

Cost-efficient 3D Face Reconstruction from a Single 2D Image

Juseung Yun*, Jaeyoung Lee*, Dongyoon Han*, Jeongwoo Ju*, Junmo Kim*

*School of Electrical Engineering, KAIST (Korea Advanced Institute of Science and Technology), South Korea
st24hour@kaist.ac.kr, hellojy@kaist.ac.kr, dyhan@kaist.ac.kr, veryju@kaist.ac.kr, junmo.kim@kaist.ac.kr

Abstract—We propose a three-dimensional (3D) face-modelling method from a single two-dimensional (2D) face image using a gallery of 2D face images and their corresponding 3D face models. Unlike existing methods, which require human effort, we provide a simple way to reconstruct 3D face models without user interaction. Our main approach is based on the idea that a particular coefficient that linearly combines vectors of 2D face images and outputs a vector that approximates the input image vector in terms of the vector norm can be reused in 3D models. Therefore, the pair of a 2D image and its 3D model plays an important role in our algorithm. Using the FaceGen software allows us to avoid the employed in previous works procedure whereby the 3D model is generated in a labor-intensive, expensive way. As a result, we are able to easily establish our 2D and 3D database. In this paper, we present a method for adopting the coefficients in 3D models and demonstrate the results of our algorithm.

Keywords—Facial modelling, 3D face reconstruction, 2d face fitting, face landmark matching, linear combination

I. INTRODUCTION

The possibility of converting a two-dimensional (2D) face image into a three-dimensional (3D) face model has attracted attention in various research fields, such as 3D animation, augmented reality, the game industry, and graphics. Most face-modeling techniques have relied on manual methods because they were considered the most accurate way to model a realistic 3D face. These manual methods typically require drawing outlines of a face or setting landmarks on a face. However, these user-interactive methods have a crucial drawback, since they require a lot more time than automatic methods do. Therefore, some researchers have sought to develop automatic methods for this procedure.

A method called photometric stereo is frequently used to obtain 3D face models [1]; this estimates the surface normal of an object by observing it under different lighting conditions, considering the number of point sources, intensity, direction, diffusion, and so on. However, photometric stereo has several limitations. For instance, it requires multiple images of each lighting condition and only works with specific constraints

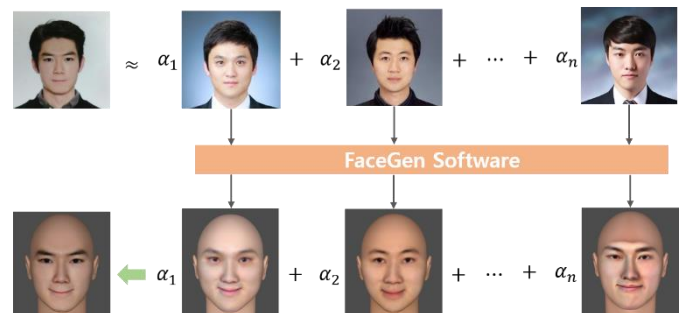


Figure 1. Overall process of three-dimensional (3D) modelling from a two-dimensional (2D) image. The upper row shows that a 2D face can be approximated by the linear combination of other faces. The arrows in the middle represents process of making a 3D face dataset using FaceGen. The lower row shows the 3D faces that can be reconstructed by the linear combination of weights obtained from 2D faces and the 3D face dataset.

that are far from reality for many types of materials. Another popular approach is a morphable model, which can be formulated with the linear combination of 3D models from a database [3]. These approaches have produced accurate results, but they also entail a complicated structure and demand a large number of calculations and parameters that are not ideal for industries or commercial use.

Our goal is to reconstruct a 3D face model with only a single face image and no user interaction. Compared to the morphable model, which uses shape and texture vectors to combine a 3D model with elaborate calculations [3], we reduce the complexity by reusing the coefficients obtained after morphing in the 2D image space.

II. RELATED WORK

There have been several attempts to find a 3D model corresponding with 2D face images. Two major approaches have been used to reconstruct 3D models without user interaction. First, photometric stereo is widely used to obtain 3D models from 2D images. Second, the set of reconstructions can be learned from a 3D model database.

A. Photometric Stereo

The traditional approach without user interaction is shape from shading (SFS), which represents a special photometric stereo case where the data are a single image. Photometric stereo is a way to find a 3D shape by integrating multiple object images under different lighting conditions. It was introduced by Woodham [1], where it is only available under certain conditions, namely Lambertian reflectance and uniform albedo. The basic method is inverting the following linear equation to estimate the surface normal at each pixel of the images:

$$I = n \cdot L,$$

where I is a vector of m observed intensities, n is the surface normal vector, and L is a known $3 \times m$ matrix of normalized light directions. Because of the lighting conditions requirements, that is, albedo or boundary conditions, various approaches have been introduced, such as estimating lighting directions [4, 5] and treating the number of light sources as one [6].

B. The Morphable Model

In a morphable model, an arbitrary face can be represented as a linear combination of “basis” faces. This idea was first proposed for image compression in telecommunications in 1991 [9]. Here, the researchers treated the morphable model in 2D space only to synthesize a new facial image. Then, Volker Blanz et. al. [3] applied this idea into 3D space to generate a 3D facial model. They defined a shape vector and texture vector in the 3D facial space and generated a prototype 3D model by linearly combining the “basis” 3D face models. They reconstructed the 3D face model from a single image, but they used a complicated manual approach when they attempted to match the identity of the face image with 3D face model.

The major problem with the morphing approach is the correspondence problem, which means the number of vertices of all basis face models should be the same; in addition, the locations of all landmarks of the face should be matched among all basis faces. In [3], the researchers accomplished full correspondence of all basis faces using optic flow.

III. METHOD

In this section, we introduce our new method of reconstructing a corresponding 3D face from an input 2D face. Using a dataset composed of frontal 2D face images, the input face is approximated by weighted sum; following this, the 3D mesh and texture map is reconstructed from our 3D face dataset. A schematic illustration is shown in Figure 1.

A. 2D Image Foreground Masking and Alignment

For 2D faces, both the alignment and background subtraction steps are initially performed on the face images.

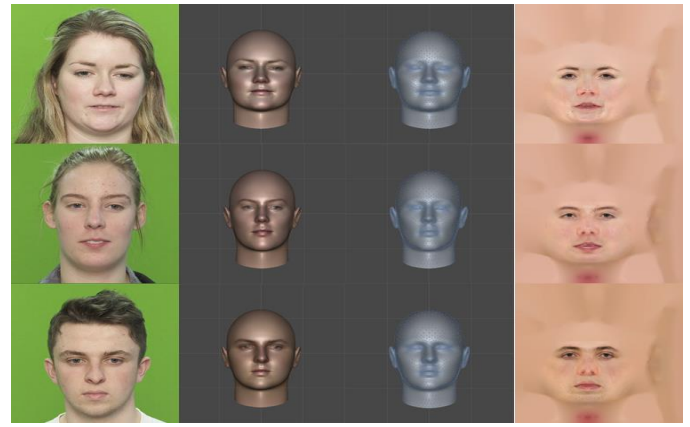


Figure 2. The first column on the left represents the input and the others are outputs of FaceGen: The second column shows 3D face, the third shows 3D meshes, and the fourth shows the texture map. Input images were taken from the TCD-TIMIT dataset [7].

Since face images include both facial parts and noisy parts, background subtraction is necessary. We manually construct the background subtraction mask for foreground masking with the average face and use it for both target face and the faces in our dataset. It should be noted that the average face is created with a facial dataset that only includes facial parts. Then, masked faces are aligned to accurately approximate a facial component of the 2D target face in the following step.

B. Facial Point Extraction and 2D Face Approximation

To approximate the target face, we first extract facial points from the target face. We use the supervised descent method (SDM) [8] to accomplish this. It should be noted that the faces in our dataset have facial points pre-extracted by SDM; this information is cached with the corresponding faces in the dataset.

The target face is then approximated by a linear combination of the faces in the dataset using

$$\mathbf{x} \approx \hat{\mathbf{x}} = \mathbf{D}\boldsymbol{\alpha},$$

where $\mathbf{x} \in \mathbb{R}^m$ is the target face, $\hat{\mathbf{x}} \in \mathbb{R}^m$ is the approximated face, $\mathbf{D} \in \mathbb{R}^{m \times n}$ is a dataset where each column corresponds to a face, $\boldsymbol{\alpha} \in \mathbb{R}^n$ is a weight vector, and n is the number of images in the dataset. We use an L-2 loss regularizer to decay the weights to avoid large elements in $\boldsymbol{\alpha}$ as follows:

$$\operatorname{argmin}_{\boldsymbol{\alpha}} \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_2^2,$$

where λ is a regularization parameter and we use 1. The regularizing term is used to prevent $\boldsymbol{\alpha}$ from becoming too big, because this would mean overfitting had occurred.

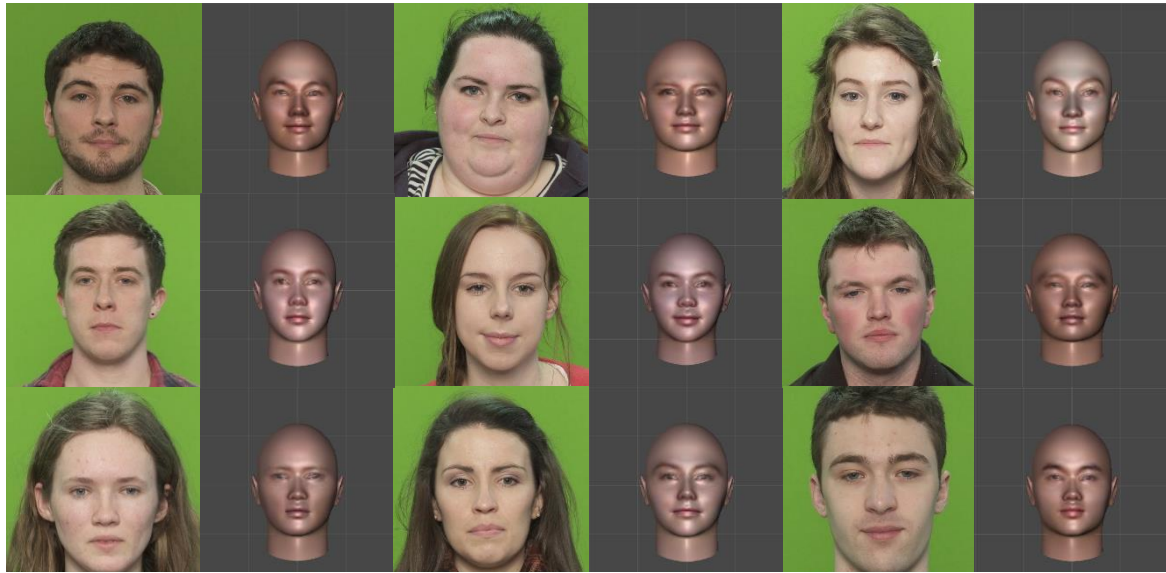


Figure 3. The qualitative results of our 3D face reconstruction algorithm. Without any manual setting, each 3D face is automatically constructed from a single 2D face. Odd columns are 2D facial target images, even columns are reconstructed 3D models from each left column. We can see that the 3D models are effectively reconstructed with details like skin color, facial shape, thickness of lips, and shape of eyebrows preserved.

C. Construction of the 3D Mesh Dataset

Utilizing our face dataset (i.e., a 2D face image dataset), we construct a new 3D mesh dataset to reconstruct 3D faces. We use the FaceGen [10] software program, which is commercial software that takes a single 2D image as input and reconstructs a 3D face composed of a 3D mesh and texture map. However, in FaceGen, manual setting, such as landmark point setting, is a prerequisite to obtain a satisfactory 3D face from an input 2D image. Thus, we utilize FaceGen to construct a new 3D mesh dataset that includes 3D mesh point clouds and texture maps of each corresponding 2D face image. This dataset is employed in our method to reconstruct a 3D face automatically.

With the 3D mesh dataset, we assume that the 3D face to be reconstructed is extremely closely related to given 2D face

image. Thus, the approximation coefficient produced by 2D face approximation can be utilized to reconstruct 3D faces.

D. Reconstruction of the 3D Face

We follow the approach mentioned in the previous subsection, using α to create a 3D face with 3D meshes and texture maps. An input 2D face can be approximated by linear combination of 2D faces in the dataset, as stated above. We can construct 3D meshes and texture maps;

$$\begin{aligned} \mathbf{x}_{\text{mesh}} &\approx \mathbf{M}\alpha \\ \mathbf{x}_{\text{tex}} &\approx \mathbf{T}\alpha. \end{aligned}$$

Algorithm 1 summarizes the overall procedure in our algorithm.

Algorithm 1 3D Face Reconstruction from a Single 2D Image using FaceGen. Here, m is dimension of aligned 2D facial data, n is the number of data.

Data: 2D face images
Output: 3D mesh $\mathbf{x}_{\text{mesh}}$ and 2D texture map $\mathbf{x}_{\text{tex}}$

- 1 Make data matrix $\mathbf{D} \in \mathbb{R}^{m \times n}$
 // by aligning 2D face images and subtracting
 // their backgrounds(Sec. III.A)
- 2 Find α minimizing $\|\mathbf{x} - \mathbf{D}\alpha\|_2^2 + \lambda\|\alpha\|_2^2$ (Sec. III.B)
- 3 Make 3D mesh $\mathbf{M}$ // using FaceGen
- 4 Make 2D texture $\mathbf{T}$ // using FaceGen (Sec. III.C)
- 5 $\mathbf{x}_{\text{mesh}} \approx \mathbf{M}\alpha$
- 6 $\mathbf{x}_{\text{tex}} \approx \mathbf{T}\alpha$
- 7 Mapping $\mathbf{x}_{\text{tex}}$ to $\mathbf{x}_{\text{mesh}}$ (Sec. III.D)

IV. EXPERIMENTS

In this section, we present qualitative results of the reconstructed 3D faces in Figure 3. Target images were taken from the TCD-TIMIT dataset [7], an audio-visual corpus of continuous speech, which consists of high-quality audio and video footage of 62 speakers reading a total of 6,913 phonetically rich sentences. The results showed successful reconstruction of details of the 2D query images, such as eyebrows, lip, skin color, and eye shape. The qualitative results show some limitations; for example, skin troubles like freckles, acne, and so on are not entirely recovered; however, this does not greatly affect the reconstruction performance.

V. CONCLUSIONS

In this work, we developed a simple, cost-efficient method for reconstructing a 3D face model from a 2D single face image without employing user interactions or a laser scanner.

We demonstrated that the proposed method could reconstruct the 3D model using FaceGen and taking a single 2D image as an input. The reconstructed results are plausible for the most of the various 2D images. We observed minor limitations in our algorithm in that it cannot reconstruct skin issues. In future work, we aim to extend this work to generate a video sequence corresponding to audio speech, which has the additional input of an audio speech signal.

ACKNOWLEDGEMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea MSIP. (No.B0101-15-0551, Technology Development of Virtual Creatures with Digital DNA).

REFERENCES

- [1] Woodham, R.J., Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
- [2] Atick, J.J., P.A. Griffin, and A.N. Redlich, Statistical approach to shape from shading: Reconstruction of 3D face surfaces from single 2D images. *Neural Computation*, 8(6):1321–1340, 1996.
- [3] Blanz, V., and T. Vetter, A morphable model for the synthesis of 3D faces. *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co., :187–194, 1999.
- [4] Pentland, A.P., Finding the illuminant direction. *Journal Optical, Society of America*, 72(4):448–455, 1982.
- [5] Hernandez, C., G. Vogiatzis, and R. Cipolla, Multi-view photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):548–554, 2008.
- [6] Rama Chandran, V., Perception of shape from shading. *Nature*, 331:163–166, 1988.
- [7] Harte, N., and E. Gillen, TCD-TIMIT: An audio-visual corpus of continuous speech. *IEEE Transactions on Multimedia*, 17(5):603–615, 2015.
- [8] Xiong, X., and F. de la Torre, Supervised descent method and its application to face alignment. *Proceedings of the IEEE conference on computer vision and pattern recognition*, :532–539, 2013.
- [9] Choi, C.S. et al. A system of analyzing and synthesizing facial images. *IEEE International Symposium on Circuits and Systems*. IEEE, 1991.
- [10] Inversions, Singular. FaceGen Modeller, Version 3.5. (2010).



Dongyoon Han received his BS and MS degrees in electrical engineering in 2011 and 2013, respectively, from KAIST, Daejeon, Rep. of Korea. He is a PhD student in the School of Electrical Engineering at KAIST. His research interests include machine learning and data mining.



Jeongwoo Ju received his BS degree in aerospace engineering from Chungnam National University, Daejeon, Rep. of Korea in 2015. He received his MS degree at KAIST, Daejeon, Rep. of Korea. He is a PhD student at KAIST. His research interests include pattern recognition, machine learning, and computer vision.



Junmo Kim received his BS degree from the University of Seoul, Seoul, Rep. of Korea in 1998. He received his MS and PhD degrees in 2000 and 2005, respectively, from the Massachusetts Institute of Technology, Cambridge, US. He won the Gold Prize at the National Mathematics Competition in 1993, served as a member of the Korea Foundation for Advanced Studies in 1998, and won the Best Paper Award, Samsung Tech Conference and Signal Processing in both 2007 and 2008. He is an associate professor at KAIST. His research areas are statistical signal processing, image processing, computer vision, and information theory.



Juseung Yun received his BS degree in electrical engineering from Sogang University, Seoul, Rep. of Korea in 2016. His research interests include statistics, machine learning, computer vision, and information theory.



Jaeyoung Lee received his BS degree in electrical and computer engineering from the University of Seoul, Seoul, Rep. of Korea in 2015. His research interests include machine learning, reinforcement learning, computer vision, and information theory.