3D Mesh Generation and Landmark Transfer

The art of producing caricatures from real faces is as human of an activity as any art. It requires a delicate eye for distinguishing the particular characteristics that make a certain person's face appear unique. The caricature artist must then produce a work that exemplifies these characteristics in a way that exaggerates, yet still allows the photo to still be identifiable to the human eye. A standard caricature, however, implies a two-dimensional work of art. However, a three-dimensional caricature, or perhaps an exaggerated model of the original face, would then be exponentially more difficult, likely impossible, for a human artist. To then produce such a complex and human work of art programmatically would thus only be truly capable via machine learning or some other form of artificial intelligence. Using advanced neural networks, we will show that it is possible to produce a three-dimensional caricature from a flat, still image of a human face.

First, however, we must start with some theory. As noted above, a caricature represents the unique features of a person. It is obvious to the human eye that a caricature looks distinctly like the human it was drawn to resemble. But what exactly is it about a caricature that communicates these traits? "It has been claimed that humans recognize faces by the amount they differ from the norm and that by exaggerating these perceived differences we can create a caricature" (Noble 2). Susan Brennan at Cornell University's Department of Psychology first described this phenomenon in her 1985 publication "Caricature Generator" and referred to it as the *mean face assumption*. The idea of the mean face assumption is the basis with which we understand and identify not only the unique features of a human face but the exaggerations of that human face's caricature. Thereby, it should be understood that to create a caricature of a face, we must first identify the face's unique features by *deviation* from the utmost average human face.

Figure 1: Caricature of Abe Lincoln

Attaining this average face is no small task. Because we unfortunately do not have any concrete evidence to what exactly the most average human face would be, it must be generated. This is outside the scope of this project, but it luckily has already been accomplished by a number of academic institutions. The most effective way to produce an "average" face would be to very literally average many scans of human faces, optimally of many varying appearances. One such institution, the Max-Plank Institute for Biological Cybernetics, has openly released an averaged face in their "Face Database" of which is an averaged face from 200 laser-scanned heads at 7 different views. While they provide options for gender segregation, for the purposes of this project, we will only use the general human average face.

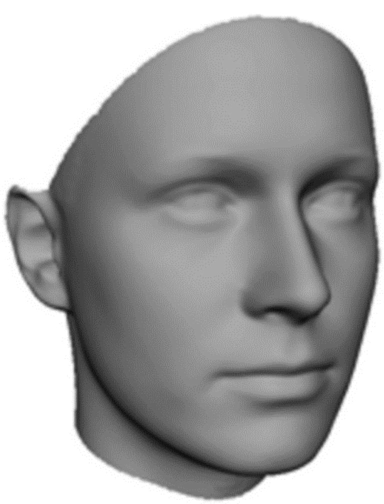


Figure 2: Averaged Faced Mesh

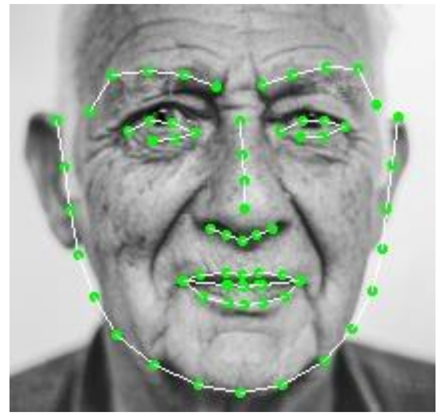


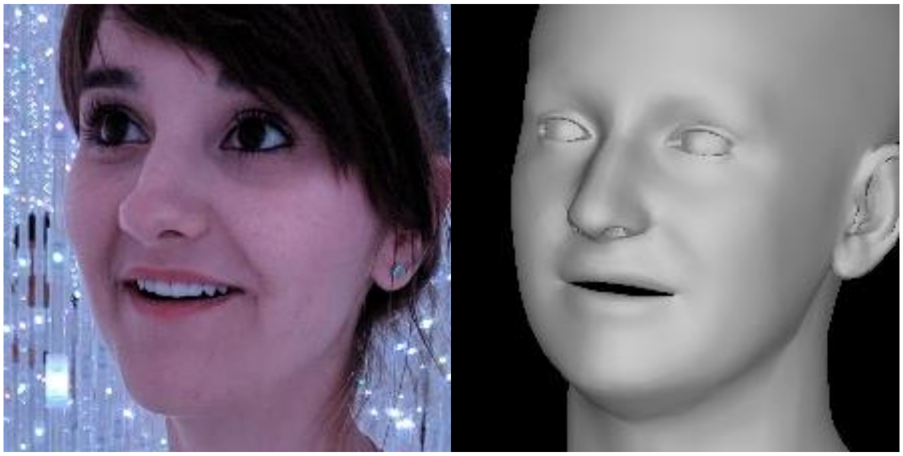
Figure 3: Landmarks laid on face

Next, the concept of "landmarks" must be elucidated. One of the most crucial aspects of facial recognition, motion capture, and many other technologies relies on the basis of identifying key points on a subject's face. By tracing these so called "landmarks," it is possible to both programattically detect expressions or changes, or to use them to morph reform the face. The example given in figure 3 shows landmarks placed on a two-dimensional photo. Landmarks are not limited to the second dimension, however. They may also be placed on "meshes" or three-dimensional polyhedrons used in modelling and rendering. Taking advantage of landmarks laid on a mesh allows for an animator, digital artist, programmer, etc. to "push and pull" the landmarks and the associated mesh's features to the desired result. They may also be used for judging shape, size, distance of the vector, etc. on the mesh. For the purposes of this project, we will use landmarks to judge deviation from the average and reform the mesh.

Using the knowledge discussed above, we will attempt to generate a three-dimensional caricature from a flat, still image of a real human face. The step necessary to accomplish this are as follows: a) generate a base mesh from the flat image, b) place landmarks on the mesh, c) calculate the delta of the landmark vectors from the average, d) exaggerate the landmark vector distance by some given percentage, and e) reform the mesh to the newly acquired landmarks for a final three-dimensional caricature. The percentage used is user-defined; the greater the exaggeration, the more a like a caricature the result is. However, there is an upper limit as once a caricature's features are too far exaggerated, it begins to appear inhuman to the human eye. An example of this spectrum is given in figure 4.

Figure 4: Spectrum of feature exaggeration

To generate a base mesh, we will use existing research. As discussed previously, generating a three-dimensional mesh from a flat image is outside the scope of this project. One recent and very successful learned neural network is the *DECA: Detail Expression Capture Animation* which we will use for producing the base mesh. The model is built on the *FLAME*



library for three-dimensional face mesh creation, of which is developed using the famous *Torch* tensor library. Using this model, we are able to create a starting point to work from. Figure 5 shows an example of one such mesh.

Figure 5: Generated three-dimensional mesh

Next, landmarks must be placed on the mesh for the purposes of assessing deviation and reforming. The benefit of using the *FLAME* library is that it also supports landmark predictors. Using both the mesh and the flat image itself, we can automatically lay landmarks on the mesh itself or even the flat image if necessary. Notice that in figure 6, the landmarks follow the contours of the subject's key identification characteristics such as jawline, the shape of the eyebrows and the shape of the bridge of the nose. These are all necessary points that must be exaggerated to produce a believable caricature.

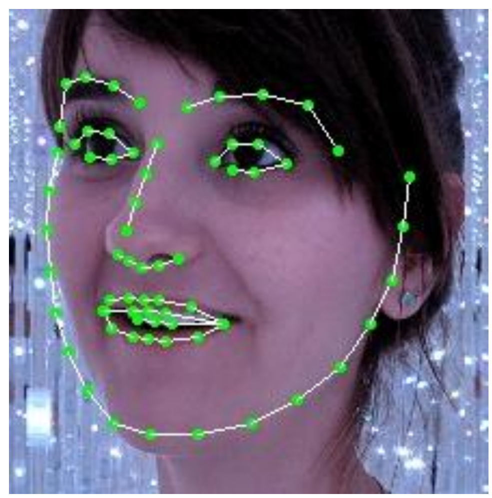


Figure 6: Generated landmarks

Finally, we will replace the landmarks and reform the mesh to produce an altered mesh. This will be done to exhibit the concept and proof of landmark transfer. Here, the reforming of the mesh may also be accomplished using again the *FLAME* neural network. The beginning of the process involves generating a mesh of a different subject first. Using this mesh, we repeat the same process to produce three-dimensional landmarks on the mesh. Once we have landmarks for the secondary subject, we can simply replace them on the primary subject. The result is a mesh with ill alligned landmarks. We can,

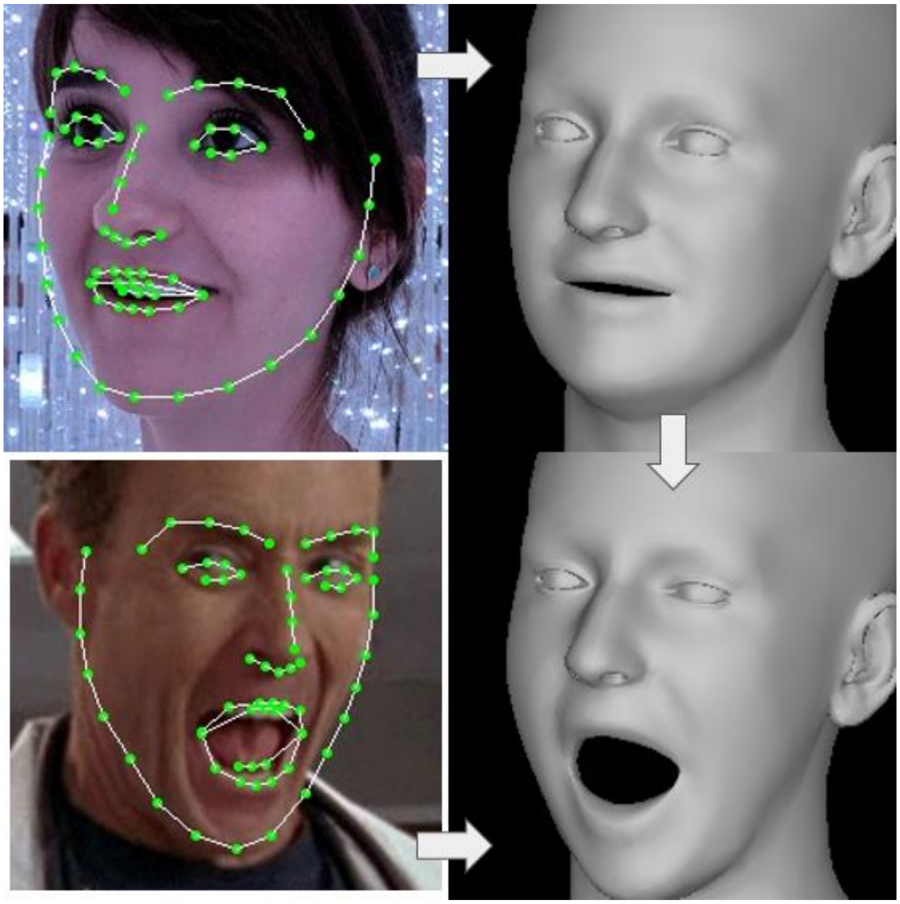


Figure 7: Landmark transfer and mesh reform

however, reform the mesh to follow the new landmark to produce a newly morphed mesh. This can be seen in figure 7.

Following this research and proof of concept, I would develop a deviation metric algorithm for determining the greatest indicators for a particular person's characteristics. To determine how far the landmarks deviate from a "normal," The Max-Plank Institute's average face as discussed previously would be used as a baseline and a difference in the vectors for the landmarks could be calculated. Using the delta in this landmark vector, along with a user-defined exaggeration percentage, a new set of landmarks may be produced with which contain the exaggerated characteristics of the subject. Some empirical data would be collected on varying percentages to determine the most effective metric to produce a caricature: too little and the face simply looks irregular, too great and the face begins to look inhuman.

References

Brennan, S.E.: Caricature generator: The dynamic exaggeration of faces by computer. Leonardo 18(3), 170–178 (1985)

Feng, Yao et al. "Learning an Animatable Detailed 3D Face Model from In-The-Wild Images."

Li, Tianye et al. "Learning a model of facial shape and expression from 4D scans". ACM Transactions on Graphics, (Proc. SIGGRAPH Asia) 36. 6(2017).

Paul Noble, et al. "Multi-level 3D caricatures with relative feature displacement and head shapes." Proceedings of Computer Graphics International 2011.

Sanyal, Soubhik et al. "Learning to Regress 3D Face Shape and Expression from an Image without 3D Supervision." Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).