Pattern Recognition 44 (2011) 1470–1486

Contents lists available at ScienceDirect

Pattern Recognition

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

A SfM-based 3D face reconstruction method robust to self-occlusion by using a shape conversion matrix

Sung Joo Lee [a], Kang Ryoung Park [b], Jaihie Kim [a,*]

[a] School of Electrical and Electronic Engineering, Yonsei University, Biometrics Engineering Research Center, Seoul, South Korea
[b] Division of Electronics and Electrical Engineering, Dongguk University, Biometrics Engineering Research Center, South Korea

ARTICLE INFO

Article history:
Received 8 June 2010
Received in revised form
18 October 2010
Accepted 16 November 2010
Available online 27 November 2010

Keywords:
Structure from motion (SfM)
3D face reconstruction
Self-occlusion
Shape conversion matrix

ABSTRACT

This paper presents a 3D face reconstruction method using multiple 2D face images. Structure from motion (SfM) methods, which have been widely used to reconstruct 3D faces, are vulnerable to point correspondence errors caused by self-occlusion. In order to solve this problem, we propose a shape conversion matrix (SCM) which estimates the ground-truth 2D facial feature points (FFPs) from the observed 2D FFPs corrupted by self-occlusion errors. To make the SCM, the training observed 2D FFPs and ground-truth 2D FFPs are collected by using 3D face scans. An observed shape model and a ground-truth shape model are then built to represent the observed 2D FFPs and the ground-truth 2D FFPs, respectively. Finally, the observed shape model parameter is converted to the ground truth shape model parameter via the SCM. By using the SCM, the true locations of the self-occluded FFPs are estimated exactly with simple matrix multiplications. As a result, SfM-based 3D face reconstruction methods combined with the proposed SCM become more robust against point correspondence errors caused by self-occlusion, and the computational cost is significantly reduced. In experiments, the reconstructed 3D facial shape is quantitatively compared with the 3D facial shape obtained from a 3D scanner, and the results show that SfM-based 3D face reconstruction methods with the proposed SCM show a higher accuracy and a faster processing time than SfM-based 3D face reconstruction methods without the SCM.

1. Introduction

Modeling 3D faces is a very useful technology in computer vision for its various applications, including pose-invariant face recognition [1,2], age-invariant face recognition [3], 3D person-specific game and movie character generation [4,5], teleconferencing, surgical simulation, etc. One approach used to model a 3D face requires special hardware or multi-cameras, such as a 3D laser scanner, stereo cameras, or a structured light [6]. However, applications using this approach are restrictive because of high costs and camera calibration. As an alternative, 3D face reconstruction methods by using an image sequence have been researched intensively. These methods can be categorized into model-based and SfM-based methods.

Model-based methods build 3D morphable face models offline to represent facial shape and texture, illumination, and camera geometry, with a large number of model parameters [1,7]. When 2D facial images are inputted, the methods find the model parameters of the 3D morphable face model iteratively in order to minimize the texture residual between the 2D facial image synthesized from the model parameters and the inputted 2D facial image. If the optimal model parameters are found, a detailed 3D facial shape can then be reconstructed from these model parameters. However, these methods require a high computational costs and possibly fall into a local minima because the number of model parameters is considerably large [8,9]. If they fall into a local minima, the reconstructed 3D face tends to be close to the mean face because the parameter optimization begins at the mean face [8,9]. In order to reduce the computational complexities of the original 3D morphable model, simplified morphable models representing only facial shape have been proposed in [3,10–12].

SfM-based methods estimate a 3D facial shape and a projection matrix from the corresponding 2D facial feature points (FFPs) of multiple facial images [8,9,13–16]. The basic idea of SfM is that the corresponding 2D FFPs can be factorized into the 3D facial shape and the projection matrix by using some reasonable constraints, such as the rank constraint [17]. Compared to model-based methods, these methods do not require the parameter optimization started from the mean of the 3D faces. As a result, a person-specific 3D facial shape can be reconstructed by using SfM [8,9].

SfM-based methods can be categorized into dense correspondence-based methods and sparse correspondence-based methods according to the density of the corresponding 2D FFPs. Dense correspondence-based methods find dense corresponding 2D FFPs,

* Corresponding author.
E-mail addresses: sungjoo@yonsei.ac.kr (S.J. Lee), parkgr@dongguk.edu (K.R. Park), jhkim@yonsei.ac.kr (J. Kim).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.11.012

as shown in Fig. 1(a), and reconstruct a dense 3D facial shape from them [8,9]. If dense corresponding 2D FFPs can be correctly found, these methods can reconstruct a detailed 3D facial shape. However, in general, it is difficult to find dense corresponding 2D FFPs since some facial regions, such as the cheek and the forehead, have no salient texture patterns. When point correspondence errors occur, the reconstructed 3D face appears rough. In order to obtain a smooth reconstructed 3D face, [8,9] used the mean 3D face to regularize the reconstructed 3D face.

Sparse correspondence-based methods find a predetermined number of sparse and salient corresponding 2D FFPs, as shown in Fig. 1(b), and reconstruct a sparse 3D facial shape from them [13–16]. A dense 3D mean face is then adapted to the reconstructed sparse 3D facial shape [13,14]. By using these methods, a smooth and dense 3D face can be reconstructed without dense corresponding 2D FFPs. In addition, in some applications, such as 3D game character generation, sparse corresponding 2D FFPs can be found with the assistance of the user. Therefore, these methods are very useful in such cases. The proposed 3D reconstruction method belongs to this category.

One of the fundamental problems of SfM-based methods is self-occlusion which means some facial parts occlude other facial parts when head rotation occurs. Fig. 2(b) shows the ground-truth 2D FFPs that do not include point correspondence errors, and the observed 2D FFPs when the head orientation is frontal. Fig. 2(a) and (c) show them when the head is significantly rotated. It can be seen from Fig. 2 that the error between the observed 2D FFPs and the ground truth ones increases as head rotation increases due to self-occlusion. This kind of correspondence error can degrade the performance of the SfM algorithms.

In order to solve this problem, previous SfM methods found the initial 3D facial shape based on all the observed 2D FFPs, and then reconstructed a detailed 3D facial shape iteratively by minimizing the shape residual of the visible 2D FFPs [18–20]. The matrix completion method estimated the true locations of the self-occluded FFPs by minimizing both the shape residual of the visible 2D FFPs and the nuclear norm of the estimated matrix [35,36]. However, these methods have the following limitations: Firstly, these methods are sensitive to point correspondence errors because of the reduced number of useful 2D FFPs. According to Szeliski and Kang [21], the SfM algorithms were less sensitive to point correspondence errors as the number of FFPs and the amount of object rotation increase if each FFP has an equivalent correspondence error. Unfortunately, in the face, it is difficult to obtain both a large number of FFPs and a sizable head rotation simultaneously, because more FFPs become self-occluded as head rotation increases. Secondly, these methods basically require a lot of iterations which lead to high computational costs. Finally, these methods require an additional self-occlusion detector in order to find the visible FFPs. As a result, performance is dependent on the self-occlusion detector.

In order to solve these problems, we propose a shape conversion matrix (SCM) to estimate the true locations of the self-occluded FFPs. In other words, the SCM transforms the observed facial shape to a converted facial shape that is closer to the ground truth facial shape shown in Fig. 2.

In order to make the SCM, observed 2D FFPs and ground-truth 2D FFPs were collected from subjects and two shape models were built to represent these 2D FFPs using principal component analysis (PCA). Then, two shape parameters were found by projecting the two types of 2D FFPs onto their corresponding shape models. The SCM was found to convert the observed shape parameter to the ground truth shape parameter by using a least square method. By using the SCM, the true locations of the self-occluded FFPs are estimated exactly with simple matrix multiplications. As a result, SfM-based 3D face reconstruction methods combined with the proposed SCM become more robust against correspondence errors with only a small additional computational cost. Furthermore, the proposed method does not require a self-occlusion detector to find the visible points because the self-occluded points are estimated from all of the observed FFPs by using the SCM. Table 1 shows a comparison between previous 3D face reconstruction methods and the proposed method.

The remainder of the paper is organized as follows: Section 2 describes the general procedures for sparse correspondence-based 3D face reconstruction and previous solutions for the self-occlusion problem. Section 3 presents the proposed method using the SCM. Section 4 lays out our experimental environment and the quantitative and qualitative results. Finally, conclusions are given in Section 5.

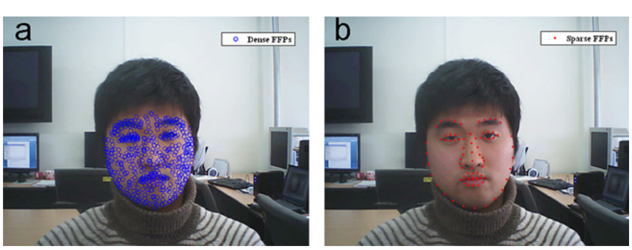

Fig. 1. Examples of 2D FFPs: (a) dense 2D FFPs and (b) sparse 2D FFPs.

2. Related works

2.1. General procedure for sparse correspondence-based 3D face reconstruction

As shown in Fig. 3, the general procedure for sparse correspondence-based 3D face reconstruction consists of sparse 2D FFPs extraction, 3D reconstruction of the sparse FFPs using SfM, 3D dense mean model adaptation, and texture mapping.

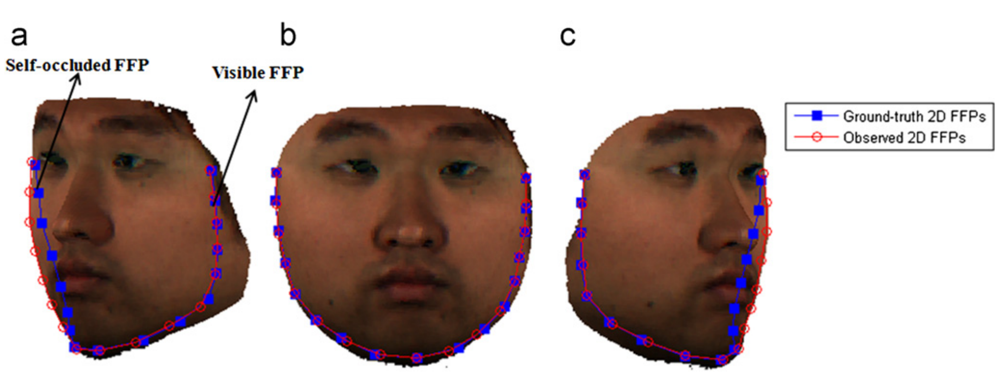

Fig. 2. 2D facial contour FFPs when head rotation is: (a) -45°, (b) 0°, and (c) 45°.