9.60e−5	4.58e−7	0

constraints result in better smoothness of the registered mesh model but also lead to increased fitting errors. Conversely, smaller constraints yield smaller fitting errors but at the cost of reduced smoothness. As illustrated in Fig. 7, the introduction of biharmonic constraints in addition to regularization constraints ensures small fitting errors while maintaining the smoothness of the registered mesh model. Fig. 8 demonstrates that with lower regularization constraint weights and the introduction of biharmonic constraints, the fitting errors of the registered mesh model are significantly reduced.

With weak regularization constraints, where $\lambda_Q = 0.01$, the introduction of biharmonic constraints ensures both minimal average fitting errors and improved smoothness, as evidenced in Figs. 9 and 10. As indicated in Table 2, a quantitative comparison of the local smoothness of the registered mesh models is conducted, both before and after optimization of parameter $\mathcal{B}_V$, and under conditions with and without optimized skinning weights. The table suggests that there is a significant improvement in local smoothness with optimization under both scenarios. Local smoothness is defined by the absolute displacement of vertices before and after applying Laplacian smoothing (using the Cotangent Laplacian operator in $\bar{T}$), followed by the computation of the average area distribution within the 0–1 cm range. A larger area in the high-displacement interval indicates poorer smoothness. After optimization, the area with smoothness errors exceeding 5 mm decreases by 83.2%.

5.3. Optimization of triangular parametric deformation

In the training of the human body's shape deformation S and T^A, we introduce a triangular based parametric model such as BlendSCAPE. This effectively reduces the deformation space for the shape deformation. We compare the body shape reconstruction results of A-pose with and without the triangular parametric model. As illustrated in Fig. 11, it is clearly observable that the reconstruction results under A-pose are more plausible when the triangular parametric model is incorporated.

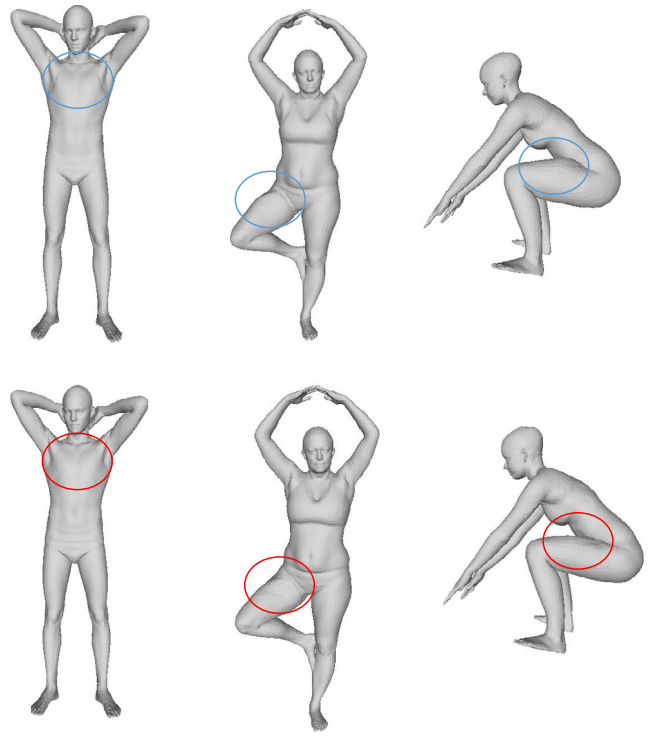

Fig. 7. The first row shows the reconstruction results with a regularization weight of 5.0, from which it is obvious that despite the larger weight setting, local distortions are still present. The second row, with a regularization weight of 0.05 and the introduction of biharmonic constraints, significantly enhances the local smoothness of the registered mesh and improves the fidelity of local deformations.

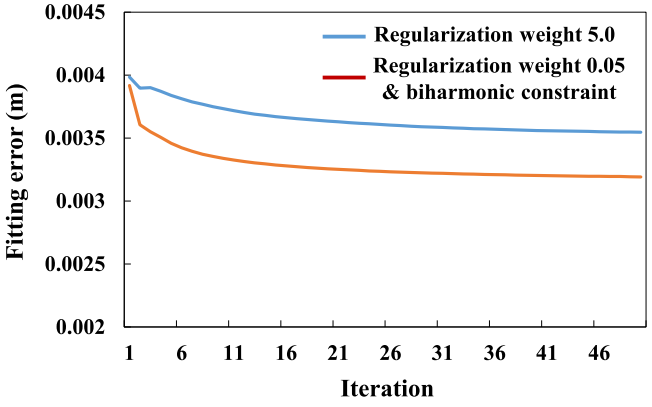

Fig. 8. Fitting error of the registered mesh model when only setting a larger regularization weight, compared to setting a smaller regularization weight and considering biharmonic constraints.

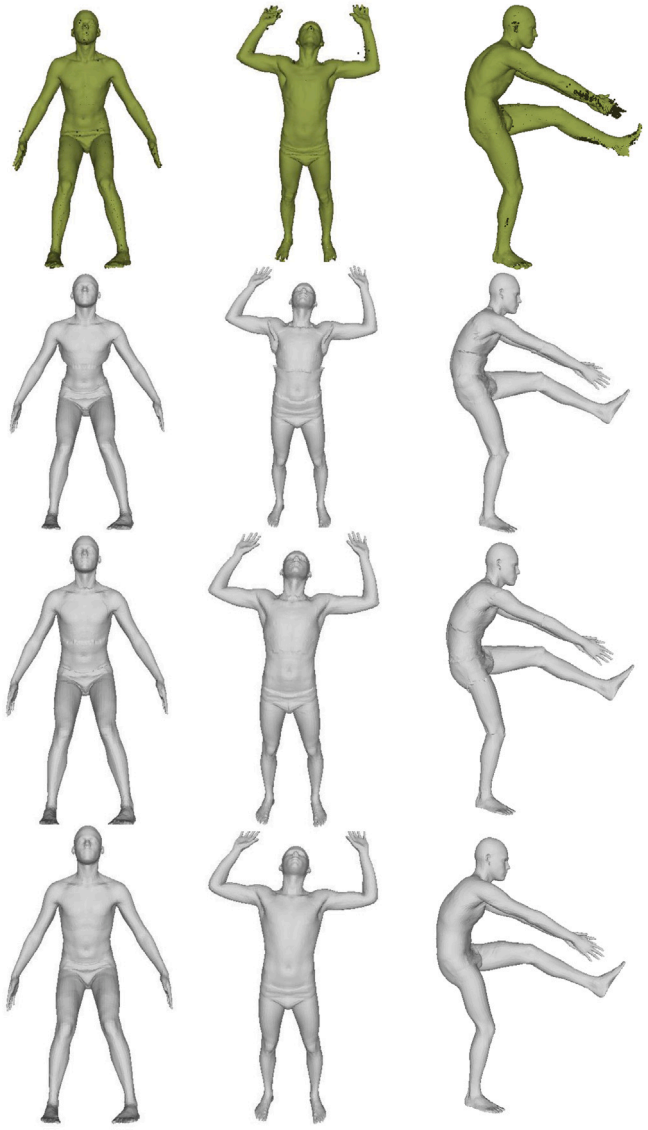

Fig. 9. Comparison of male model reconstruction results using different methods. The first row presents the original scan point cloud of a male in three poses. The second row shows the registration result without optimization. The third row illustrates the registration results after optimizing skinning weights. The fourth row shows the results achieved by optimizing both skinning weights and non-rigid deformation parameters.

5.4. Comparison of reconstruction results with other parametric models

The comparative analysis of reconstruction results between the proposed model and the BlendSCAPE [3] model is illustrated in Fig. 12. It can be noted that the mesh's smoothness generated by our model is comparable to that of the BlendSCAPE model. Furthermore, Fig. 13 presents the comparison of reconstruction results between our model and the SMPL [4] model . It is discernible that the mesh reconstructed by the our model demonstrates superior smoothness. We also compare the fitting error of our model with SMPL and STAR models on a dataset consisting of SCAPE, CAESAR and FAUST. This dataset contains a total of 6350 scanned data. We calculate the average point-to-point distance between the generated parameterized mesh and the registrated mesh obtained during coregistration as the fitting error. In Fig. 14, we calculate the fitting errors of different shape coefficients on the dataset. It is evident that the model proposed in this paper has smaller fitting errors, demonstrating stronger representational ability.

We also evaluate our model against human body models SMPL-X [8] and SUPR-Body [21]. We sample 100 male and 100 female subjects from the CAESAR dataset. Each subject contains three poses. We register the template to all the scans with SMPL-X, SUPR and our model respectively by minimizing the vertex-to-vertex loss between the model surface and the corresponding registration. We report the mean absolute error (mabs) exclude the face, the feet and the hands for up to 20 shape components in Fig. 16 and show a qualitative sample of the model fits in Fig. 15. Our model shows a lower error than SMPL-X and SUPR.

Finaly, we compare our model with SKEL [22] model on the DFAUST [16] dataset and their SMPL fits with 10 shape parameters. Similar to SKEL, we fit template to each of the SMPL meshes by optimizing pose and shape parameters. We compute the mean and the

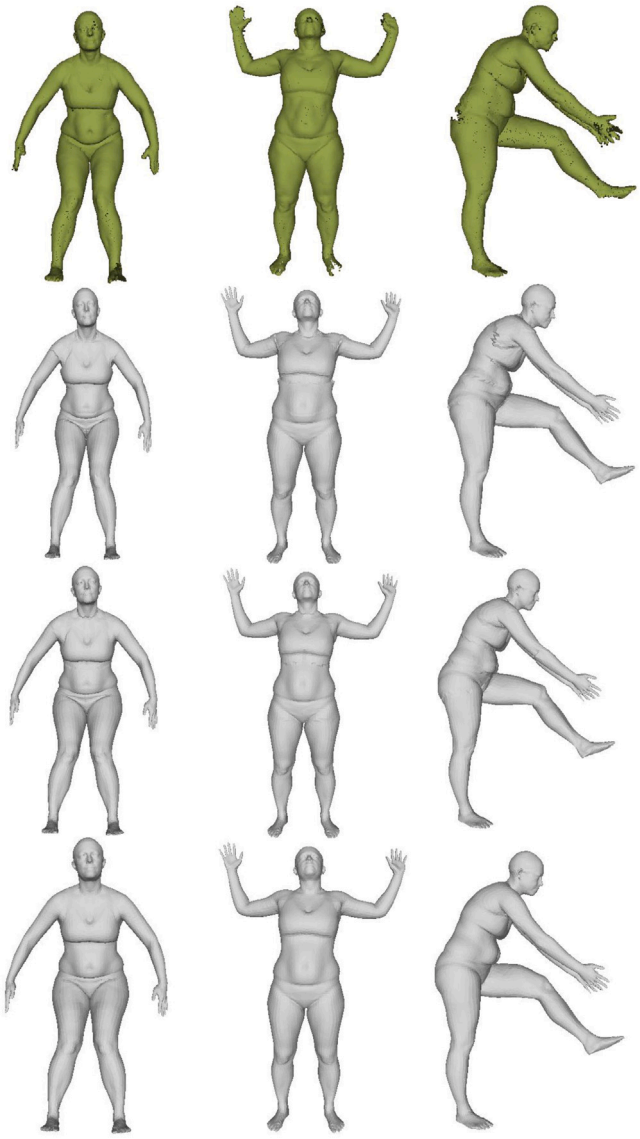

Fig. 10. Comparison of female model reconstruction results using different methods. The first row presents the original scan point cloud of a female in three poses. The second row shows the registration result without optimization. The third row illustrates the registration results after optimizing skinning weights. The fourth row shows the results achieved by optimizing both skinning weights and non-rigid deformation parameters.

Table 3
Comparison of fitting errors with SKEL [22] on DFAUST dataset [16] (cm).

Metric Sex	Ave-mean Male	Ave-mean Female	Ave-max Male	Ave-max Female
SKEL	1.1	2.5	0.9	1.9
Our model	0.60	2.61	0.51	2.23

max absolute error between the vertices of fitted template and the target SMPL vertices over all the frames. Then calculate the weighted average of all the above values, with the weight being the area of the vertices. The quantitative comparison results are shown in Table 3. A visualization of these errors on the fitted template is shown in Fig. 17. It can be seen that the model proposed in this article is comparable or superior to the SKEL model in terms of both average mean(ave-mean) and max(ave-max) absolute error metrics.

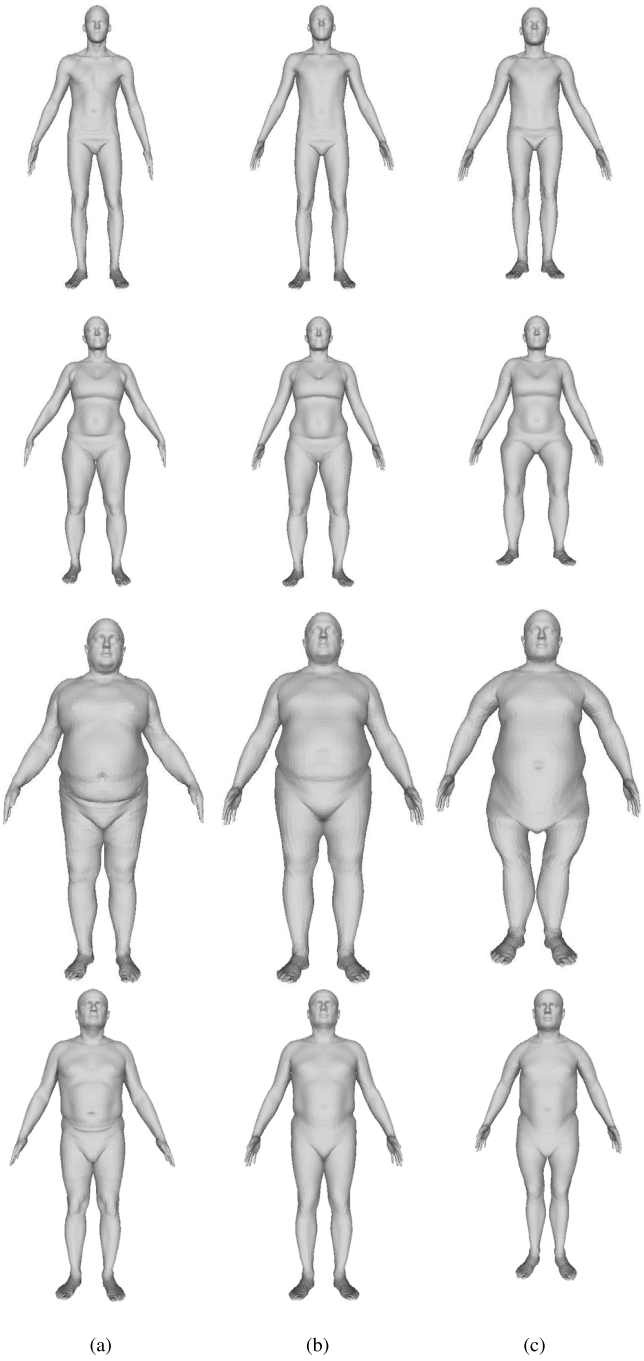

Fig. 11. Body shape reconstruction results of A-pose with and without introducing the triangular parametric model for four different individuals. Each row represents a single individual. (a) Approximated A-pose from original scan data; (b) Standard A-pose model reconstructed by coregistration with triangular parametric model; (c) Standard A-pose model reconstructed by coregistration without triangular parametric model.

5.5. Experimental setup and performance analysis

The experiments were conducted on a Lenovo Kaitian M420 desktop with a 12-core CPU, 3.2 GHz. Table 4 shows the average execution times for the each training sub-steps of our model, including the training of skinning weights $\mathcal{W}_V$, non-rigid deformation parameters

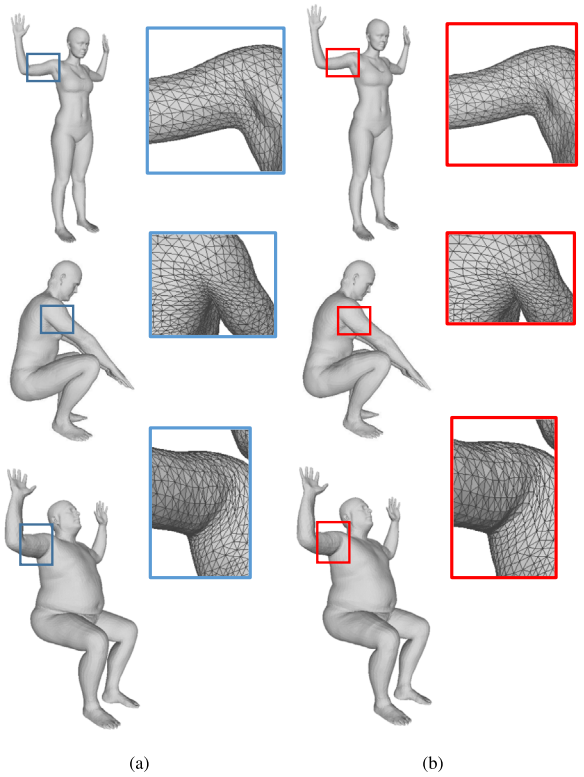

Fig. 12. The reconstruction results with (a) BlendSCAPE [3] model and (b) ours. It can be noted that the mesh's smoothness generated by our model is comparable to that of the BlendSCAPE model.

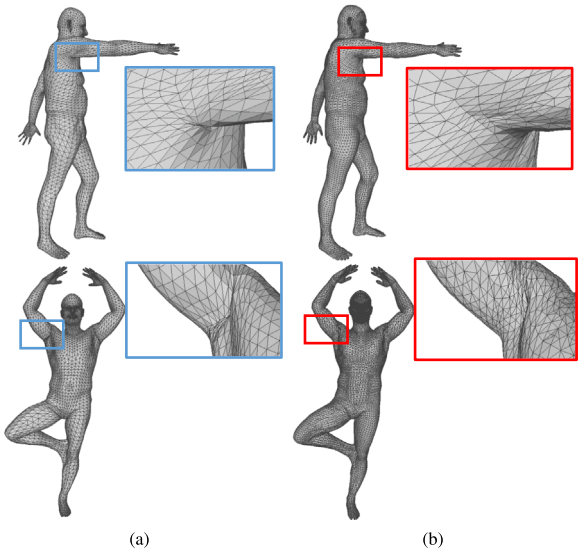

Fig. 13. The reconstruction results with (a) SMPL [4] model and (b) ours. It is discernible that the mesh reconstructed by our model demonstrates superior smoothness in the articular area.

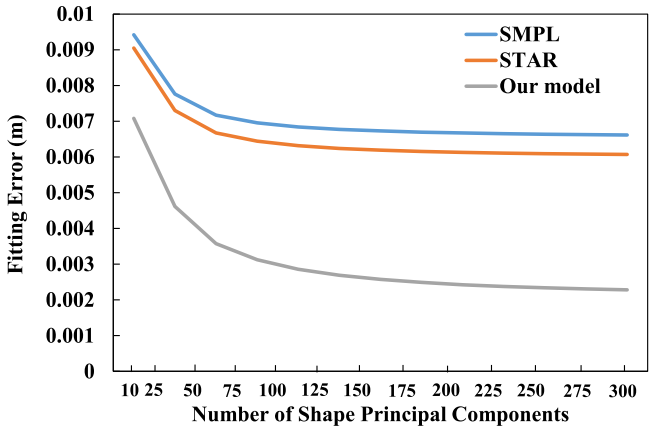

Fig. 14. Comparison of the fitting errors between SMPL [4], STAR [20], and the method proposed in this paper on datasets consisting of SCAPE, CAESAR and FAUST.

Table 4
Each sub-step time consuming in training process.

Sub-step	$\mathcal{W}_V$	$\mathcal{Q}_F$	$\mathcal{B}_V$	$\{S\}$ $\{T^A\}$	$\{T^k\}$
Time (s)	23.22	42.14	13.57	129.54	130.74

for triangles $\mathcal{Q}_F$, non-rigid deformation parameters for vertices $\mathcal{B}_V$, body shape deformation parameters $\{S\}$ and shape in A-pose $\{T^A\}$, and the registration of model $\{T^k\}$. None of the sub-steps employed GPU acceleration. It is observed that the training of the body shape deformation parameters and the registration of model $\{T^k\}$ are the most time-consuming steps.

The former is time-consuming because of two reasons. One is that for each individual, it is necessary to optimize both the shape deformation S and the shape under A-pose T^A simultaneously, with an unknown number of $|F|\times 3+|V|$, as detailed at the end of Section 4.5. The other is that the training dataset has multiple individuals, and each individual needs to be optimized separately. The reason for the latter being time-consuming is similar. One reason is that for each point cloud data, it is necessary to optimize all vertex positions simultaneously, with an unknown number of $|V|$, as detailed in Section 4.6. The other reason is that each point cloud data in the training dataset needs to be optimized separately. Performance of our implementation, and a comparison against BlendSCAPE, is shown in Fig. 18.

5.6. More results

Fig. 19 displays the mesh models created with new body shapes and poses using our parametric model. It is obvious that the mesh models, generated with the optimized skinning weights and non-rigid deformation parameters, are more realistic and smooth. Fig. 20 shows the generation results of more different individuals in different poses using our model.

6. Conclusion and future work

Existing 3D human parametric models possess robust expressive capabilities, capable of generating realistic 3D human models with simple pose and shape parameters. However, they do not consider the local similarity of parameters within the model. This paper addresses this problem by proposing a biharmonic 3D human parametric model. Experimental results indicate that by introducing biharmonic constraints to the training process, the area with fitting errors exceeding 5 mm decreases by 17.6% after skinning weight optimization. And

Fig. 15. We evaluate our method against baselines. We sample 100 males and 100 females from the CAESAR dataset to evaluate SMPL-X [8], SUPR [21] and our BiharmonicSMPL using 20 shape components. Blue: 0 cm, Red 1 cm.

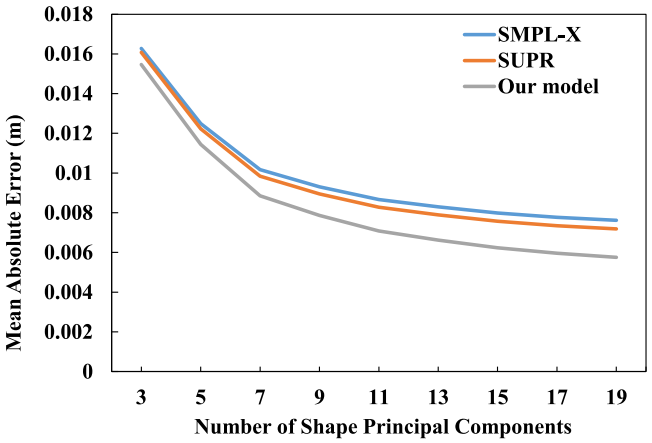

Fig. 16. Evaluating the generalization of our method against SMPL-X [8] and SUPR [21]. We report the mean absolute vertex-to-vertex error (m) as a function of the number of the shape coefficients.

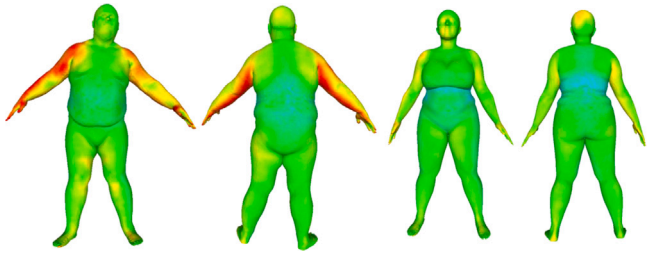

Fig. 17. Mean absolute vertex distance between our model and SMPL fit to the males and females of the DFAUST [16] dataset. Blue: 0 cm, Red 1 cm.

the area with smoothing errors exceeding 5 mm decreases by 83.2% after optimization of non-rigid deformation parameters. Additionally,

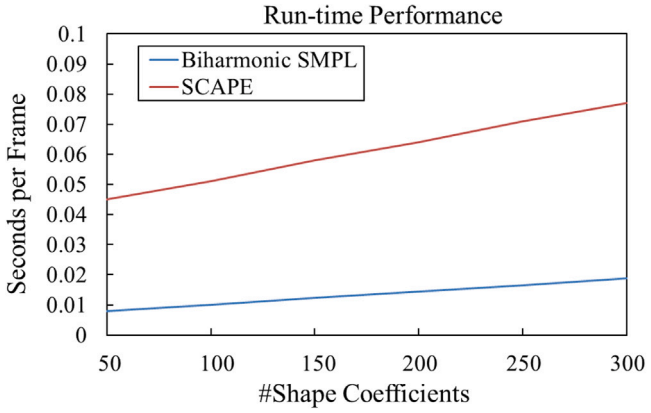

Fig. 18. Performance of SMPL and BlendSCAPE vary with the number of body shape coefficients used. Performance shown here is from a Lenovo Kaitian M420 desktop with a 12-core CPU, 3.2 GHz.

this paper presents a coregistration framework that transforms the process of model training and mesh registration into solving a series of optimization problems.

Our method also has some limitations. Firstly, while the biharmonic constraints enhance the model's performance, they also increase the training time. This is because the vertex displacement-based human parametric model simultaneously considers the triangular-based parametric model during the training process, necessitating the solution of complex multivariable optimization problems. Future work could consider training only the vertex displacement-based human parametric model. Secondly, the non-rigid vertex displacement deformations in this paper essentially apply the same deformation model to all human bodies, without accounting for individual differences that lead to person-specific non-rigid displacements. In fact, different individuals may have different non-rigid displacements even under the same pose. To solve this problem, it is necessary to consider establishing separate non-rigid displacement models for different individuals. One feasible approach is to characterize each individual's pose guided non-rigid

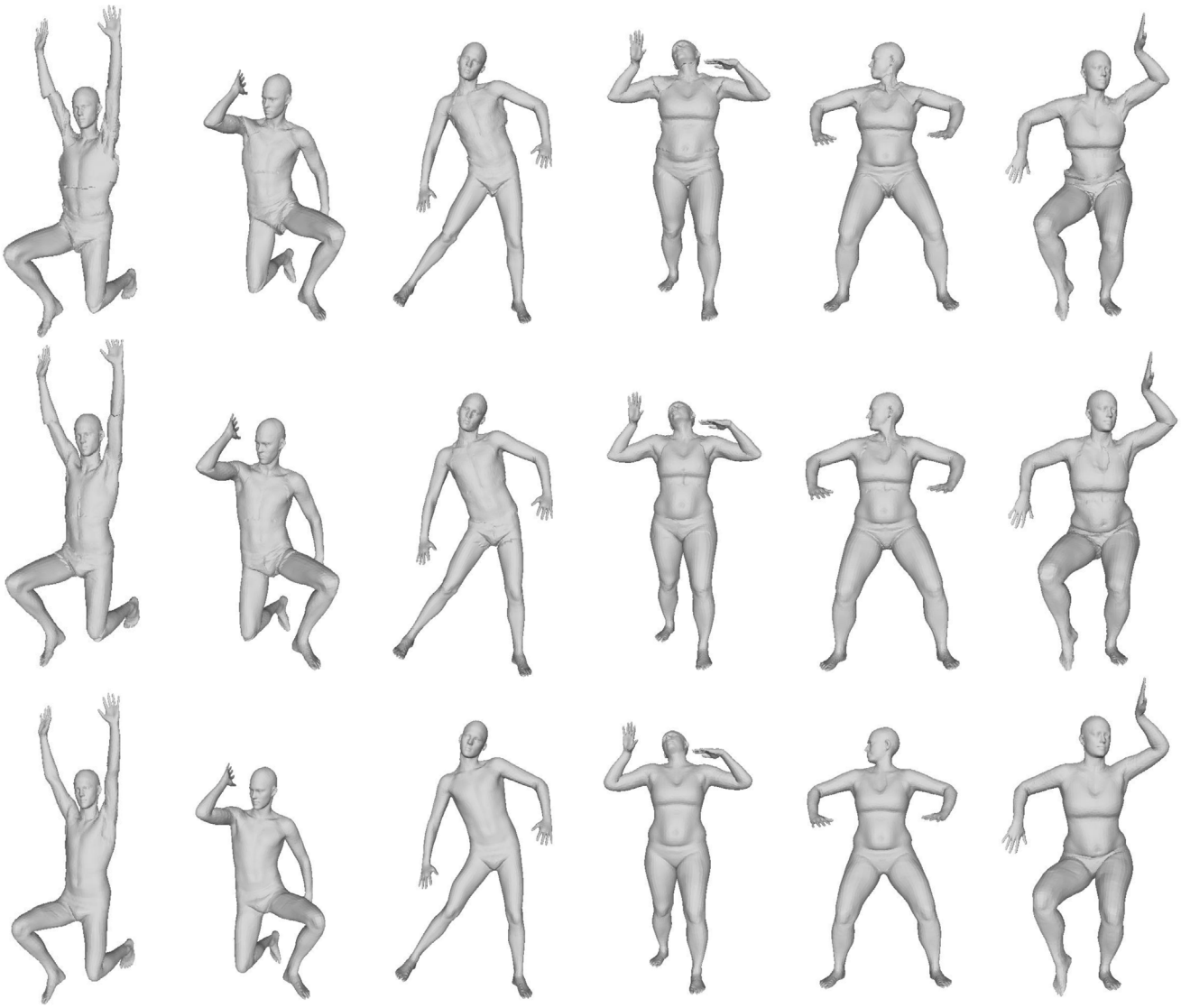

Fig. 19. The first row shows the mesh generated by the SMPL parametric model without any optimization. The second row shows the mesh generated by the parametric model with skinning weight optimization. The third row shows the mesh generated by the parametric model with optimization of both skinning weights and non-rigid deformation parameters. It is evident that the mesh quality in the third line is the best.

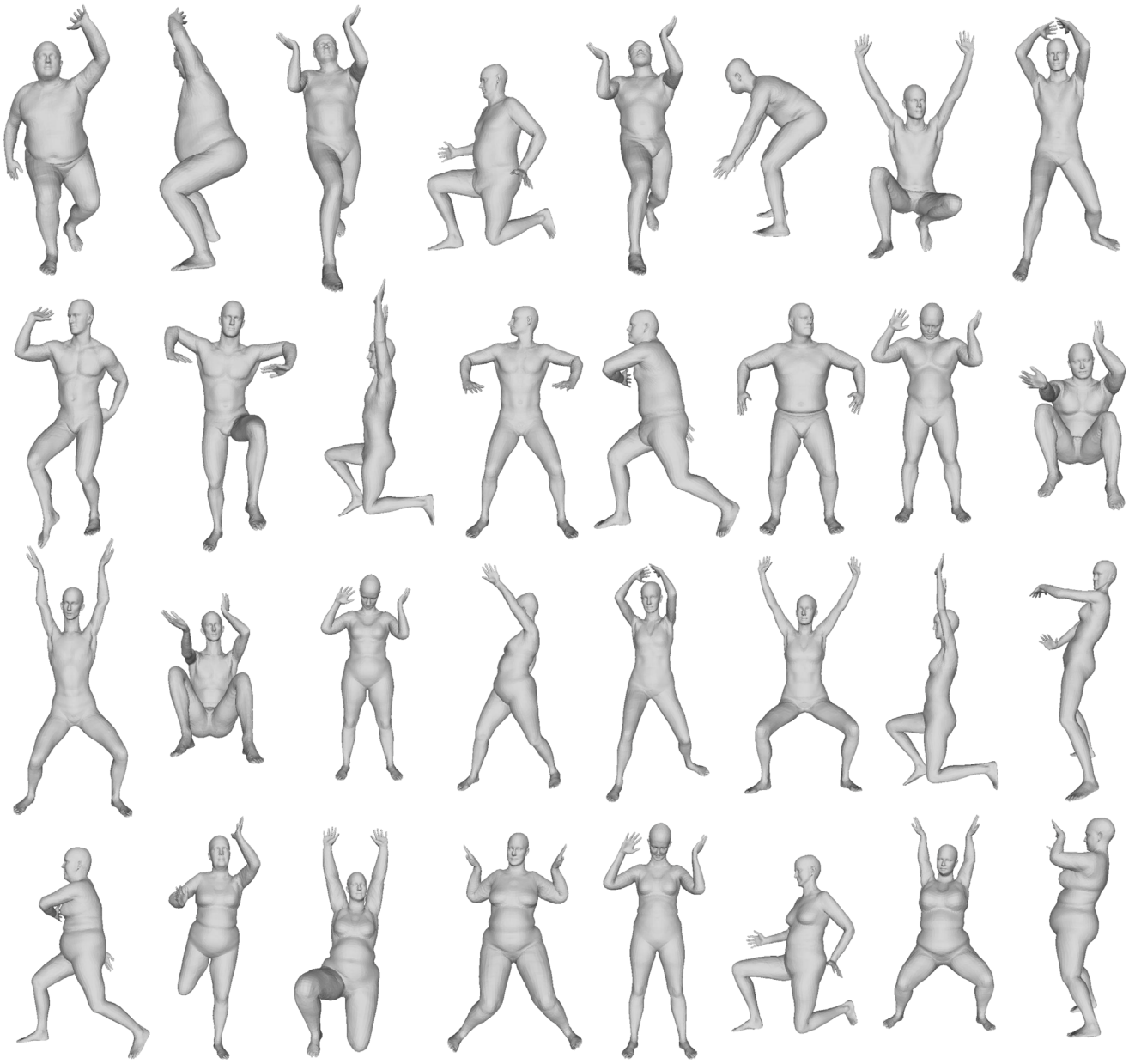

Fig. 20. More results of different individuals in different poses using our Biharmonic SMPL model.

displacement model as a function that varies with body shape. Another potential research direction is to use the proposed model as prior knowledge for human reconstruction based on depth or color images, aiming to achieve more accurate 3D human models.

CRediT authorship contribution statement

Yin Chen: Project administration, Methodology. **Yuping Ye:** Writing – original draft. **Weiwei Xu:** Writing – review & editing. **Qiliang Yang:** Supervision, Conceptualization. **Qizhen Zhou:** Visualization.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Acknowledgments

This work is supported by National Key Research and Development Program of China (2023YFC3107100) and Natural Science Foundation of Jiangsu Province, China (BK20211226).

Data availability

The authors do not have permission to share data.

References

[1] Cheng Z-Q, Chen Y, Martin RR, Wu T, Song Z. Parametric modeling of 3D human body shape—A survey. Comput Graph 2018.

[2] Anguelov D, Srinivasan P, Koller D, Thrun S, Rodgers J, Davis J. Scape: shape completion and animation of people. In: ACM SIGGRAPH 2005 papers. 2005, p. 408–16.

[3] Hirshberg DA, Loper M, Rachlin E, Black MJ. Coregistration: Simultaneous alignment and modeling of articulated 3D shape. In: Computer vision–ECCV 2012: 12th European conference on computer vision, florence, Italy, October 7-13, 2012, proceedings, part VI 12. Springer; 2012, p. 242–55.

[4] Loper M, Mahmood N, Romero J, Pons-Moll G, Black MJ. SMPL: A skinned multi-person linear model. In: Seminal graphics papers: pushing the boundaries. vol. 2, 2023, p. 851–66.

[5] Chen Y, Song Z, Xu W, Martin RR, Cheng Z-Q. Parametric 3D modeling of a symmetric human body. Comput Graph 2019.

[6] Bogo F, Black MJ, Loper M, Romero J. Detailed full-body reconstructions of moving people from monocular RGB-d sequences. In: Proceedings of the IEEE international conference on computer vision. 2015, p. 2300–8.

[7] Yu T, Zheng Z, Guo K, Zhao J, Dai Q, Li H, Pons-Moll G, Liu Y. Doublefusion: Real-time capture of human performances with inner body shapes from a single depth sensor. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 7287–96.

[8] Pavlakos G, Choutas V, Ghorbani N, Bolkart T, Osman AA, Tzionas D, Black MJ. Expressive body capture: 3d hands, face, and body from a single image. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019, p. 10975–85.

[9] Guan P, Weiss A, Balan AO, Black MJ. Estimating human shape and pose from a single image. In: 2009 IEEE 12th international conference on computer vision. IEEE; 2009, p. 1381–8.

[10] Jacobson A, Baran I, Popovic J, Sorkine O. Bounded biharmonic weights for real-time deformation. ACM Trans Graph 2011;30(4):78.

[11] 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.

[12] Chen Y, Liu Z, Zhang Z. Tensor-based human body modeling. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2013, p. 105–12.

[13] Chen Y, Cheng Z-Q, Lai C, Martin RR, Dang G. Realtime reconstruction of an animating human body from a single depth camera. IEEE Trans Vis Comput Graphics 2015;22(8):2000–11.

[14] Bogo F, Romero J, Loper M, Black MJ. FAUST: Dataset and evaluation for 3D mesh registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, p. 3794–801.

[15] Tsoli A, Mahmood N, Black MJ. Breathing life into shape: Capturing, modeling and animating 3D human breathing. ACM Trans Graph 2014;33(4):1–11.

[16] Bogo F, Romero J, Pons-Moll G, Black MJ. Dynamic FAUST: Registering human bodies in motion. In: Computer vision & pattern recognition. 2017.

[17] Li T, Bolkart T, Black MJ, Li H, Romero J. Learning a model of facial shape and expression from 4D scans. ACM Trans Graph 2017;36(6):1–17.

[18] Romero J, Tzionas D, Black MJ. Embodied hands: modeling and capturing hands and bodies together. ACM Trans Graph 2017;36(6). http://dx.doi.org/10.1145/3130800.3130883.

[19] Joo H, Simon T, Sheikh Y. Total capture: A 3d deformation model for tracking faces, hands, and bodies. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 8320–9.

[20] Osman AAA, Bolkart T, Black MJ. STAR: Sparse trained articulated human body regressor. 2020.

[21] Osman AAA, Bolkart T, Tzionas D, Black MJ. SUPR: A sparse unified part-based human representation. In: European conference on computer vision. 2022.

[22] Keller M, Werling K, Shin S, Delp S, Pujades S, Liu CK, Black MJ. From skin to skeleton: Towards biomechanically accurate 3D digital humans. ACM- Trans Graph 2023;42(6):12.

[23] Bhatnagar BL, Sminchisescu C, Theobalt C, Pons-Moll G. Loopreg: Self-supervised learning of implicit surface correspondences, pose and shape for 3d human mesh registration. Adv Neural Inf Process Syst 2020;33:12909–22.

[24] Deng B, Lewis JP, Jeruzalski T, Pons-Moll G, Hinton G, Norouzi M, Tagliasacchi A. Nasa neural articulated shape approximation. In: Computer vision–ECCV 2020: 16th European conference, glasgow, UK, August 23–28, 2020, proceedings, part VII 16. Springer; 2020, p. 612–28.

[25] Chen X, Zheng Y, Black MJ, Hilliges O, Geiger A. Snarf: Differentiable forward skinning for animating non-rigid neural implicit shapes. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021, p. 11594–604.

[26] Chibane J, Alldieck T, Pons-Moll G. Implicit functions in feature space for 3D shape reconstruction and completion. In: 2020 IEEE/CVF conference on computer vision and pattern recognition. CVPR, 2020.

[27] Shen K, Guo C, Kaufmann M, Zarate JJ, Valentin J, Song J, Hilliges O. X-Avatar: Expressive human avatars. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. CVPR, 2023, p. 16911–21.

[28] Tian Y, Zhang H, Liu Y, Wang L. Recovering 3d human mesh from monocular images: A survey. IEEE Trans Pattern Anal Mach Intell 2023.

[29] Weiss A, Hirshberg D, Black MJ. Home 3D body scans from noisy image and range data. In: 2011 international conference on computer vision. IEEE; 2011, p. 1951–8.

[30] Zheng J, Zeng M, Cheng X, Liu X. SCAPE-based human performance reconstruction. Comput Graph 2014.

[31] Bălan AO, Black MJ. The naked truth: Estimating body shape under clothing. In: Computer vision–ECCV 2008: 10th European conference on computer vision, marseille, France, October 12-18, 2008, proceedings, part II 10. Springer; 2008, p. 15–29.

[32] Zhang C, Pujades S, Black MJ, Pons-Moll G. Detailed, accurate, human shape estimation from clothed 3D scan sequences. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 4191–200.

[33] Ye M, Yang R. Real-time simultaneous pose and shape estimation for articulated objects using a single depth camera. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, p. 2345–52.

[34] Zheng Z, Zhao X, Zhang H, Liu B, Liu Y. Avatarrex: Real-time expressive full-body avatars. ACM Trans Graph 2023;42(4):1–19.

[35] Shetty K, Birkhold A, Jaganathan S, Strobel N, Kowarschik M, Maier A, Egger B. Pliks: A pseudo-linear inverse kinematic solver for 3d human body estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 574–84.

[36] Kushwaha M, Choudhary J, Singh DP. 3DPMesh: An enhanced and novel approach for the reconstruction of 3D human meshes from a single 2D image. Comput Graph 2024.

[37] Yuan Y-J, Lai Y-K, Wu T, Xia S, Gao L. Data-driven weight optimization for real-time mesh deformation. Graph Model 2019;104:101037.

[38] TaubinÝ G. Geometric signal processing on polygonal meshes. In: Proceedings of eUROGRAPHICS. 2000.

[39] Baran I, Popović J. Automatic rigging and animation of 3d characters. ACM Trans Graph 2007;26(3):72–es.

[40] Li H, Adams B, Guibas LJ, Pauly M. Robust single-view geometry and motion reconstruction. ACM Trans Graph (ToG) 2009;28(5):1–10.

[41] Horn BK, Hilden HM, Negahdaripour S. Closed-form solution of absolute orientation using orthonormal matrices. Josa A 1988;5(7):1127–35.