Donghan Kim (김동한)
Media Lab, Korea University

3D Human Texture Inference Report

https://github.com/donghankim/3D-Texture-Inference

This report contains a detailed summary regarding my efforts in applying texture inference on human 3D models.

Overview:

Multiple efforts have been made in the field of human 3D reconstruction. In particular, Facebook FAIR's research *Tex2shape* [https://github.com/thmoa/tex2shape], *Densepose* [https://github.com/facebookresearch/DensePose] and USC's research *PIFu* [https://github.com/shunsukesaito/PIFu] are all attempts in generating a 3D human model from a single RGB image. My aim was to continue the development made by Tex2shape, by adding RGB values (texturing) to the surface of the outputted human 3D model.

Texture Inference Network (PIFu):

The term texture inference was introduced in the PIFu paper. The idea was to train a deep neural network to inference texture (RGB) values directly from the 3D surface (as opposed to using a 2D parameterization of the surface). A convolutional neural network was trained using the DeepFashion and RenderPeople dataset where the inputs were the individual vertices and a depth value denoted by z. While the contribution of this paper is impressive, I found the results to be a little lacking for general use.

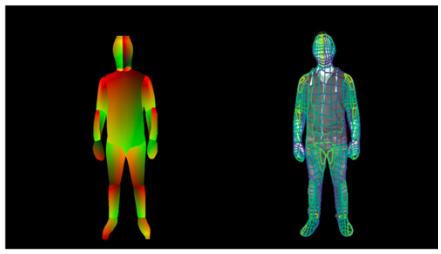
As a result, the main idea behind this project was to use the texture inference model and apply it to the output of the Tex2shape output. The Tex2shape model produces a more accurate and a higher resolution 3D human model, but without any texture. Because the texture inference network used in PIFu does not rely on a proprietary 3D surface, applying the PIFu texture inference network directly onto the Tex2shape 3D surface will not be an issue.

Donghan Kim (김동한)
Media Lab, Korea University

Method:

Input Image

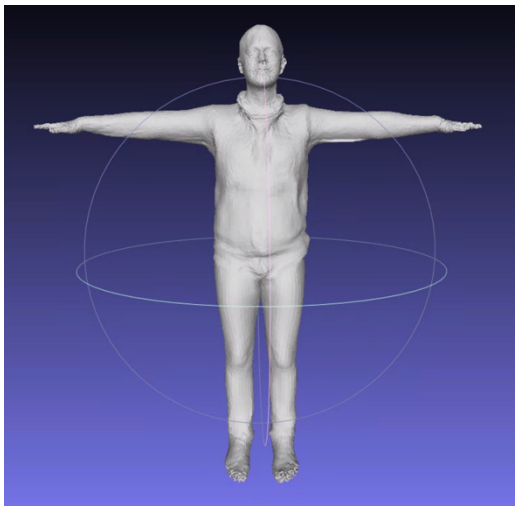
Background removed image
& mask
(for PIFu)



Densepose Output
(for tex2shape)

In order to run the Tex2shape algorithm, we first have to apply DensePose to our input image to obtain the corresponding surface UV maps (for more detail refer to the DensePose paper). Using the outputs from DensePose, we can obtain our 3D parametrized human model. From here, for every vertex in our 3D model, we apply the texture inference network to obtain the texture for the entire surface of our Tex2shape model (please keep in mind that the texture inference network was not additionally trained for this project).

Result:



The results shown are poor to say the least. Surfaces of the human body where the input image does not cover (in this case the arms) are not even queried. In addition, areas of the surface that are shown in the input image are not accurately textured. Although the overall color of the texture is correctly outputted, this is not nearly good enough to compare with the outputs of PIFu.

Potential Issues:
One possible issue is the fact that the output from the Tex2shape model takes a different form from that of the input image. All outputs of Tex2shape come out with the arms spread out as shown above. This will make it difficult for the PIFu texture inference network to accurately produce RBG values since the 3D surface model produced from PIFu follows the same shape as the human in the input image.

Another issue is training. Although deep learning has shown a lot of potential in the past, we cannot expect the texture inference network to accurately infer RGB values without any additional training. However, free or open-sourced datasets that are currently available are not adequate for this task. Not only do we need high resolution 3D human models in a variety of different poses and clothing, but we also need their corresponding 2D RGB image. Therefore, acquiring the right dataset can be cumbersome or expensive.

Potential Improvements:
Addressing the first potential issue, because Tex2shape uses Denspose for UV parameterization of the human body, it should be possible to output the human 3D model following the pose of the input image. If the output 3D model takes the same pose as the input 2D image, then applying the texture inference network on the Tex2shape output should produce better results.

In regard to the second issue (dataset) there really isn't an easy solution. 3D human data is not as readily available as 2D images such as ImageNet. RenderPeople does provide a large dataset of 3D human models (and their corresponding 2D image), however, it is a paid service and the cost of acquiring this data may be expensive. I have tried looking at the BUFF dataset provided by the Max-Planc Institue, but the subjects are limited and don't come with any 2D images.