Deep 3D Modeling of Human Bodies from Freehand Sketching

Kaizhi Yang, Jintao Lu, Siyu Hu, and Xuejin Chen(✉)

University of Science and Technology of China, Hefei 230026, China
{ykz0923,ljt123,sy891228}@mail.ustc.edu.cn, xjchen99@ustc.edu.cn

Abstract. Creating high-quality 3D human body models by freehand sketching is challenging because of the sparsity and ambiguity of hand-drawn strokes. In this paper, we present a sketch-based modeling system for human bodies using deep neural networks. Considering the large variety of human body shapes and poses, we adopt the widely-used parametric representation, SMPL, to produce high-quality models that are compatible with many further applications, such as telepresence, game production, and so on. However, precisely mapping hand-drawn sketches to the SMPL parameters is non-trivial due to the non-linearity and dependency between articulated body parts. In order to solve the huge ambiguity in mapping sketches onto the manifold of human bodies, we introduce the skeleton as the intermediate representation. Our skeleton-aware modeling network first interprets sparse joints from coarse sketches and then predicts the SMPL parameters based on joint-wise features. This skeleton-aware intermediate representation effectively reduces the ambiguity and complexity between the two high-dimensional spaces. Based on our light-weight interpretation network, our system supports interactive creation and editing of 3D human body models by freehand sketching.

Keywords: Freehand sketch · Human body · 3D modeling · Skeleton joint · SMPL

1 Introduction

The human body is one of the commonest subjects in virtual reality, telepresence, animation production, etc. Many studies have been conducted in order to produce high-quality meshes for human bodies in recent years. However, 3D modeling is a tedious job since most interaction systems are based on 2D displays. Sketching, which is a natural and flexible tool to depict desired shapes, especially at the early stage of a design process, has been attracting a great deal of research works for easy 3D modeling [14,21,24]. In this paper, we propose a powerful sketch-based system for high-quality 3D modeling of human bodies.

Creating high-quality body models is non-trivial because of the non-rigidness of human bodies and articulation between body parts. A positive thing is that human bodies have a uniform structure with joint articulations. Many parametric
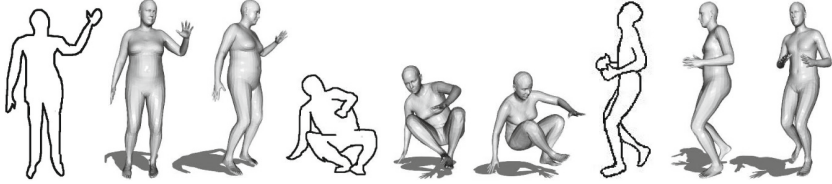
Fig. 1. Our sketch-based body modeling system allows common users to create high-quality 3D body models in diverse shapes and poses easily from hand-drawn sketches.

representations have been proposed to restrict the complicated body models to a manifold [13,15]. Taking advantage of deep neural networks (DNNs), a large number of techniques have been developed to reconstruct human bodies from 2D images [10,12,25–28].

However, hand-drawn sketches show different characteristics compared to natural images. The inevitable distortions in sketch lead to much more ambiguity in the ill-posed problem of 3D interpretation from a single view image. Moreover, hand-drawn sketches are typically sparse and cannot carry as rich texture information as real images which makes sketch-based modeling become more challenging. To address the ambiguity in mapping coarse and sparse sketches to high-quality body meshes, we propose a skeleton-aware interpretation neural network, taking advantage of both non-parametric skeleton construction and parametric regression for more natural-looking meshes. In the non-parametric interpretation stage, we employ a 2D CNN to encode the input sketch into a high-dimensional vector and then predict the 3D positions of joints on the body skeleton. In the parametric interpretation stage, we estimate the body shape and pose parameters from the 2D sketch features and the 3D joints, taking advantage of the SMPL [13] representation in producing high-quality meshes. The disjunctive network branches for pose and shape respectively along with the skeleton-first and mesh-second interpretation architecture efficiently facilitate the quality and accuracy of sketch-based body modeling.

While existing supervised 3D body reconstruction required paired 2D input and 3D models, we build a large-scale database that contains pairs of synthesized sketches and 3D models. Our proposed method achieves the highest reconstruction accuracy compared with state-of-the-art image-based reconstruction approaches. Furthermore, with the proposed light-weight deep neural network, our sketch-based body modeling system supports interactive creation and editing of high-quality body models of diverse shapes and poses, as Fig. 1 shows.

2   Related Work

Sketch-Based Content Creation. Sketch-based content creation is an essential problem in multimedia and human-computer interaction. Traditional methods make 3D interpretation from line drawings based on reasoning geometric constraints, such as convexity, parallelism, orthogonality to produce 3D
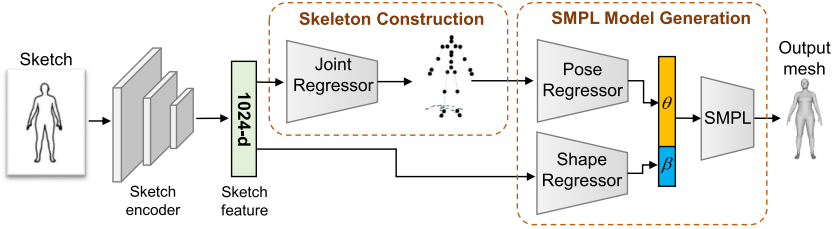
Fig. 2. Overview of our approach with the skeleton as an intermediate representation for model generation. A skeleton-aware representation is introduced before pose parameters regression.

shapes [16,24]. Recently, deep learning has significantly promoted the performance of sketch-based systems, such as sketch-based 3D shape retrieval [20], image synthesis, and editing [11]. Using a volumetric representation, [8] designs two networks to respectively predict a start model from the first view sketches and update the model according to the drawn strokes in other views. [14] predicts depth and normal maps under multiple views and fuses them to get a 3D mesh. An unsupervised method is proposed by searching similar 3D models with the input sketch and integrating them to generate the final shape [21]. Due to the limited capability of shape representation, the generated shapes by these approaches are very rough in low resolutions. Moreover, precise control of the 3D shapes with hand-drawn sketches is not supported.

Object Reconstruction from Single-View Image. With the advance of large-scale 3D shape datasets, DNN-based approaches have made great progress in object reconstruction from single images [1,3,4,6,22,23]. [1] predicts an occupancy grid from images and incrementally fusing the information of multiple views through a 3D recurrent reconstruction network. In [3], by approximating the features after image encoder and voxel encoder, a 3D CNN is used to decode the feature into 3D shapes. Also, many efforts have been devoted to mesh generation from single-view images [4,22]. Starting from a template mesh, they interpret the vertex positions for the target surface through MLPs or graph-based networks. A differentiable mesh sampling operator [23] is proposed to sample point clouds on meshes and predict the vertex offsets for mesh deformation. In comparison, we combine non-parametric joint regression with a parametric representation to generate high-quality meshes for human bodies from sketches.

Human Body Reconstruction. Based on the unified structure of human bodies, parametric representation is usually used in body modeling. Based on Skinned Multi-Person Linear Model (SMPL) [13], many DNN-based techniques extract global features from an image and directly regress the model parameters [9,18,26–28]. However, a recent study [10] shows that direct estimation of vertex positions outperforms parameter regression because of the highly nonlinear correlation between SMPL parameters and body shapes. They employ a graph-CNN to regress vertex positions and attach a network for SMPL param-
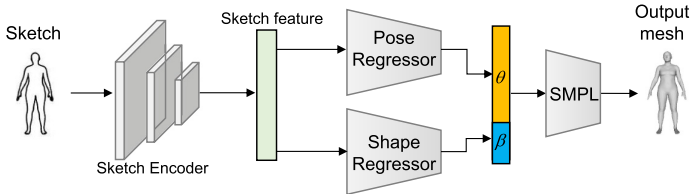
Fig. 3. A plain network (MLP-Vanilla) to regress SMPL parameters from sketches for body mesh generation.

eter regression to improve the mesh quality. Except for global features that are directly extracted from image, semantic part segmentation can also be involved to help reconstruct the body models [17,19,26,27]. Besides single images, multi-view images [12,25] or point clouds [7] can be other options to reconstruct body models. However, different from natural images with rich texture information or point clouds with accurate 3D information, hand-drawn sketches are much more sparse and coarse without sufficient details and precision, making the body modeling task more challenging.

3   Our Method

Hand-drawn sketches are typically coarse and only roughly depict body shapes and poses, leaving details about fingers or expressions unexpressed. In order to produce naturally-looking and high-quality meshes, we employ the parametric body representation, SMPL [13] to supplement shape details. Therefore, the sketch-based modeling task becomes a regression task of predicting SMPL parameters. The SMPL parameters consist of shape parameters $\beta \in \mathbb{R}^{10}$ and pose parameters θ for the rotations associated with $N_j = 24$ joints. Similar to [10], we regress the 3×3 matrix representation for each joint rotation, i.e. $\theta \in \mathbb{R}^{216}$, to avoid the challenge of regressing the quaternions in the axis-angle representation.

When mapping sketches into SMPL parameters, a straightforward method is to encode the input sketch image into a high-dimensional feature vector and then regress the SMPL parameters from the encoded sketch feature, as Fig. 3 shows. A K_s dimensional feature vector $\mathbf{f}_{sketch}$ can be extracted from the input sketch S using a CNN encoder. From the sketch feature $\mathbf{f}_{sketch}$, the shape parameters β and the pose parameters θ are then separately regressed using MLPs.

However, due to the non-linearity and complicated dependency between the pose parameters, it is non-trivial to find a direct mapping from the sketch feature space to the SMPL parameter space. We bring the skeleton as an intermediate expression, which unentangles the global pose and the significantly nonlinear articulated rotations. Figure 2 shows the pipeline of our skeleton-aware model generation from sketches.

3.1   Intermediate Skeleton Construction

Although the rough sketch could not precisely describe local details of human bodies, it effectively delivers posture information. A 3D skeleton can efficiently describe the pose of a human body, while shape details such as fingers, facial features are not considered in this stage. Based on the underlying common sparsity, the mapping from 2D sketches and 3D body skeletons is more effective to interpret. We use the skeleton structure that is composed of $N_j = 24$ ordered joints with fixed connections, as defined in [13]. In order to construct the 3D body skeleton based on the input sketch, the 3D locations of the N_j joints can be regressed using multi-layer perceptrons from the extracted sketch feature $\mathbf{f}_{sketch}$, as shown in Fig. 4. By concatenating the N_j joints' 3D coordinates, the pose parameters θ can be regressed from the $N_j \times 3$ vector using MLPs.
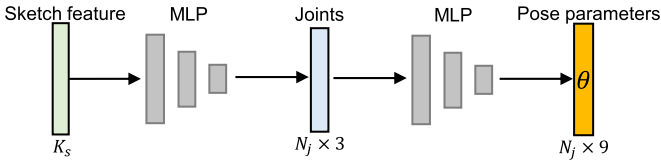

Fig. 4. MLP-Joint Network. With skeleton as an intermediate representation, we first regress the positions of N_j joints on the skeleton with three MLPs and then regress the SMPL pose parameters with another three MLPs.

3.2   Joint-Wise Pose Regression

Direct regression of the non-linear pose parameters from the joint positions inevitably abandons information in the original sketch. On the other hand, naively regression the pose parameters from the globally embedded sketch feature ignores the local articulation between joints. Since pose parameters represent the articulated rotation associated with the joints under a unified structure, we propose an effective joint-wise pose regression strategy to regress the highly non-linear rotations. First, we replace the global regression of joint positions with deformation from a template skeleton. Second, we regress the pose parameters from the embedded joint features rather than the joint positions. Figure 5 shows the detailed architecture of our pose regression network.
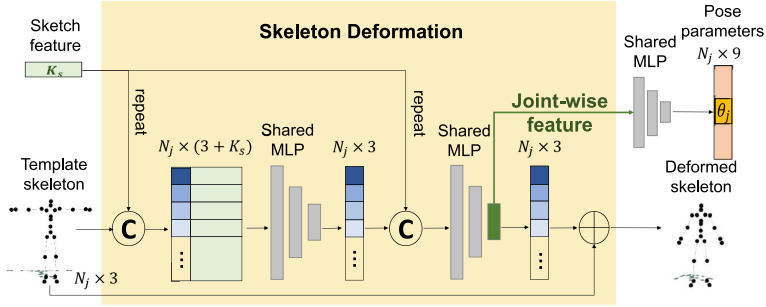
Fig. 5. Detailed architecture of our joint-wise pose regression in the Deform+JF network. We first regress the local offset of each joint from the template skeleton. The pose parameters are regressed from the last layer feature for each joint rather than the deformed joint position using shared MLPs.

Skeleton Deformation. Different from general objects, despite the complex shapes and high dimensional meshes, the human bodies have a unified structure, so a template skeleton could efficiently embed the structure and all the pose of human bodies could be concisely explained as the joints movement in skeleton instead of the deformation of all the vertices in the mesh. We adopt a two-level deformation network for skeleton construction. Each joint coordinate (x, y, z) on the template skeleton is concatenated with the K_s dimensional global sketch feature $\mathbf{f}_{sketch}$ to go through a shared MLP to estimate (d_x^1, d_y^1, d_z^1) for the first step deformation. Another deformation step is performed in a similar manner to get joint-wise deformation (d_x^2, d_y^2, d_z^2). Adding the predicted joint offset to the template skeleton, we get a deformed body skeleton.

Joint-Wise Pose Regression. Benefited from the shared MLP in the skeleton deformation network, we further present joint-wise pose regression based on the high-dimensional vector as joint features (the green vector in Fig. 5). In practice, as all the joints will go through the same MLPs in the deformation stage, this high dimensional representation contains sufficient information which not only from spatial location but also from the original sketch. The joint feature is then fed into another shared MLPs to predict the rotation matrix of each joint.

3.3 Loss

Based on the common pipeline with a skeleton as the intermediary, the above-mentioned variants of our sketch-based body modeling network can all be trained end-to-end. The loss function includes a vertex loss, SMPL parameter loss for the final output, as well as a skeleton loss for the intermediate skeleton construction.

Vertex Loss. To measure the difference between the output body mesh and the ground truth model, we apply a vertex loss L_{vertex} on the final output SMPL

models. The vertex loss is defined as

$$L_{vertex} = \frac{1}{N_{ver}} \sum_{i=1}^{N_{ver}} \|\hat{\mathbf{v}}_i - \mathbf{v}_i\|_1, \tag{1}$$

where $N_{ver} = 6890$ for the body meshes in SMPL representation. $\hat{\mathbf{v}}_i$ and $\mathbf{v}_i$ are the 3D coordinates of the i-th vertex in the ground-truth model and the predicted model respectively.

Skeleton Loss. We use the L_2 loss of the N_j ordered joints on the skeleton to constrain the intermediate skeleton construction with the joint regressor or skeleton deformation network.

$$L_{skeleton} = \frac{1}{N_j} \sum_{i=1}^{N_j} \|\hat{\mathbf{p}}_i - \mathbf{p}_i\|_2, \tag{2}$$

where $\hat{\mathbf{p}}$ is the ground truth location of each joint, and $\mathbf{p}_i$ is the predicted joint location.

SMPL Regression Loss. For the SMPL regressor, we use the L_2 losses on the SMPL parameters θ and β. The entire network is trained end-to-end with the three loss items as:

$$L_{model} = L_{vertex} + L_{skeleton} + L_\theta + \lambda L_\beta, \tag{3}$$

where $\lambda = 0.1$ to balance the parameter scales.

4   Experiments and Discussion

Our system is the first one that uses DNNs for sketch-based body modeling. There is no existing public dataset for training DNNs and make a comprehensive evaluation. In this section, we first introduce our sketch-model datasets. Then we explain more implementation details and training settings. We conduct both quantitative evaluation on multiple variants and qualitative test with freehand sketching to demonstrate the effectiveness of our method.

4.1   Dataset

To evaluate our network, we build a dataset with paired sketches and 3D models. The 3D models come from the large-scale dataset AMASS [15] which contains a huge variety of body shapes and poses with fully rigged meshes. We select 25 subjects from a large subset BMLmovi in AMASS dataset to test our algorithm for its large diversity of body shapes and poses. For each subject, we pick one model every 50 frames. In total, we collect 7097 body models with varied shapes and poses. For each subject, we randomly select 60% for training, 20% for validation, and 20% for the test. To generated paired sketch and model for training

and quantitative evaluation, we render sketches for each body model using suggestive contours [2] to imitate hand-drawn sketches. For each model, we render a sketch-style image from one view. Figure 6 shows two examples of synthesized sketches from 3D models and hand-drawn sketches by a non-expert user. We can see that the synthesized sketch images look similar to hand-drawn sketches. By rendering a large number of synthesized sketches, the cost of collecting sketch and model pairs is significantly reduced and quantitative evaluation is allowed. Our experiments demonstrate that freehand sketching is well supported in our sketch-based body modeling system without performance degradation.
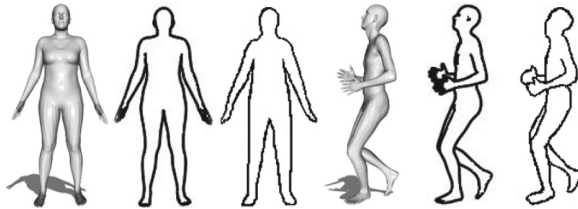

Fig. 6. Two examples of synthesized sketches (middle) from a 3D model (left) and hand-drawn sketches (right).

4.2 Network Details and Training Settings

We use ResNet-34 [5] as our sketch encoder and set the sketch feature dimension $K_s = 1024$ in our method. We introduce three variants of network to map the sketch feature $\mathbf{f}_{sketch} \in \mathbb{R}^{1024}$ into SMPL parameters $\beta \in \mathbb{R}^{10}$ and $\theta \in \mathbb{R}^{216}$. The first, called MLP-Vanilla, as shown in Fig. 3, consists MLP of (512, 256, 10) for the shape regressor and (512, 512, 256, 216) for the pose regressor. The second, call MLP-Joint, employs an MLP of (512, 256, 72) for the joint regressor and an MLP of (256, 256, 216) for pose regressor as Fig. 4 shows. The third one is our full model (named Deform+JF) involving skeleton deformation and joint-wise features for skeleton construction and pose regression, as Fig. 5 shows. In the skeleton deformation part, the two-level shared-MLP of (512, 256, 3) is used to map the $3 + 1024$ input to the 3D vertex offset for each joint. For pose regression, we use the shared-MLP of (128, 64, 9) to regress the 9 elements of a rotation matrix from the 256-D joint-wise feature for each joint.

The networks are trained end-to-end. Each network was trained for 400 epochs with a batch size of 32. Adam optimizer is used with $\beta_1 = 0.9$, $\beta_2 = 0.999$. The learning rate is set as 3e-4 without learning rate decay.

4.3 Results and Discussion

Comparison with Leading Methods. In order to evaluate the effectiveness of our method, we compare our method with two state-of-the-art reconstruction

Table 1. Quantitative evaluation of body reconstruction performance (errors in mm) on the test set. The number of network parameters and GFLOPs are also reported.

Method	RE	MPJPE	N_{para} (M)	GFLOPS
3DN [23]	119.06	95.19	151.77	190.58
CMR [10]	43.22	37.45	40.58	4.046
MLP-Vanilla	42.49	36.42	23.43	3.594
MLP-Joint	39.75	34.10	23.27	3.594
Deform+JF	**38.09**	**32.38**	23.82	3.625

techniques, training by replacing images with synthesized sketches. We select the convolutional mesh regression (CMR) [10] for single-image human body reconstruction and the 3DN model [23] which is a general modeling framework based on mesh deformation. We follow their default training settings except removing the 2D projected joint loss in CMR since it's non-trivial to estimate the perspective projection matrix from hand-drawn sketches.

Two error metrics, the reconstruction error (RE) and the mean per joint position error (MPJPE), are used to evaluate the performance of sketch-based modeling. 'RE' is the mean Euclidean distance between corresponding vertexes in the ground truth and predicted models. The 'MPJPE' reports the mean per joint position error after aligning the root joint. Table 1 lists the reconstruction performance of the three variants of our method and two SOTA methods. The three variants of our network all outperform existing image-based methods when trained for sketch-based body modeling. Taking advantage of the intermediate skeleton representation, our full model effectively extracts features from sketches and produces precise poses (32.38 mm of MPJPE) and shapes (38.09 mm of RE). Moreover, three variants of our method only have about 23M parameters (about half of CMR [10], 0.15 of 3DN [23]) and the least computational complexity.

Ablation Study. With regard to the three variants of our method, the MLP-Vanilla version directly regresses the shape pose parameters from global features extracted from the input sketch. As the existing image-based method may not suitable for sparse sketches, we set the vanilla version as our baseline. Table 1 shows that this plain network is effective in mapping sparse sketches on the parametric body space. It works slightly better than the CMR model. Bringing our main concept of using the skeleton as the intermediate representation, our MLP-Joint network reduces the reconstruction error 2.74 mm compared with MLP-Vanilla. Furthermore, by introducing joint features into the pose regressor, our Deform+JF network improves the modeling accuracy by 1.66 mm.

Qualitative Comparison. Figure 7 shows a group of 3D body models interpreted from synthesized sketches by different methods. As 3DN [23] does not use any prior body structure, it has difficulty to regress a naturally-looking body
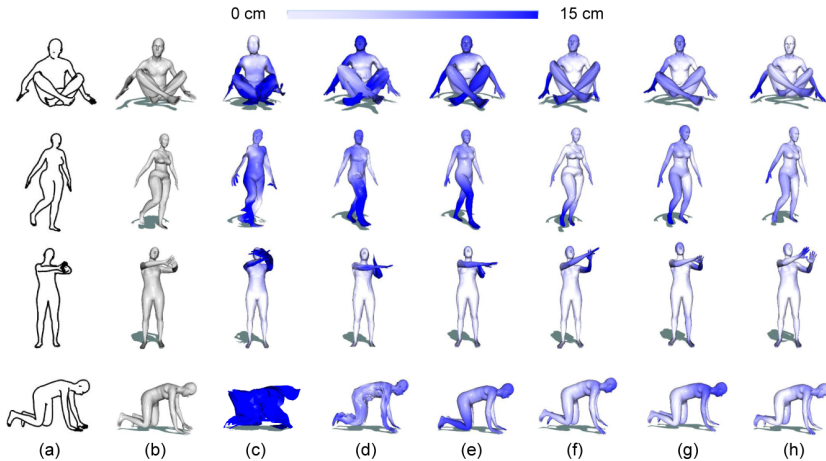
Fig. 7. 3D Meshes generated from synthesized sketches using different methods. The vertex errors are color-coded. From left to right: (a) input sketches; (b) ground truth; (c) results of 3DN [23]; (d) deformation results of CMR [10]; (e) SMPL regression results of CMR [10]; (f) results of MLP-Vanilla; (g) results of our MLP-Joint network; (h) results of our Deform+JF network which regresses parameters from joint-wise features.

model in a huge space. As one of the state-of-the-art method of image-based body reconstruction, CMR's [10] dense deformation by graph convolutions for regressing 3D locations of a large number of vertices rely on rich image features, its performance degrades in sparse and rough sketches. In comparison, as shown in the last three columns in Fig. 7, our method achieves high-quality body models from sketches. From Fig. 7 (f) to (h), the generated body models get better as the reconstruction errors get lower, especially for the head and limbs. And our Deform+JF network shows the most satisfactory visual results.

Skeleton Interpretation. We also compared the joint errors that are directly regressed from the sketch features and after SMPL regression in our network. Table 2 lists the errors of skeleton joints that directly regressed from sketch features and those calculated from the output SMPL models. We can see that our Deform+JF with SMPL could achieve the lowest MPJPE with 32.38 mm. On the other hand, although without strict geometric constraints on joints in the SMPL model, our MLP-Joint and Deform+JF's skeleton regressor still get high-quality skeletons, as shown in Fig. 8. Moreover, by introducing joint features, the Deform+JF network not only obtains more precise final SMPL models but also achieves better skeleton estimation accuracy (46.99 mm) than the MLP-Joint network (62.49 mm), which proves the effectiveness of joint-wise features.
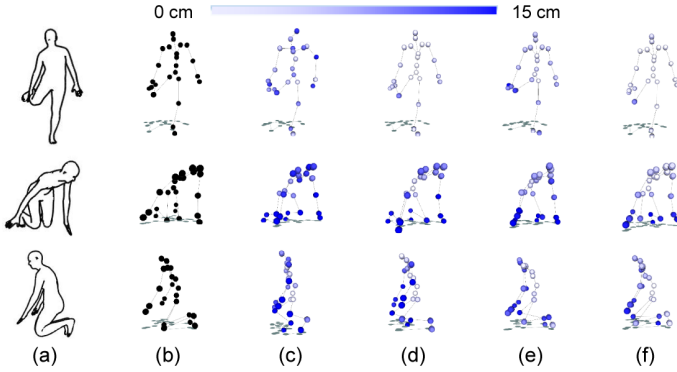
Fig. 8. Reconstructed body skeletons from sketches using different methods. From left to right: (a) input sketch; (b) ground truth; (c) regression result from MLP-Joint; (d) skeleton from the output SMPL model by MLP-Joint; (e) regression result from Deform+JF; (f) skeleton from the output SMPL model by Deform+JF.

Table 2. Comparison MPJPE (in mm) of skeletons generated by different ways.

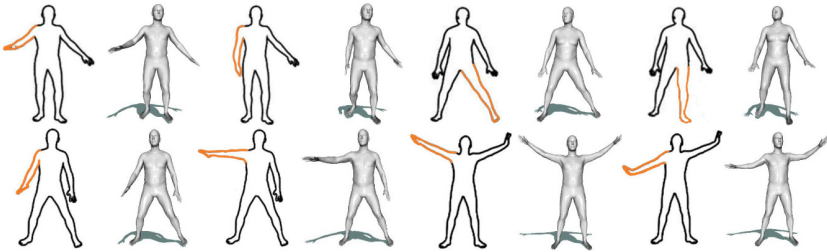| Method	MLP-Joint		Deform+JF	
	Regressed	SMPL	Regressed	SMPL
MPJPE	62.49	34.10	46.99	32.38

Fig. 9. Interactive editing by sketching. The user modifies some strokes as highlighted to change body poses.

4.4   Body Modeling by Freehand Sketching

Though training with synthetic sketches, our model is also able to interpret 3D body shapes accurately from freehand sketching. We asked several graduate students to test our sketch-based body modeling system. Figure 1 and Fig. 9 show a group of results in diverse shapes and poses. Though the sketches drawn by non-expert users are very coarse with severe distortions, our system is able to capture the shape and pose features and produce visually pleasing 3D models. As shown in Fig. 9, based on our lightweight model, users could modify the sketch in part and get a modified 3D model in real-time rate which allows non-expert users to easily create high-quality 3D body models with fine-grained control on desired shapes and poses by freehand sketching and interactive editing.

5   Conclusions

In this paper, we introduce a sketch-based human body modeling system. In order to precisely mapping the coarse sketches into the manifold of human bodies, we propose to involve the body skeleton as an intermediary for regression of shape and pose parameters in the SMPL representation which allow effective encoding of the global structure prior of human bodies. Our method outperforms state-of-the-art techniques in the 3D interpretation of human bodies from hand-drawn sketches. The underlying sparsity of sketches and shape parameters allows us to use a lightweight network to achieve accurate modeling in real-time rates. Using our sketch-based modeling system, common users can easily create visually pleasing 3D models in a large variety of shapes and poses.

References

1. Choy, C.B., Xu, D., Gwak, J.Y., Chen, K., Savarese, S.: 3D-R2N2: a unified approach for single and multi-view 3D object reconstruction. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 628–644. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_38
2. DeCarlo, D., et al.: Suggestive contours for conveying shape. In: ACM SIGGRAPH, pp. 848–855 (2003)
3. Girdhar, R., Fouhey, D.F., Rodriguez, M., Gupta, A.: Learning a predictable and generative vector representation for objects. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 484–499. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46466-4_29
4. Groueix, T., et al.: A papier-mâché approach to learning 3D surface generation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 216–224 (2018)
5. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
6. Wu, J., et al.: Learning shape priors for single-view 3D completion and reconstruction. In: Proceedings of the European Conference on Computer Vision, pp. 646–662 (2018)
7. Jiang, H., et al.: Skeleton-aware 3D human shape reconstruction from point clouds. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5431–5441 (2019)
8. Delanoy, J., et al.: 3D sketching using multi-view deep volumetric prediction. Proc. ACM Comput. Graph. Interact. Tech. **1**(1), 1–22 (2018)
9. Kanazawa, A., et al.: End-to-end recovery of human shape and pose. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7122–7131 (2018)

10. Kolotouros, N., et al.: Convolutional mesh regression for single-image human shape reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4501–4510 (2019)
11. Li, Y., et al.: LinesToFacePhoto: face photo generation from lines with conditional self-attention generative adversarial networks. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 2323–2331 (2019)
12. Liang, J., et al.: Shape-aware human pose and shape reconstruction using multi-view images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4352–4362 (2019)
13. Loper, M., et al.: SMPL: a skinned multi-person linear model. ACM Trans. Graph. **34**(6), 1–16 (2015)
14. Lun, Z., et al.: 3D shape reconstruction from sketches via multi-view convolutional networks. In: International Conference on 3D Vision, pp. 67–77 (2017)
15. Mahmood, N., et al.: AMASS: archive of motion capture as surface shapes. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5442–5451 (2019)
16. Olsen, L., et al.: Sketch-based modeling: a survey. Comput. Graph. **33**(1), 85–103 (2009)
17. Omran, M., et al.: Neural body fitting: unifying deep learning and model based human pose and shape estimation. In: International Conference on 3D Vision, pp. 484–494 (2018)
18. Tan, J.K.V., et al.: Indirect deep structured learning for 3d human body shape and pose prediction (2017)
19. Venkat, A., et al.: HumanMeshNet: polygonal mesh recovery of humans. In: Proceedings of the IEEE International Conference on Computer Vision Workshops (2019)
20. Wang, F., et al.: Sketch-based 3D shape retrieval using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1875–1883 (2015)
21. Wang, L., et al.: Unsupervised learning of 3D model reconstruction from hand-drawn sketches. In: Proceedings of the 26th ACM International Conference on Multimedia, pp. 1820–1828 (2018)
22. Wang, N., et al.: Pixel2Mesh: generating 3D mesh models from single RGB images. In: Proceedings of the European Conference on Computer Vision, pp. 52–67 (2018)
23. Wang, W., et al.: 3DN: 3D deformation network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1038–1046 (2019)
24. Zeleznik, R.C., et al.: SKETCH: an interface for sketching 3D scenes. In: ACM SIGGRAPH Courses, pp. 9-es (2006)
25. Zhang, H., et al.: DaNet: decompose-and-aggregate network for 3D human shape and pose estimation. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 935–944 (2019)
26. Zheng, Z., et al.: DeepHuman: 3D human reconstruction from a single image. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 7739–7749 (2019)
27. Xu, Y., et al.: DenseRaC: joint 3D pose and shape estimation by dense render-and-compare. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 7760–7770 (2019)
28. Kolotouros, N., et al.: Learning to reconstruct 3D human pose and shape via model-fitting in the loop. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2252–2261 (2019)