Pacific Graphics (2022)
Y. Yang, A. D. Parakkat, B. Deng, and S.T. Noh (Editors)

Poster

Human Face Modeling based on Deep Learning through Line-drawing

Y. Kawanaka[1] S. Sato[2,4] K. Sakurai[3] S. Gao[1] and Z. Tang[1]

[1]University of Toyama, Japan [2]Hosei University, Japan [3]DWANGO Co., LTd., Japan [4]Prometech CG research, Japan

Abstract

This paper presents a deep learning-based method for creating 3D human face models. In recent years, several sketch-based shape modeling methods have been proposed. These methods allow the user to easily model various shapes containing animal, building, vehicle, and so on. However, a few methods have been proposed for human face models. If we can create 3D human face models via line-drawing, models of cartoon or fantasy characters can be easily created. To achieve this, we propose a sketch-based face modeling method. When a single line-drawing image is input to our system, a corresponding 3D face model are generated. Our system is based on a deep learning; many human face models and corresponding images rendered as line-drawing are prepared, and then a network is trained using these datasets. For the network, we use a previous method for reconstructing human bodies from real images, and we propose some extensions to enhance learning accuracy. Several examples are shown to demonstrate usefulness of our system.

CCS Concepts
• ***Computing methodologies*** → ***Mesh models;***

1. Introduction

Recently, demands for 3D human models are increasing due to the attention of a metaverse which is a 3D virtual space. In a metaverse, human models are used as avatars which represents hes/her-self. A copy of hes/her-self is often used for the avatar, but other types are also used, for examples, cartoon or fantasy characters. However, creating 3D models are quite difficult for non-expert users because 3D computer graphics software (3DCG software) have complex user-interface and professional knowledge is required to master them. A system for making characters might be implemented in some video games, but variation of characters is limited.

To easily build human body/face models, many reconstruction methods from photographs have been proposed [SHN*19]. However, these methods do not consider cartoon or fantasy characters. In contrast, several sketch-based human face modeling methods have been developed [HGY17]. These methods create face models like cartoon via line-drawing. However, a resultant face model is often partially different from an input line-drawing.

For resolving this problem, we propose a deep learning-based method to achieve accurate modeling of 3D human face via line-drawing. We prepare many 3D human face models and corresponding rendered images as dataset. These images are rendered as line-drawing using a method for non-photorealistic rendering. Our network is based on Saito's method which reconstructs 3D models of a human body from real images [SHN*19], and we extend the method to enhance accuracy for our case using line-drawing as an input. Our method can easily create a human face model only to input a single line-drawing of a face. We demonstrate the usefulness of our method across various examples containing Fig. 1.

Figure 1: *Our result. The inset image is an input line-drawing.*

2. Our Method

We extend Saito's method [SHN*19] for generating a human face model from a line-drawing of a face.

We prepare datasets which contain various 3D human face models and corresponding images rendered as line-drawing. Face models are created using 3DCG software, and the number of datasets

DOI: 10.2312/pg.20221239

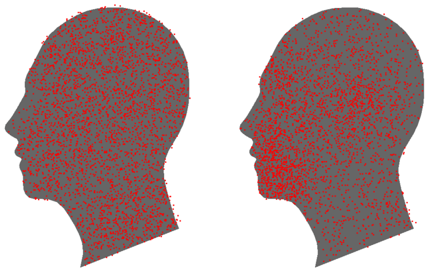

Figure 2: *Comparison with a uniform sampling (left) and our adaptive sampling (right). Red dots indicate sampling points.*

are increased by varying a size of these models. Non-photorealistic rendering is also performed on the software to obtain line-drawings corresponding to each face model. These line-drawings are rendered at multiple viewpoints: viewpoints are rotated centering around the model one by one degree from 0 to 360 degrees. Furthermore, we render the line-drawings with various line widths. By doing this, our system becomes robust for a line width.

The same as Saito's method, our network consists of a convolutional neural network (CNN) and a multilayer perceptron (MLP). The CNN part extracts features of an input image. Using the features, the MLP part predicts a implicit function which represents a 3D face model corresponding to the input image. This implicit function is defined on a 3D space whose range of value is $[0,1]$, and it defines the iso-surface, out of the surface, and inside the model with 0.5, $[0,0.5)$, and $(0.5,1]$, respectively. At an learning phase, these two networks are learned by minimizing a loss function defined as differences between predicted values and correct values corresponding to the input. To enhance a learning accuracy, we adaptively change the number of sample points, which are used for evaluating a loss function, based on curvatures of meshes of a face model. Concretely, each mesh is divided into two groups according to it's curvature value: many/a few sample points are placed around meshes with a large/small curvature. A threshold of the curvature value is manually set by the users. Fig. 2 shows a comparison with a previous uniform sampling and our adaptive sampling. Sample points concentrate around a jaw, mouth, and eyes which have a high curvature value.

3. Experimental Results

This section shows some results created using our method. In the experiment, the number of epochs is set to 8, and we prepare 100 face models for training and 4 models for testing. The rendered images are prepared 144000 and 5760 for training and testing, respectively. We use Adam optimizer for training our network.

First, we show a result using datasets containing multiple line-drawings with a different line width. An inset image in the left one in Fig. 3 is an input line-drawings. This type of line widths is not contained in both datasets. When a single type of line widths is used, a result cannot accurately generate (a left image in Fig. 3). In contrast, when multiple types of line widths are used, a resultant face model is successfully generated (a right image in Fig. 3).

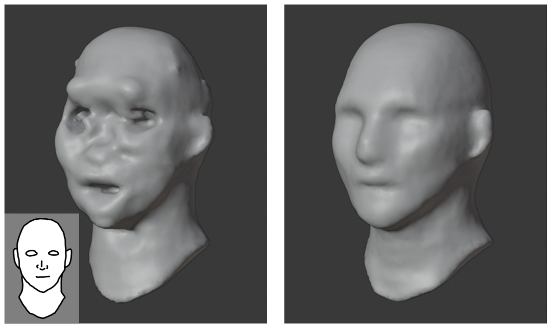

Figure 3: *A comparison with datasets containing line-drawings with multiple types of line widths. The left uses a single type, the right uses 4 types.*

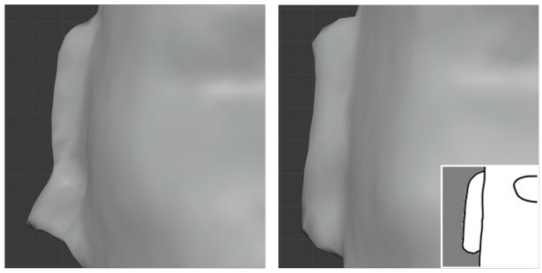

Figure 4: *A comparison with different sampling approaches: a uniform sampling (left) and our adaptive sampling (right).*

Next, a comparison with different sampling methods is shown in Fig. 4. The total number of sample points are the same in both results. A result with a uniform sampling is not similar with the input line-drawing (the inset image in the figure), but our adaptive sampling can generate the similar shape.

Finally, Fig. 1 shows a result obtained using our method. Our system can successfully obtain the 3D face models similar with the input. Other results are contained in our supplementary material.

4. Conclusions

We have proposed the deep learning-based method for modeling 3D human face via a line-drawing. Our network is based on Saito et al. [SHN*19], and we extended this method to enhance a learning accuracy in our case using a line-drawing as an input.

Generating hairs using our method is difficult because these are often represented using particles instead of meshes. We are planning to develop a method to generate hairs via a line-drawings.

References

[HGY17] HAN X., GAO C., YU Y.: Deepsketch2face: A deep learning based sketching system for 3d face and caricature modeling. *ACM Trans. Graph. 36*, 4 (2017). 1

[SHN*19] SAITO S., HUANG Z., NATSUME R., MORISHIMA S., KANAZAWA A., LI H.: Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *The IEEE International Conference on Computer Vision* (2019), vol. 25, pp. 624–630. 1, 2