DIG: Draping Implicit Garment over the Human Body
— Supplementary Material —

Ren Li[1][0000−0003−2998−7104], Benoît Guillard[1][0000−0002−8747−6153], Edoardo Remelli[2][0000−0002−8506−9191], and Pascal Fua[1][0000−0002−5477−1017]

[1] CVLab, EPFL, Switzerland
{ren.li, benoit.guillard, pascal.fua}@epfl.ch
[2] Meta Reality Labs Research, Zurich, Switzerland
edoremelli@fb.com

1 Dataset Pre-Processing

The garment mesh template in CLOTH3D [2] is designed for T-pose body with shape $\beta \neq 0$. To transform it back to the body with neutral shape ($\beta = 0$), for each garment vertex $\mathbf{v}_g$, we first find its closest body vertex $\mathbf{v}_b$ and get the corresponding shape displacement $B_s(\mathbf{v}_b, \beta)$ from SMPL [4], and then $\mathbf{v}_g$ is replaced by the new position $\mathbf{v}_g - B_s(\mathbf{v}_b, \beta)$, which gives us the garment on the neutral body. We also apply a Laplacian operator with $\lambda = 0.2$ afterwards to smooth the transformed garment.

2 Weight coefficients

Following [6], we use $\lambda_{grad} = 0.1$ and $\lambda_{reg} = 0.001$ in Eq. 2. In Eq. 11, we use $\lambda_{deform} = 1$, $\lambda_{interp} = 0.1$ and $\lambda_{order} = 0.01$. When the values of λ_{interp} and λ_{order} are too large, the deformation accuracy decreases because of the trade-off between loss terms. Moreover, using a larger value for λ_{interp} does not prevent interpenetration around the armpits as shown in the main paper Fig. 7(c) since arms can get very close to the body for some poses and self-collision can even happen on the GT body mesh as illustrated by Fig. 6(a).

3 Additional Evaluation

3.1 Examples of Reconstruction

In Fig. 1 and 2, we show some samples of shirts and trousers reconstructed by our SDF model.

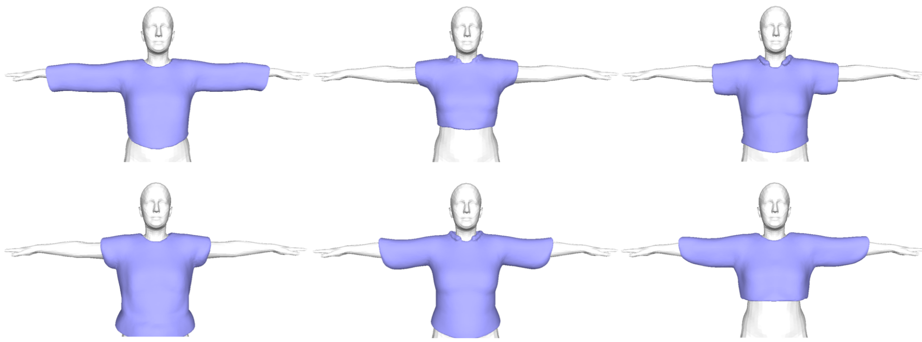

Fig. 1. The samples of shirts reconstructed by our SDF model.

Fig. 2. The samples of trousers reconstructed by our SDF model.

3.2 Ablation Study of L_{order}

In Table. 1, we compare the deformation results for our models trained with and without L_{order}. Using L_{order} can decrease the deformation accuracy but the margin is small. Considering the efficacy of L_{order} in resolving garment self-intersections as illustrated in Fig. 4(b) of the main paper, it is still necessary for us to have it for the better visual quality.

Shirt	CD ($\times 10^{-4}$)	NC (%)	IR (%)
w/o L_{order}	**2.62**	**87.3**	**1.0**
w/ L_{order}	3.78	84.7	1.5

Table 1. The evaluation results of our deformation models trained w/ and w/o L_{order} on test sequences.

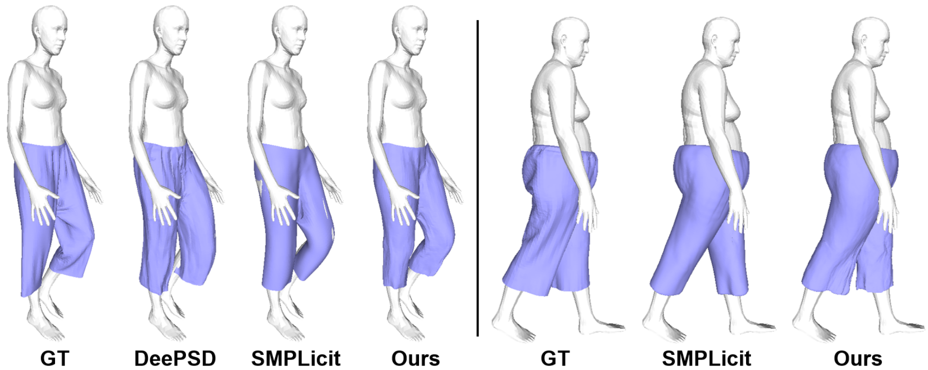

Fig. 3. The skinning results for the ground-truth trousers (left) and the SDF reconstructed trousers (right). Since the input of DeePSD should be the point cloud of the mesh template, we only evaluate it with the unposed ground-truth mesh. Compared to DeePSD and SMPLicit, our method can produce more realistic details and have less body-garment interpenetrations.

3.3 Qualitative Results of Deformation

In Fig. 3, we show the deformation results of DeePSD [1], SMPLicit [3] and ours for the trousers. Similar to the results on shirts, our method outperforms the baselines and shows more realistic details and less body-garment interpenetrations.

Fig. 4 shows the interpolation of the body shape with the same shirt. Notice that our method can produce reasonable deformation results for bodies with fat/slim and short/tall figures. Fig. 5 shows the draping results of different shirts (randomly generated by our SDF model) on body in different shapes. The reconstructed shirts have different styles and lengths, but our model is still able to produce natural dynamics for them.

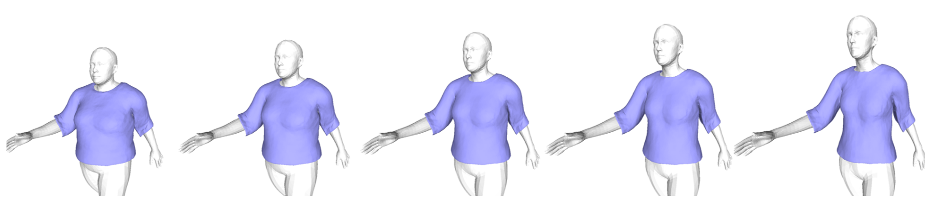

Fig. 4. The draping results of the same shirt on body with different shapes.

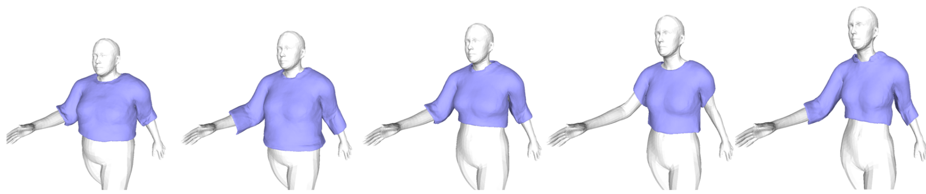

Fig. 5. The draping results of different shirts, randomly generated by our SDF model, on body with different shapes.

3.4 The Accuracy of Body Reconstruction

We conducted an experiment to measure the accuracy of body reconstruction. We use the data from CLOTH3D [2] since it has the ground-truth body and garment mesh. We run [5] to estimate the SMPL parameters $\hat{\beta}$ and $\hat{\theta}$ for the input image, and use them as the initial values for β and θ to perform the optimization of Eq. 15. We use Mean Vertex Error (MVE) to measure the error between the ground-truth body mesh and the reconstruction. The MVE of the initial body estimated by [5] is 127.2mm, while the MVE of our recovered mesh is 43.2mm. This demonstrates that our optimization can correct inaccuracies in the initial body mesh.

4 Limitations

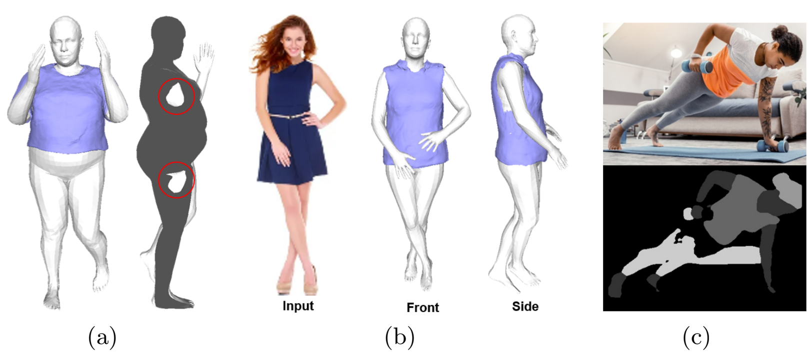

Fig. 6. (a) Self-collision on body mesh (red circles). (b) A fitting example with inaccurate recovery of dress and pose due to the absence of dress in the training data and depth ambiguity. (c) Noisy segmentation result of [7].

Body-garment interpenetration can happen around armpits due to the small distance between arms and the body under specific poses (Fig. 6(a)). The performance of our model is a function of the mesh data used in training, so it is hard for us to recover the garment type not included in the training set (e.g., dress as Fig. 6(b)). Besides, our fitting strategy requires segmentation masks estimated by the segmentation algorithm (e.g., [7]), which can fail under challenging scenarios and is not able to resolve depth ambiguity for pose as shown in Fig. 6(b) and 6(c). In the future we plan to address interpenetration at armpits by operating in coordinates relative to the underlying SMPL template and depth ambiguity by considering multi-view or depth observations.

References

1. Bertiche, H., Madadi, M., Tylson, E., Escalera, S.: DeePSD: Automatic Deep Skinning and Pose Space Deformation for 3D Garment Animation. In: International Conference on Computer Vision (2021)
2. Bertiche, H., Madadi, M., Escalera, S.: CLOTH3D: clothed 3d humans. In: European Conference on Computer Vision. pp. 344–359 (2020)
3. Corona, E., Pumarola, A., Alenya, G., Pons-Moll, G., Moreno-Noguer, F.: Smplicit: Topology-Aware Generative Model for Clothed People. In: Conference on Computer Vision and Pattern Recognition (2021)
4. Loper, M., Black, M.: Opendr: An Approximate Differentiable Renderer. In: European Conference on Computer Vision. pp. 154–169 (2014)
5. Rong, Y., Shiratori, T., Joo, H.: Frankmocap: Fast monocular 3d hand and body motion capture by regression and integration. In: arXiv Preprint (2020)
6. Saito, M., Okatani, T., Deguchi, K.: Application of the Mean Field Methods to MRF Optimization in Computer Vision. In: Conference on Computer Vision and Pattern Recognition (June 2012)
7. Yang, L., Song, Q., Wang, Z., Hu, M., Liu, C., Xin, X., Jia, W., Xu, S.: Renovating parsing R-CNN for accurate multiple human parsing. In: European Conference on Computer Vision. pp. 421–437 (2020)