Pattern Recognition 48 (2015) 73–85

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

# Single-view-based 3D facial reconstruction method robust against pose variations

Jaeik Jo [a], Heeseung Choi [b], Ig-Jae Kim [b], Jaihie Kim [a,*,1]

[a] School of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, Republic of Korea
[b] The Image Media Research Center, Korea Institute of Science and Technology, Seoul 136-130, Republic of Korea

ARTICLE INFO

*Article history:*
Received 26 December 2013
Received in revised form
8 July 2014
Accepted 12 July 2014
Available online 1 August 2014

*Keywords:*
3D facial reconstruction
Structure from motion
Morphable model
Single view
Self-occlusion
3D model fitting

ABSTRACT

The 3D Morphable Model (3DMM) and the Structure from Motion (SfM) methods are widely used for 3D facial reconstruction from 2D single-view or multiple-view images. However, model-based methods suffer from disadvantages such as high computational costs and vulnerability to local minima and head pose variations. The SfM-based methods require multiple facial images in various poses. To overcome these disadvantages, we propose a single-view-based 3D facial reconstruction method that is person-specific and robust to pose variations. Our proposed method combines the simplified 3DMM and the SfM methods. First, 2D initial frontal Facial Feature Points (FFPs) are estimated from a preliminary 3D facial image that is reconstructed by the simplified 3DMM. Second, a bilateral symmetric facial image and its corresponding FFPs are obtained from the original side-view image and corresponding FFPs by using the mirroring technique. Finally, a more accurate the 3D facial shape is reconstructed by the SfM using the frontal, original, and bilateral symmetric FFPs. We evaluated the proposed method using facial images in 35 different poses. The reconstructed facial images and the ground-truth 3D facial shapes obtained from the scanner were compared. The proposed method proved more robust to pose variations than 3DMM. The average 3D Root Mean Square Error (RMSE) between the reconstructed and ground-truth 3D faces was less than 2.6 mm when 2D FFPs were manually annotated, and less than 3.5 mm when automatically annotated.



## 1. Introduction

Researchers have focused considerable attention on 3D facial reconstruction technologies because they are useful in various applications such as pose-invariant facial recognition [1,2,38], age-invariant facial recognition [3], 3D person-specific game and movie-character generation [4,5], frontal facial synthesis of suspects in surveillance camera systems, teleconferencing, surgical simulation, etc.

These 3D facial reconstruction technologies can be divided into two approaches. The first is a hardware-based approach that reconstructs a 3D face by using additional hardware such as a stereographic camera [39], structured light [40], depth sensors [41], or a 3D laser scanner [6]. This method can be used to obtain accurate 3D facial data; however, it incurs additional costs and requires image preprocessing and camera calibration. To address these issues, a monocular camera-based approach was proposed to reconstruct the 3D face. This approach can be categorized into single-view and multiple-view-based approaches. The representative methods of these approaches are the Structure from Motion (SfM) and the 3D Morphable Model (3DMM), respectively. Shape-From-Shading (SFS) is a method to reconstruct a 3D face from the brightness variations in a single image. Although considerable research has been done on SFS-based methods, they have impractical constraints because the Lambertian reflectance model and a known light source direction need to be assumed to produce accurate results [32–35,42–44].

The multiple-view-based SfM method calculates the 3D facial shape and projection matrix using the corresponding 2D Facial Feature Points (FFPs), which are extracted from 2D facial images captured from various angles [8,9,13–16,22–24,26,45]. The 2D FFPs can be factorized into the 3D facial shape and projection matrix using Singular Value Decomposition (SVD) and the rank theorem [17]. SfM can accurately reconstruct a user-specific 3D face because more 2D facial information is available than the single-view-based approach. Moreover, a training process is not required, and the reconstructed face is not biased toward the 3D mean face model. However, SfM requires multiple-view images from various angles and the corresponding points between these facial images have to be located.

* Correspondening author. Tel.: +82 22123 2869; fax: +82 2312 4584.
*E-mail addresses:* jaeik@yonsei.ac.kr (J. Jo), hschoi@imrc.kist.re.kr (H. Choi), kij@imrc.kist.re.kr (I.-J. Kim), jhkim@yonsei.ac.kr (J. Kim).
[1] Postal address: B619, 2nd Engineering-Building, Yonsei University, 134 Sinchon-dong, Seodaemun-gu, Seoul 120-749, Korea.

http://dx.doi.org/10.1016/j.patcog.2014.07.013
0031-3203/

The single-view-based approach using the 3D model builds the 3D face model in a training process to represent facial shape, texture, illumination, and camera geometry with a number of model parameters [1,7,29–30,36–37,46]. Given a 2D facial image, the model-based method continuously optimizes 3D facial model parameters to minimize the shape and texture residuals between the 2D facial image input and a 2D facial image synthesized from model parameters. The final 3D face can be reconstructed from the obtained optimal parameters. This method enables reconstruction of a 3D face from a single facial image; moreover, it can work relatively well under various constraints. However, this method has some limitations. For one, it incurs high computational costs because the process identifies the numerous parameters through an iterative parameter optimization process [8,9]. In addition, the reconstructed 3D face can be biased toward a 3D mean face from the training data. Furthermore, the reconstruction accuracy decreases as the rotation angle of the input face increases. This decrease in accuracy is because of fitting errors caused by occluded FFPs due to head rotation.

To reduce computational complexity, which is the primary limitation of the 3DMM method, the Simplified 3DMM (S3DMM) method was proposed [3,10-12,19–21,25]. Unlike 3DMM, which reconstructs 3D facial shape, texture, and illumination, S3DMM reconstructs only a 3D facial shape. Additionally, S3DMM uses a sparse shape composed of dozens of vertices, whereas 3DMM uses a dense shape comprised of thousands of vertices. Therefore, the S3DMM method is not required to search dense FFPs; consequently, it has low computational complexity due to the decreased number of model parameters. One disadvantage of S3DMM is that it is limited by the constraint that its input facial image should be a frontal view image [11,19]. Although previous research methods [3,20,21] have reconstructed a 3D face from arbitrary views, the work remains insufficiently robust to pose variations due to vulnerability to self-occlusion errors.

For example, when a face is rotated, one side of the face is self-occluded by the other side; the FFPs in the occluded region are invisible. Therefore, the located FFPs in the self-occluded region always include location errors. This decreases the accuracy of 3D facial reconstruction by S3DMM. Lee et al. [25] decreased errors due to self-occlusion by excluding FFPs detected in the occluded region, and by using only residual FFPs that were visible after detection. However, their approach did not resolve the problem of decreased 3D reconstruction performance arising from pose variations. Many fitting errors are generated in the process of fitting the residual FFPs to a 3D model because the number of occluded FFPs to be excluded increase as the face rotation angle increases.

To address the issues mentioned above, we propose a novel single-view-based 3D facial reconstruction method. In our proposed method, the 3D facial shape is reconstructed by SfM using mirrored side-view image and frontal face image generated via bilateral symmetry and S3DMM. The primary contributions of the proposed method are as follows. First, the proposed method is robust to pose variations, unlike previous model-based methods that are vulnerable to pose variations because of fitting errors caused by occluded FFPs due to head rotation. In the proposed method, the occluded FFPs in the original view-image are revealed in the different view-images generated by bilateral symmetry and S3DMM. Subsequently, these corresponding 2D FFPs are used to calculate the 3D facial shape using SfM. Thus, the proposed method is more robust to pose variations than the previous model-based methods. Second, the proposed method is person-specific and not biased toward a mean face. The reconstruction results of the previous model-based methods, which are obtained by optimizing the parameters from the 3D mean face, can be biased toward the 3D mean face when parameter optimization falls into local minima. In contrast, the reconstruction results of the proposed method is person-specific because reconstruction is done by calculating the 3D facial shape using 2D FFPs in the different view-images of the corresponding subject. Table 1 provides a comparison of the advantages and disadvantages of our proposed method and previous 3D facial reconstruction methods.

The rest of this paper is organized as follows. In Section 2, the proposed 3D facial reconstruction method is described. In Section 3, our experimental environment and results are outlined. Conclusions and future work are presented in Section 4.

## 2. Proposed method

The proposed method is composed of several steps, which include FFP extraction, head pose estimation, occlusion point compensation, 3D facial shape reconstruction by combining S3DMM and SfM methods, dense 3D facial shape reconstruction, and texture mapping. A schematic diagram and flowchart of the proposed method are presented in Figs. 1 and 2, respectively.

**Table 1**
Comparison of proposed method and previous methods.

| | Previous SfM-based methods [8,9,13–16, 22–24,26,45] | Previous 3DMM-based method [1,3,7,10-12, 19–21,25,29–30,36–37,46] | Proposed method (SfM+3DMM) |
|---|---|---|---|
| **Overview** | • Reconstruct the 3D face from multiple-view 2D facial images<br>• 3D facial shape and projection matrix are estimated from corresponding 2D FFPs in the multiple-view facial images | • Reconstruct the 3D face from a single-view 2D facial image<br>• 3D facial shape is reconstructed by fitting 2D FFPs to a 3D mean face model | • Reconstruct the 3D face from a single-view 2D facial image<br>• Mirrored side-view image and frontal face image are generated through bilateral symmetry and S3DMM; the 3D facial shape is ultimately reconstructed by SfM. |
| **Strength** | • Person-specific, accurate 3D face can be reconstructed when multiple-view facial images are provided<br>• Training data is not required | • Can reconstruct the 3D face from only a single-view facial image | • Person-specific 3D face can be reconstructed from only a single-view facial image<br>• Robust to pose variations |
| **Weakness** | • Multiple-view facial images are required for accurate 3D facial reconstruction | • Vulnerable to pose variations<br>• Reconstructed results can be biased otowa the 3D mean face when parameter optimization falls into a local minima<br>• Requires training data to create a 3D mean face model<br>• High computational cost | • Incurs a high computational cost due to fusion of SfM and S3DMM<br>• Requires training dat |