Modeling Non-photorealistic 3D Characters
PhD Research Proposal

by Tianyu Luan

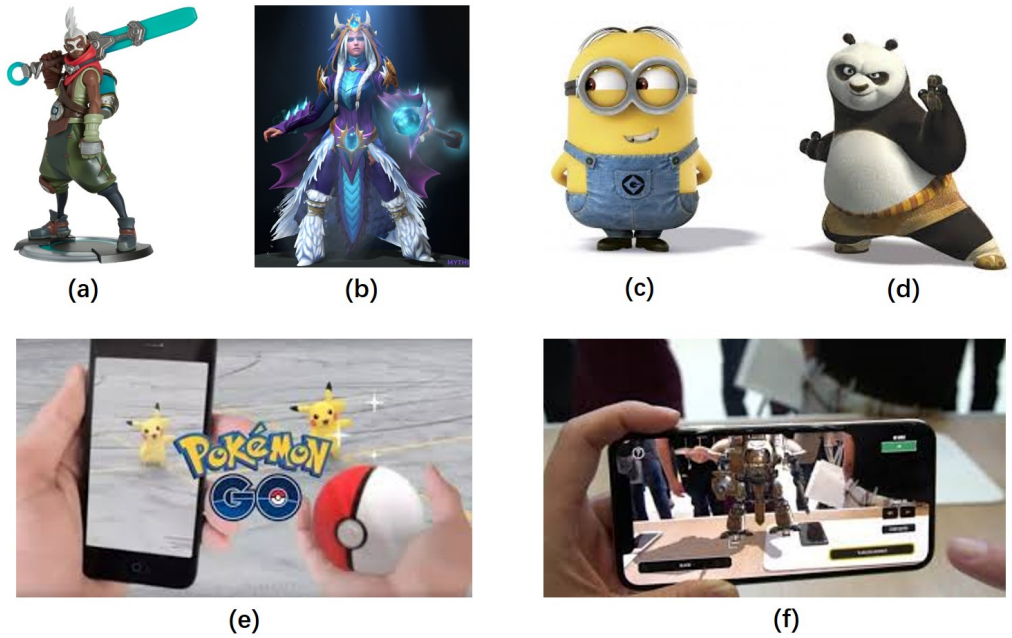

Figure 1: Application examples of non-photorealistic 3D characters, including (a)(b) computer games, (c)(d) cartoon videos, and (e)(f) AR/VR applications.

1 Introduction

Photorealistic 3D human reconstruction have always been an important topic for industrial application. In modern computer games, such as NBA2K (1) and GTA5 (2), realistic 3D human avatars are vital components; In the application of AR and VR, a fast approach of human mesh reconstruction along with texture is also required. Recently, remarkable research progress has been made on realistic 3D human reconstruction based on parametric models and deep learning.

However, realistic human avatars are not sufficient enough to feed the industrial needs. Non-photorealistic 3D virtual characters have been widely used in industries such as computer games, cartoon videos, and AR/VR applications. Fig. 1 shows a few examples of such applications. Instead of precise human avatars, the design these characters aims for non-photorealistic human characters with various designing styles. Typically, those characters would require professional designers to handcraft them case-by-case, which is expansive, time consuming and not available for massive production. On the contrary, reconstructing them by computers can avoid those disadvantages. Thus, with the growth of computer game and the AR/VR industry, the automatic generation of non-photorealistic 3D characters is now becoming a meaningful topic.

There are a few challenges that need to be addressed in designing non-photorealistic 3D characters, including (1) modeling the large shape and style variations among different characters, (2) combined representation of character's body and "decorations" (such as clothes, accessories, hairstyle, weapons, etc.), (3) the abstraction and transfer of designing style. Overcoming these challenges would allow us

to achieve tasks such as 3D modeling from sketch (2D-3D lifting) and style transfer. We will further discuss these tasks in Section 4.

2 Research Objective

This project will focus on modeling non-photorealistic 3D characters. Specifically, we will design general representations for non-photorealistic models and learning-based frameworks to achieve sketch-based 3D character modeling and style transfer.

3 Related Works

3.1 Parametric 3D human representation.

Parametric character reconstruction is a field that has not yet been touched, but the highly related task, parameterized 3D human reconstruction, has been profoundly researched. Multiple human body parametric models, such as SCAPE (6), SMPL-based (27; 33; 3; 31; 9), DenseRaC (41) and STAR (40), are designed for human bodies of various shapes and poses. These methods have been proved to be functional in modeling coarse outlooks of realistic human. Thus, given the latent similarity between human and character reconstruction, it is our best choice to design the character parameterized model based on human parameterized model.

3.2 Sketch based 3D object modeling.

Sketch-based 3D object modeling (SBM) is a widely researched task. It is highly related to our task of sketch-based virtual character reconstruction. Initial works such as (14; 21; 29; 22; 19) managed to reconstruct simple 3D objects from drawings of few lines, using determined mapping from lines to faces. Since these methods mainly focus on simple objects, their effectiveness and robustness are questionable when applied to more complex objects such as game or cartoon characters. More recent works (10; 8; 39) leveraged skeletons to regulate the reconstruction, and optimized energy functions to find the best possible surface and color. Although it is possible for them to handle complex input sketches, color and shape defects always occur due to the optimization approach's poor understanding of the input image. In deep learning era, (37; 43; 11; 26; 16; 15; 12) use Convolutional Neural Networks (CNNs) to generate 3D models in an end-to-end manner. Some public datasets such as (17; 25; 24) are also widely used for deep neural network training. The robustness and capability of deep learning methods handling complex stretch largely exceeded traditional methods.

However, all the above methods are designed to retrieve arbitrary objects. When applying those methods to virtual characters, the results could be sometimes unreasonable in the depth dimension, given the ill-posed setting of this task and the lack of prior information of human-alike characters. In our work, we will obtain more reasonable and robust performance by introducing a parameterized model of the character, which serves as a human-alike prior and regulation.

3.3 3D human reconstruction.

Another task related to 3D character reconstruction is realistic 3D human reconstruction. Previous works could be categorized into two approaches: parametric-based human reconstruction and voxel-based human reconstruction. In terms of the former approach, (4; 3; 5; 9; 42) have realized realistic and rigged human avatars with good robustness and efficiency. But despite the advantages, these methods have shown poor adaptability for human clothing with large topological variance. As for voxel-based approached, such as (30; 34; 35), directly use voxel as the representation of mesh instead of parametric models. These methods can handle the large variance of human clothing, but the reconstruction results are less stable due to the lack of regularization of parametric model.

In our task of 3D character reconstruction, we need the character to be in well-regulated shapes, while the decoration of the characters requires free representation. Therefore, we improve our own approach from both ideas: for character's body, we will use parametric reconstruction; for representing decoration, we will add voxel-based attachments to that parametric model.

Figure 2: A few examples of collected 3D character dataset.

3.4 Virtual try-on.

Style transfer of 3D character models is still a new research area. But a similar task, a.k.a. human clothes virtual try-on & style transfer, is becoming a popular field. Most previous frameworks on virtual try-on, such as (32; 9; 28; 44; 7), represent human clothing with a parametric model, and try to learn those parameters through data-driven methods. But this kind of approach is perhaps not suitable for characters' style transfer. Characters in cartoons and games have much wilder design than human clothes, so the large variance of character styles would require more flexible representation. Thus, in the field of 3D virtual characters, we plan to use voxel-based method for style transfer to feed the needs of variation of characters' decoration.

3.5 2D image style transfer.

We will learn from the idea of 2D image style transfer to design our 3D character style transfer framework. In particular, previous works such as (23; 38) applied auto-encoder for image style transfer, as well as using Generative Adversarial Network (GAN) (18) for more realistic outputs. In our task, we plan to do similar auto-encoder process in voxel level.

4 Methods and Plans

4.1 3D character datasets (6 months).

We need to build datasets of 3D virtual characters for training. Generally, the dataset should include mesh and texture of 3D virtual characters. After obtaining the 3D dataset, we can render 2D sketches of various angles. For designing 3D character data, we can seek help and cooperation from art design companies, independent designers, and computer game cooperation. At initial stage of experiment, we could collect publicly available data for art design forum for simple verification. Up to now, we have collected over 300 3D characters for experiments. Fig. 2 shows a few examples.

Though we have ways to build a 3D character dataset, it is still complex and expensive for mass acquisition of 3D character data. The typical capacity of 3D human datasets is between 100 and 1000, which is very likely to be a similar case to 3D character datasets. In order to solve the problem of insufficient data, we will build a dataset of virtual characters with only 2D images and videos for self-supervised training. In practice, the 3D dataset would be used for pretraining and the 2D dataset would be used for self-supervision to further adjust the model. Given the easy access of acquiring 2D character images and videos, we could build 2D dataset with much larger scale than 3D dataset, so that the model could be feed with a much larger variety of data to achieve better performance and robustness.

This work package would take about 6 months.

4.2 3D character representation (12 months).

Although previous work has done in-depth research on parametric models of real human bodies, there are no previous studies that focus on the representation of non-photorealistic human-alike characters. As shown in Fig. 3, compared with real human, the parameterized virtual character representation has the following differences: a. the ratio of character body parts is different from that of the real

Figure 3: Special features of non-photorealistic characters, including (a)(b) body ratio, (c)(d) body parts deformation, and (e)(f) large variance of decorations.

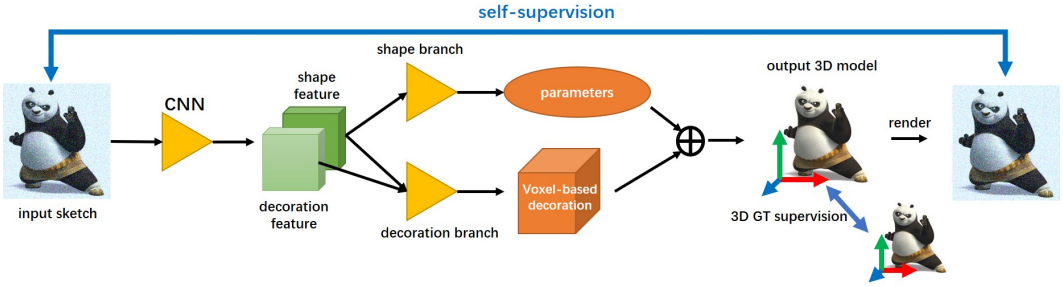

Figure 4: General framework of 3D character reconstruction from scratch.

human body; b. aggravated deformation is also occurred within body parts. c. the decoration of non-photorealistic characters, such as the design of clothes, accessories, weapons, etc., has relatively large variance compared to the decoration of real people. These differences are crucial considerations in designing parameterized character model.

In order to overcome these challenges, we will propose a combined character representation model, including 3 parts: a basic parameterized character mesh, a deformation field defined on the mesh surface, and a voxel-based attachment for characters' decorations. First of all, the basic parameterized model is used to characterize the overall shape of characters, including the basic body shape, body parts proportions, movements, etc. In this part, we will improve the existing human body parameterized models by adding new dimensions for flexible body proportions. Second, we will define a deformation field to adapt to some exaggerated changes in the physical details of the virtual character (such as face shape, nose, tail, etc.). Finally, we will define a voxel-based decorations, so they could be more flexible. The design of the parametric model would take about 12 months.

4.3 3D character reconstruction from scratch (12 months).

Sketch-based 3D character reconstruction has good prospective in industrial application. For art designers, it is complicated and challenging to directly design 3D virtual characters due to technical limitations. On the contrast, drawing a 2D sketch is a relatively simple job to do. Given the parametric 3D character representation, we could build a learning-based framework to reconstruct 3D characters from 2D sketched. The framework would use 2D sketch as input, output 3D characters' mesh and color, and use 3D characters ground truth as supervision. In implementation, auxiliary modules will be applied to separate character's body feature from decoration feature. The overall design of the network is shown in Fig. 4. The encoded 2D features are divided into two branches. The feature containing body shape information are fed to body branch to regress body parameters and deformation field, while another branch handles the occupation and color of voxel-based decoration. The result mesh of both parts will be put together and supervised by the mesh and color of 3D ground truth. Moreover, we could also render the result back to 2D image and implement a self-supervision approach with the input image. Finally, in inference, we spliced the results of both parts together to get the final 3D character mesh and color. This part will take about 12 months.

4.4 3D character style transfer (18 months).

3D character style transfer is also a common problem in application. For instance, transferring a photorealistic human model into a non-photorealistic character would be very helpful in making

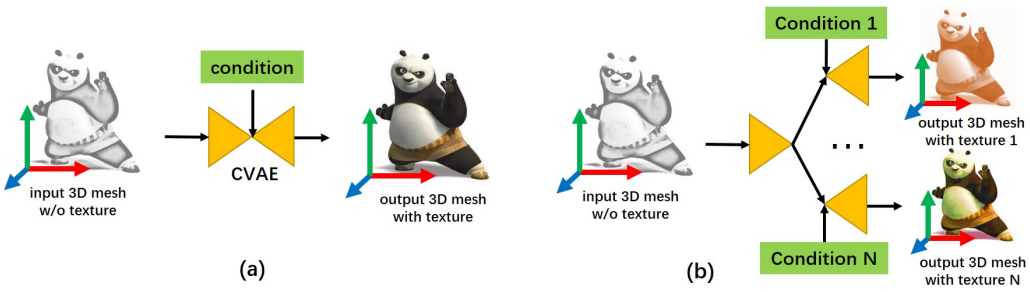

Figure 5: Texture repainting network.(a) In training, we select one condition for each training data and fix them together. (b) In inference, we randomly change the condition to obtain different color designs.

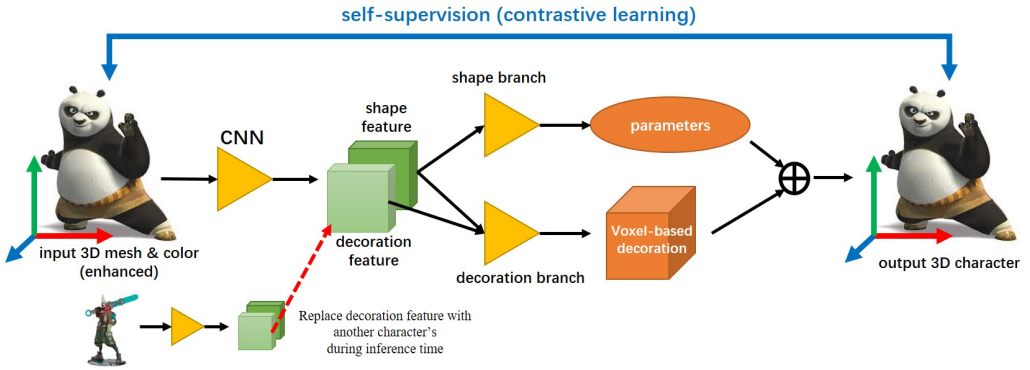

Figure 6: Decoration transfer network. We would apply a contrastive learning approach to separate shape and decoration feature of a given character. In inference, we could achieve decoration transfer by simply replacing the decoration feature with another character's.

compute games or animation films. Other applications such as transferring the decoration of one character to another could also highly improve the efficiency of design. Considering industrial needs and the resource we may have, we plan to focus on two style transfer tasks: texture (or color of the mesh) repainting and decoration transfer.

We plan to design the texture repainting network following Fig. 5. First, we convert the mesh to a voxel-based structure. Second, we feed the 3D voxel map (representing occupation) into a conditional variational auto-encoder (CVAE) (36) network and obtain 3D voxel-based color. We hope this structure could learn the coloring pattern in ground truth data, so that the inference repainting results could have a sense of design. Specifically, CVAE selects a fixed condition for each input data during training, so that by changing that condition in inference, we can obtain different color designs.

For 3D character decoration transfer, we design a framework similar to Fig. 4. We replace the input with a 3D colored mesh, and encode it into shape feature and decoration feature. Then, we follow the idea of contrastive learning (13; 20). By randomly altering the decorations of the characters (data enhancement), the network will better separate shape feature from decoration feature. The network design is shown in the Fig. 6. In inference, we could transfer decoration by replacing the decoration feature of the original characters with the decorative feature of the target characters.

These two tasks will take about 18 months.

References

[1] https://www.nba2k.com.
[2] https://www.rockstargames.com/games/V.
[3] Thiemo Alldieck, Marcus Magnor, Bharat Lal Bhatnagar, Christian Theobalt, and Gerard Pons-Moll. Learning to reconstruct people in clothing from a single rgb camera. In *CVPR*, pages 1175–1186, 2019.
[4] Thiemo Alldieck, Marcus Magnor, Weipeng Xu, Christian Theobalt, and Gerard Pons-Moll. Video based reconstruction of 3d people models. In *CVPR*, pages 8387–8397, 2018.

[5] Thiemo Alldieck, Gerard Pons-Moll, Christian Theobalt, and Marcus Magnor. Tex2shape: Detailed full human body geometry from a single image. In *ICCV*, pages 2293–2303, 2019.

[6] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: shape completion and animation of people. In *SIGGRAPH*, pages 408–416. 2005.

[7] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. Cloth3d: Clothed 3d humans. In *ECCV*, pages 344–359. Springer, 2020.

[8] Mikhail Bessmeltsev, Will Chang, Nicholas Vining, Alla Sheffer, and Karan Singh. Modeling character canvases from cartoon drawings. *Transactions on Graphics (2015)*, 34(5), 2015.

[9] Bharat Lal Bhatnagar, Garvita Tiwari, Christian Theobalt, and Gerard Pons-Moll. Multi-garment net: Learning to dress 3d people from images. In *ICCV*, pages 5420–5430, 2019.

[10] Philip Buchanan, Ramakrishnan Mukundan, and Michael Doggett. Automatic single-view character model reconstruction. pages 5–14, 07 2013.

[11] Jiaxin Chen and Yi Fang. Deep cross-modality adaptation via semantics preserving adversarial learning for sketch-based 3d shape retrieval. In *ECCV*, pages 605–620, 2018.

[12] Jiaxin Chen, Jie Qin, Li Liu, Fan Zhu, Fumin Shen, Jin Xie, and Ling Shao. Deep sketch-shape hashing with segmented 3d stochastic viewing. In *CVPR*, pages 791–800, 2019.

[13] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020.

[14] Maxwell B Clowes. On seeing things. *Artificial intelligence*, 2(1):79–116, 1971.

[15] Guoxian Dai, Jin Xie, and Yi Fang. Deep correlated holistic metric learning for sketch-based 3d shape retrieval. *IEEE Transactions on Image Processing*, 27(7):3374–3386, 2018.

[16] Johanna Delanoy, Mathieu Aubry, Phillip Isola, Alexei A Efros, and Adrien Bousseau. 3d sketching using multi-view deep volumetric prediction. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 1(1):1–22, 2018.

[17] Mathias Eitz, Ronald Richter, Tamy Boubekeur, Kristian Hildebrand, and Marc Alexa. Sketch-based shape retrieval. *ACM Trans. Graph.*, 31(4):31–1, 2012.

[18] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, pages 2672–2680, 2014.

[19] Cindy Grimm and Pushkar Joshi. Just draw it! a 3d sketching system. 2012.

[20] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, pages 9729–9738, 2020.

[21] David A Huffman. Impossible object as nonsense sentences. *Machine intelligence*, 6:295–324, 1971.

[22] Takeo Igarashi, Satoshi Matsuoka, and Hidehiko Tanaka. Teddy: a sketching interface for 3d freeform design. In *SIGGRAPH Courses*, pages 11–es. 2006.

[23] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, pages 1125–1134, 2017.

[24] Bo Li, Yijuan Lu, Afzal Godil, Tobias Schreck, Benjamin Bustos, Alfredo Ferreira, Takahiko Furuya, Manuel J Fonseca, Henry Johan, Takahiro Matsuda, et al. A comparison of methods for sketch-based 3d shape retrieval. *Computer Vision and Image Understanding*, 119:57–80, 2014.

[25] Bo Li, Yijuan Lu, Chunyuan Li, Afzal Godil, Tobias Schreck, Masaki Aono, Martin Burtscher, Hongbo Fu, Takahiko Furuya, Henry Johan, et al. Shrec'14 track: Extended large scale sketch-based 3d shape retrieval. In *Eurographics workshop on 3D object retrieval*, volume 2014, 2014.

[26] Changjian Li, Hao Pan, Yang Liu, Xin Tong, Alla Sheffer, and Wenping Wang. Robust flow-guided neural prediction for sketch-based freeform surface modeling. *ACM Transactions on Graphics (TOG)*, 37(6):1–12, 2018.

[27] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 34(6):1–16, 2015.

[28] Aymen Mir, Thiemo Alldieck, and Gerard Pons-Moll. Learning to transfer texture from clothing images to 3d humans. In *CVPR*, pages 7023–7034, 2020.

[29] Jun Mitani, Hiromasa Suzuki, and Fumihiko Kimura. 3d sketch: sketch-based model reconstruction and rendering. In *International Workshop on Geometric Modelling*, pages 85–98. Springer, 2000.

[30] Ryota Natsume, Shunsuke Saito, Zeng Huang, Weikai Chen, Chongyang Ma, Hao Li, and Shigeo Morishima. Siclope: Silhouette-based clothed people. In *CVPR*, pages 4480–4490, 2019.

[31] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3d hands, face, and body from a single image. In *CVPR*, 2019.

[32] Gerard Pons-Moll, Sergi Pujades, Sonny Hu, and Michael J Black. Clothcap: Seamless 4d clothing capture and retargeting. *ACM Transactions on Graphics (TOG)*, 36(4):1–15, 2017.

[33] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6), Nov. 2017.

[34] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *CVPR*, pages 2304–2314, 2019.

[35] Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *CVPR*, pages 84–93, 2020.

[36] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *NeurIPS*, pages 3483–3491, 2015.

[37] Fang Wang, Le Kang, and Yi Li. Sketch-based 3d shape retrieval using convolutional neural networks. In *CVPR*, pages 1875–1883, 2015.

[38] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, pages 8798–8807, 2018.

[39] Chung-Yi Weng, Brian Curless, and Ira Kemelmacher. Photo wake-up: 3d character animation from a single photo. pages 5901–5910, 06 2019.

[40] Jun Xu, Yingkun Hou, Dongwei Ren, Li Liu, Fan Zhu, Mengyang Yu, Haoqian Wang, and Ling Shao. Star: A structure and texture aware retinex model. *IEEE Transactions on Image Processing*, 29:5022–5037, 2020.

[41] Yuanlu Xu, Song-Chun Zhu, and Tony Tung. Denserac: Joint 3d pose and shape estimation by dense render-and-compare. In *ICCV*, pages 7760–7770, 2019.

[42] Tiancheng Zhi, Christoph Lassner, Tony Tung, Carsten Stoll, Srinivasa G Narasimhan, and Minh Vo. Texmesh: Reconstructing detailed human texture and geometry from rgb-d video. In *ECCV*, pages 492–509. Springer, 2020.

[43] Fan Zhu, Jin Xie, and Yi Fang. Learning cross-domain neural networks for sketch-based 3d shape retrieval. In *AAAI*, 2016.

[44] Heming Zhu, Yu Cao, Hang Jin, Weikai Chen, Dong Du, Zhangye Wang, Shuguang Cui, and Xiaoguang Han. Deep fashion3d: A dataset and benchmark for 3d garment reconstruction from single images. *arXiv preprint arXiv:2003.12753*, 2020.