Examining the Accuracy of Stereophotogrammetry for Generating 3D Face Models

Eric Sandoval Ruezga, *University of California, Santa Cruz*

Abstract—This paper presents a pipeline that produces meshes using photogrammetry and state-of-the-art methods such as machine learning to generate meshes given limited input photos and uses a 3D morphable face model to recover missing sections. The goal is to use this pipeline to analyze and visually compare the error between these meshes to a ground truth mesh using different uncertainty visualization methods given images of faces from a dataset of high-resolution images of faces taken from different angles and photos taken from different angles everyday phone cameras.

Index Terms—3D morphable face model, stereophotogrammetry, uncertainty visualization, machine learning

1 Introduction

STEREOPHOTOGRAMMETRY is a classic method for generating meshes given a set of photos of an object taken at different angles and is an alternative to laser scanning. Results are very accurate when using enough photos taken at a variety of angles with high resolutions cameras, but this is not required to create believable digital replications of everyday objects (or people). Commercial products such as Agisoft Metashape provide a streamlined photogrammetry pipeline to professionals and consumers. From preliminary work using this software, it is clear that stereophotogrammetry is accurate, although imperfect, at generating meshes of faces using just three photographs of a face...

2 Motivation

Having accurate 3D models of a face is helpful in the medical field to plan procedures, such as dentistry, or fit medical devices to a person's face, such as masks used for CPAP machines, a type of ventilator used to treat sleep apnea. Similarly, it is helpful to personalize products such as glasses to a person's face and head. Given these use cases, it is helpful to compare stereophotogrammetry to newer techniques so that researchers and developers can make better decisions on methods to generate 3D face models.

3 Related Work

The FaceScape research project provides a dataset of high-resolution images of faces taken from several angles of many volunteers. The pipeline presented in this paper will use the FaceScape dataset to generate meshes [1]. The pipeline will also use the basic idea of the morphable model initially preset by Volker Blanz and Thomas Vetter in the paper "A Morphable Model For The Synthesis Of 3D Faces." Given a set of 3D faces, the model uses a linear combination to generate a new face [2]. The morphable model in this paper will use the FaceScape dataset to fill in the gaps of meshes generated using stereophotogrammetry and other methods. Lodha *et al.*, in "Visualizing Geometric Uncertainty of Surface Interpolants," summarize and analyze several visualization techniques [3]. A few of the techniques are used in the paper to help analyze the accuracy of the different methods for generating 3D faces.

4 Proposed Research Directions

The first step of this paper is to develop a pipeline for generating the final meshes to analyze. Firstly using stereophotogrammetry, machine learning, and other current methods, a rough mesh will be generated. This mesh needs to be aligned to the mean face of the morphable model that is based on the FaceScape dataset, and then the correspondence between points can be found. It is then an optimization problem to use the morphable model to predict the rough mesh. This predicted mesh is used to fill in the gaps to finally recover a complete mesh of a person's head. This mesh will be compared to the ground truth through different methods of uncertainty visualization. (not sure how relevant this section is for the checkpoint)

5 Methods

5.1 Dataset

As mentioned earlier, this paper uses the FaceScape dataset. It includes the set of images taken of the volunteers, which serve as input to the different methods compared in this paper to generate meshes. Also, the dataset includes 3D meshes generated from these images, which are used as the ground truth for the presented comparisons [1].

5.2 Prediction Methods

The primary method used here for predicting a mesh from a set of photos is stereophotogrammetry. The specific software used for this is Agisoft Metashape Pro which provides a photogrammetry pipeline accessible through a graphical user interface.

(still need to talk about other methods)

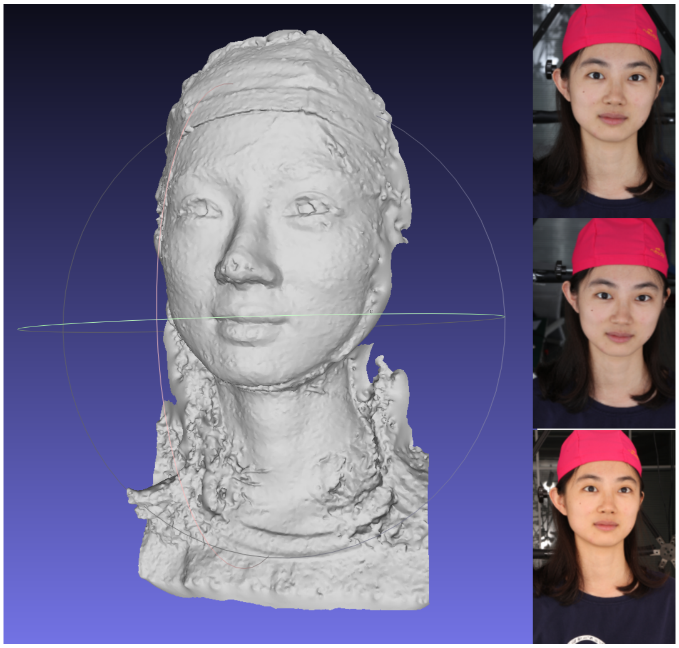

Fig. 1. A mesh that was generated with Agisoft Metashape Pro using only three images as input.

5.3 Pipeline

The software pipeline used in this paper is broken up into three stages. The first stage involves generating a mesh using stereophotogrammetry (or some other machine learning or state-of-the-art method). The next stage recovers a denoised mesh that's in one-to-one point correspondence to the ground-truth mesh. The last stage involves calculating the mean error as well as creating a comparative visualization.

5.3.1 Prediction

At least two images of a person's face need to be used when generating a mesh using Metashape Pro. The maximum settings that this software supports are generally used through this photogrammetry pipeline, which means matching photos using the highest accuracy level and building depth maps using the highest quality level with aggressive filtering.

(other methods to be added)

5.3.2 Recovery

For this pipeline step, it is necessary to recover a denoised mesh. This is done by first taking the dataset of meshes, which are all in correspondence, reshaping them into column vectors, and concatenating them into a matrix. PCA of this matrix is then performed by computing the SVD. The predicted mesh then needs to be aligned through ICP and put in correspondence to the mean face of this matrix. It can then be multiplied by the inverse of the principal components, which gives a vector of weights. They are multiplied by the principal components to generate a smoother mesh that is also in one-to-one correspondence to the ground truth mesh from the database.

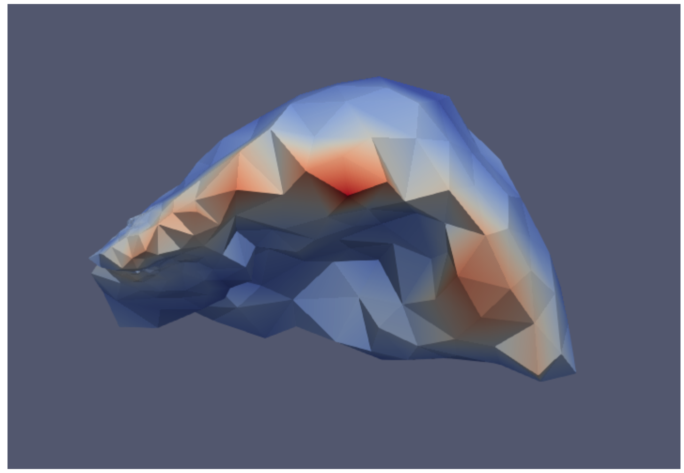

Fig. 2. This is an example heatmap visualization between a predicted and ground-truth mesh.

5.3.3 Error Analysis

The mean error between the two meshes is calculated by comparing every point in the ground truth to the closest point of the recovered mesh. Alternatively, it can also be computed between corresponding points (should also add something about L1 and L2 norm). Paraview, an application for interactive scientific visualization software, is used to analyze further the error between the two meshes through comparative visualizations that use displacement glyphs, pseudo-coloring, and difference surfaces.

6 (Preliminary) Results

See figure 2 for a preliminary visualization.

7 Conclusions

Acknowledgments

References

[1] H. Yang, H. Zhu, Y. Wang, M. Huang, Q. Shen, R. Yang, and X. Cao, "FaceScape: A Large-Scale High Quality 3D Face Dataset and Detailed Riggable 3D Face Prediction," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[2] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," *qroceedings of the 26th annual conference on Computer graphics and interactive techniques* - SIGGRAPH 99, 1999.

[3] S. Lodha, R. Sheehan, A. Pang, en C. Wittenbrink, "Visualizing Geometric Uncertainty of Surface Interpolants", in *Proceedings of the Graphics Interface 1996 Conference, May 22-24, 1996, Toronto, Ontario, Canada*, 1996, bll 238–245.