3-Dimensional Face from a Single Face Image with Various Expressions

Yu-Jin Hong[1,2], Gi Pyo Nam[2], Heeseung Choi[2], Junghyun Cho[2], and Ig-Jae Kim[1,2(✉)]

[1] Department of HCI and Robotics,
University of Science and Technology, Daejeon, Korea
[2] Imaging Media Research Center,
Korea Institute of Science and Technology, Daejeon, Korea
{hyj,keepsl201,hschoi,jhcho,kij}@imrc.kist.re.kr

Abstract. Generating a user-specific 3D face model is useful for a variety of applications, such as facial animation, games or movie industries. Recently, there have been spectacular developments in 3D sensors, however, accurately recovering the 3D shape model from a single image is a major challenge of computer vision and graphics. In this paper, we present a method that can not only acquire a 3D shape from only a single face image but also reconstruct facial expression. To accomplish this, a 3D face database with a variety of identities and facial expressions was restructured as a data array which was decomposed for the acquisition of bilinear models. With this model, we represent facial variances as two kinds of elements: expressions and identities. Then, target face image is fitted to 3D model while estimating its expression and shape parameters. As application example, we transferred expressions to reconstructed 3D models and naturally applied new facial expressions to show the efficiency of the proposed method.

Keywords: 3D face reconstruction · Bilinear models · Facial animation

1 Introduction

The acquisition of 3D face geometry is an important topic in the field of computer graphics and computer vision. Especially, user-specific 3D faces are valuable in industries such as gaming, animation, and film. There are two categories of obtaining 3D facial shapes. The first is to use a 3D scanner, and the second is to use several photographs to model a 3D face. A 3D scanner is the most accurate way to capture a 3D face, and a great deal of low-cost equipment has recently been introduced, increasing its usefulness; however, it is still costly, and the actual modeling target has the burden of travelling to a place where the equipment is located. Therefore, an image-based modeling method is useful in terms of accessibility (Fig. 1).

In this regard, existing image-based modeling methods use multiple photographs to perform 3D modeling, and these methods are still often used in VFX studios. However, there are still many situations in which 3D face reconstruction from only a single photograph is needed. The most outstanding method for 3D face modeling using a

N. Streitz and P. Markopoulos (Eds.): DAPI 2016, LNCS 9749, pp. 202–209, 2016.
DOI: 10.1007/978-3-319-39862-4_19

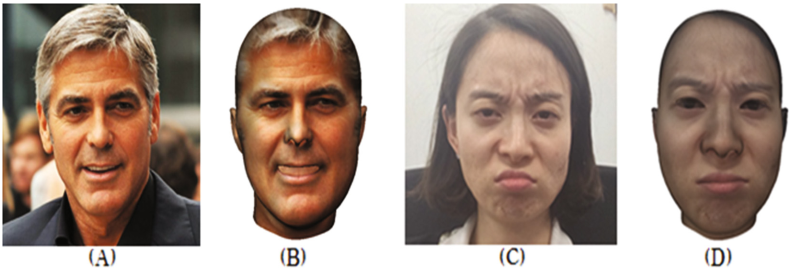

Fig. 1. (A, C) 2D input face images. (B, D) 3D faces reconstructed with the proposed method

single photo is 3DMM [1]. Through statistical techniques based on 3D face scan data from 200 people, this method enables personalized face modeling that mimics the input image. More recently, many excellent researches have introduced to generate a 3D face based on a single photograph. However, there are still limitations for reconstructions of faces showing expressions.

Facial expressions are shown in most actual individual profile photographs, including those of celebrities, who are a major target for 3D modeling. As such, there is a great need for a method that can quickly and accurately reconstruct a 3D face from a single photograph with a facial expression.

To accomplish this, our paper introduces a method that can not only reconstruct a 3D shape from a single face image but also reconstruct various facial expressions from the image. This is achieved through bilinear models, and to acquire models, we used 3D face scans with unique shape and identity structures featuring a variety of expressions and individuals. These models are structured in a data array and HOSVD (Higher Order Singular Vector Decomposition) is applied to construct bilinear models and we used these models to depict facial changes in the 3D faces as expressions and shapes. To show the efficiency of this reconstruction method, we used application example wherein a variety of expressions were transferred to different 3D faces.

2 Contributions

This paper introduces a method for acquiring user-specific 3D face models from face scan data, and its contributions are as follows:

- A practical level of speed (less than 2 s) was achieved to reconstruct 3D face from a single photo, this is accomplished while reducing the position errors between the landmarks on 3D model and the 2D image. An optimized method was used for computational efficiency.
- Through the properties of the bilinear models, the identity of the reconstructed 3D model and expression elements can be controllable and changed into a new appearance. Therefore, the user-specific blendshapes needed to create a person's expression animations can be captured from a single face photograph without 3D scanner equipment.

3 Related Work

3.1 3D Face Reconstruction

Research on acquiring 3D geometry from images has been performed by many excellent researchers. The method that can definitely be considered the best is the 3D Morphable Model [1]. This research used 3D face scan data from 200 people to represent a face as a linear combination of the principal components of shapes and textures, and the results were very realistic. Despite its excellent results, the original 3DMM method is unfortunately not practical. Its computation takes a long time because to reconstruct the face shape, the texture parameters, such as intensity and ambient light, of the input photograph must be calculated in addition to many unknown parameters such as camera information. To resolve these drawbacks, the landmark-based 3DMM method have introduced. This method reduces the position errors of several facial landmarks located on the face photograph and the 3D reference model. As mentioned earlier, reconstruction of the photograph textures is accompanied by complex computations, so the landmark-based method is used for the face shape, and the input texture is directly projected the result 3D face [2–5]. Recently, the novel methods for reconstructing facial shapes from shading information have introduced [6, 7].

However, these studies mostly reconstruct photographs without expressions. To resolve this problem, we present a method that not just reconstructs 3D facial geometry but also facial expressions through bilinear models.

3.2 Bilinear Models

Bilinear models were first introduced by Tenenbaum et al. [8] to separate two factors that are mixed. To separate a combination of various elements, research has been performed on the use of HOSVD (proposed by Tucker [9] and Kroonenberg et al. [10]) to show how these elements influence each other [11]. In our research, we represent our 3D face with bilinear models hence we assume that a facial shape is composed of expression and shape (identity) properties.

4 Overview

This section provides an overall description of how we used the bilinear model to acquire 3D face geometry from a single input photograph. We used a 3D face database of scan data from 150 people with 47 kinds of facial expressions to build our bilinear model. The fitting process reduces the differences between the 3D models and the 2D input images. The final section demonstrates the proposed method's efficiency by showing the example wherein a variety of facial expressions were applied to the resulting 3D faces (Fig. 2).

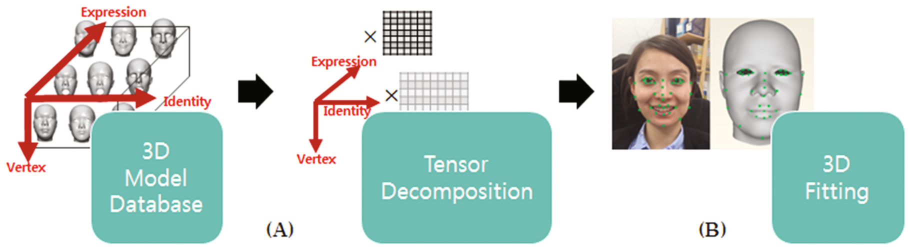

Fig. 2. Overall image of the presented method. (A) The process of building the bilinear models. (B) The process of fitting the input image on the 3D model.

5 Building Bilinear Models

The input photographs included a variety of face shapes and expressions. For this, we used the FaceWarehouse Database [12], which was constructed with 3D face scan models and includes 150 various people and 47 kinds of facial expression models per individual, each made with 11,500 vertices. We divided these models into identity and expression, and we structured a 3D array with T. Figure 3 describes the appearance of T. We then decomposed our array T with SVD to obtain the bilinear models. We wanted to create just a whole facial shape, so we excluded the vertex mode decomposition.

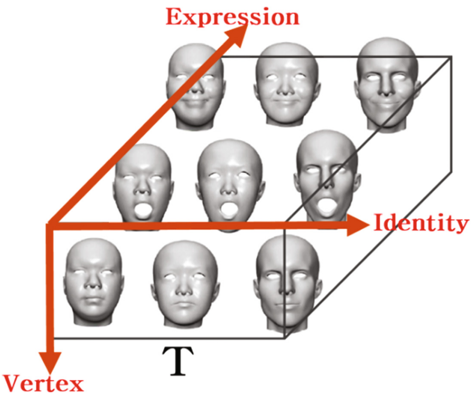

Fig. 3. Appearance of array T composed of expression, identity, and vertex

T can be written again as the following equation.

$$T = C_r x_2 U_{id} x_3 U_{exp} \tag{1}$$

C_r is called the core tensor, and it is in charge of the mutual interaction concerning how the identity and expression affect each other. U_{id} and U_{exp} are the orthonormal transformation matrixes, which represent the left singular vectors of the identity space and expression space, respectively. We call C_r a bilinear model, which can be used to express any expression on any face. This is shown in Eq. (2).

$$V = C_r x_2 w_{id}^T x_3 w_{id}^T \tag{2}$$

where w_{id} is the column vector, which refers to the weight of the identity, and w_{exp} is the column vector, which refers to the weight of the expression. V is the 3D face model, which uses these two weights and core tensor. Figure 4 shows the face model (V) and the core tensor, expression, and identity weights that compose it.

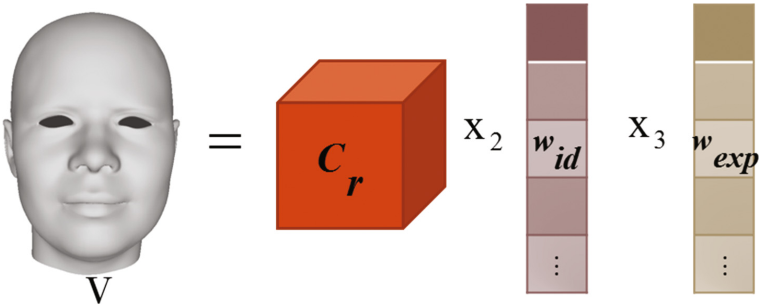

Fig. 4. 3D face V and elements in bilinear format

6 3D Face Reconstruction

As mentioned earlier, our bilinear models can be used to turn a person with any kind of expression into a 3D face. Therefore, we are able to use it to make the 3D model such that it is similar to a given image. To accomplish this, a process is carried out that estimates not only the identity and expression weights of the input photograph but also the camera information.

We initialize the reference 3D model V such that it has average identity and expression. We assume that the camera projection is weak perspective. The 3D model reference point v_k is projected onto the image space point p_k. This is shown in the following equation.

$$p_k = sRv_k + t \tag{3}$$

where s is a scaling information, R is a global rotation matrix, and t is a translation information in the image space. Translation about the Z-axis was ignored due to weak-perspective projection. The following equation shows the process for reconstructing the 3D face while minimizing the error between face landmarks in the 2D image and feature points on the 3D face.

$$\mathrm{Error}_k = \frac{1}{2}\left\| sR(C_r x_2 w_{id}^T x_3 w_{exp}^T)^{(k)} + t - s^{(k)} \right\|^2 \tag{4}$$

where $s^{(k)}$ is the facial features on the 2D image, and k is the kth feature point. We use 76 facial landmarks. The unknown values that we want to find are the pose information of the face (size, rotation, movement) and model parameter (identity and expression) information. To solve these parameters quickly and simultaneously, we used the L-BFGS [13] algorithm for optimization.

7 Results

In Fig. 5, (A) is the input face images, and (B) is the 3D face that results from the proposed method. As shown in Fig. 5(B), faces with a variety of expressions were properly reconstructed into 3D faces.

Figure 6(C) shows the results of various expressions transferred to the 3D model. It can be seen in particular that the expression was changed very naturally. Through this

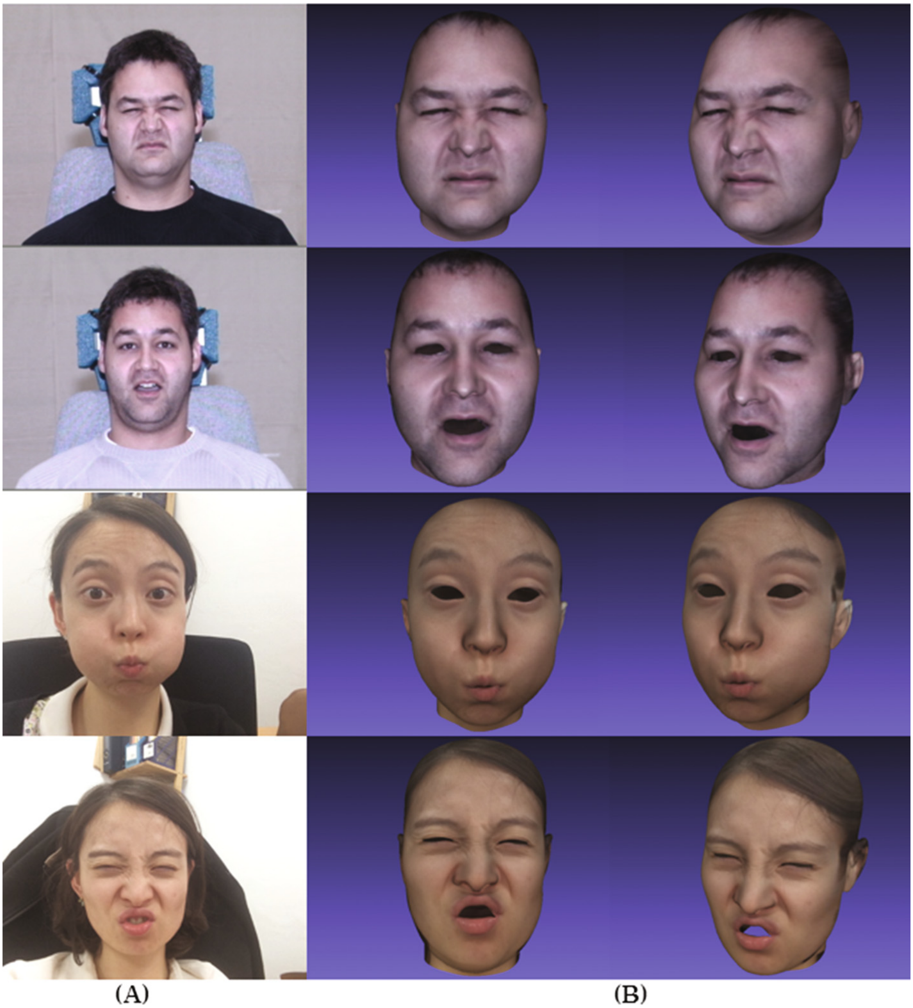

Fig. 5. (A) Input images. (B) 3D model results. The expressions in the input photos are reconstructed in 3D.

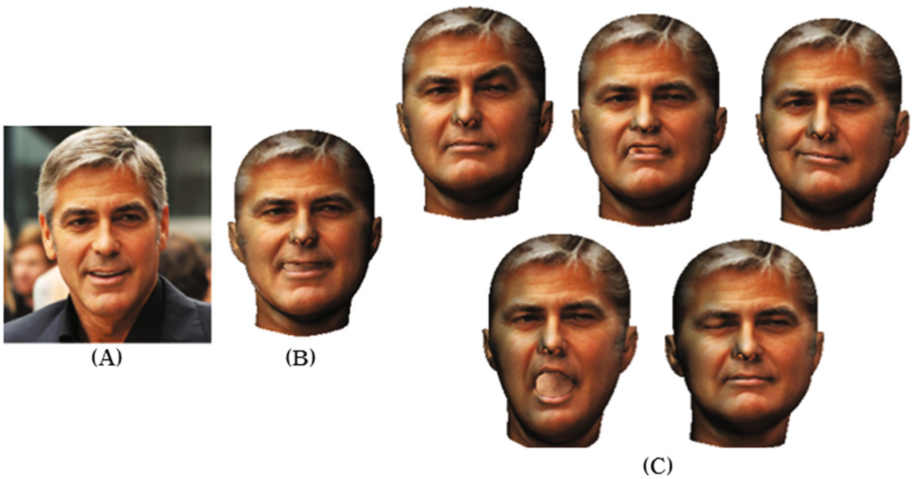

Fig. 6. (A) Input image. (B) 3D model results. (C) Expression transfer model results

method, it becomes easy to obtain the personalized blendshapes needed for making face animations from a single 3D face. Normally, to make facial animation, a set of several or several dozen expressions are made beforehand, from which a wider variety of expressions can be produced; creating such an expression set is an important task. The following formula was used to create the personalized blendshapes from a single photograph.

$$B_i = C_r x_2 w_{id} x_3 (U_{exp} a_i), 0 \leq i \leq 47 \quad (5)$$

where U_{exp} is the expression transformation matrix introduced in Sect. 4. For a_i, the expression element needed as the expression vector is set as 1, and all other vector elements are set as 0.

8 Conclusion

In this paper, we have presented a process for reconstructing a 3D face from a single photograph of a face with an expression. Because normal image-based 3D face reconstruction methods use faces without expressions as their targets, the creation of faces with expressions is difficult. To represent a facial appearance comprising identity and expression elements, we decomposed an array of these two elements and obtained a bilinear model. Using this model, we deduced the camera matrix, expression, and identity information of the input photo and turned it into a 3D face. During this process, we improved our optimization process to solve many unknown values at once so that we could enhance the efficiency of the process in terms of speed.

We also created example of various changes to expressions in reconstructed 3D models to demonstrate the possibility of obtaining personalized blendshapes from a single photograph and using them to create natural expressions for virtual reality and gaming avatars.

Acknowledgements. This work was supported by the KIST Institutional Program (Project No. 2E26450 & No. 2E25930)

References

1. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: ACM SIGGRAPH, pp. 187–194 (1999)
2. Asthana, A., Marks, T.K., Jones, M.J., Tieu, K.H., Rohith, M.V.: Fully automatic pose-invariant face recognition via 3D pose normalization. In: IEEE International Conference on Computer Vision, pp. 937–944 (2011)
3. Ding, C., Xu, C., Tao, D.: Multi-task pose-invariant face recognition. IEEE Trans. Image Process. **24**(3), 980–993 (2015)
4. Ding, L., Ding, X.: Continuous pose normalization for pose-robust face recognition. IEEE Signal Process. Lett. **19**(11), 721–724 (2012)
5. Qu, C., Monari, E., Schuchert, T., Jeyerer, J.: Fast, robust and automatic 3D face model reconstruction from videos. In: Advanced Video and Signal Based Surveillance, pp. 113–118 (2014)
6. Hassner, T.: Viewing real-world faces in 3D. In: IEEE International Conference on Computer Vision (2013)
7. Kemelmacher-Shlizerman, I., Seitz, S.M.: Face reconstruction in the wild. In: IEEE International Conference on Computer Vision (2011)
8. Tenenbaum, J.B., Freeman, W.: Separating style and content with bilinear models. Neural Comput. J. **12**, 1247–1283 (1999)
9. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika **31**(3), 279–311 (1966)
10. Kroonenberg, P.M., Leeuw, J.D.: Principal component analysis of three-mode data by means of alternating least squares algorithms. Psychometrika **45**(1), 69–97 (1980)
11. Vasilescu, M.A.O., Terzopoulos, D.: Multilinear analysis of image ensembles: TensorFaces. In: European Conference on Computer Vision, pp. 447–460 (2002)
12. Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K.: FacewareHouse: a 3D facial expression database for visual computing. IEEE Trans. Visual Comput. Graph. **20**, 413–425 (2014)
13. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Math. Program **45**, 503–528 (1989)