Neurocomputing 179 (2016) 228–237

Contents lists available at ScienceDirect

Neurocomputing

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

# 3D face modeling based on structure optimization and surface reconstruction with B-Spline

Weilong Peng [a], Chao Xu [b,*], Zhiyong Feng [a]

[a] School of Computer Science and Technology, Tianjin University, China
[b] School of Computer Software, Tianjin University, China

ARTICLE INFO

*Article history:*
Received 30 March 2015
Received in revised form
31 October 2015
Accepted 28 November 2015
Communicated by Qingshan Liu
Available online 29 December 2015

*Keywords:*
3D face reconstruction
Low rank
B-spline face
Structure optimization
SSIM

ABSTRACT

How to reconstruct 3D face model from wild photos is such a difficult issue that camera calibration is necessary and the images must be from video sequences. In this paper, a face reconstruction model with structure optimization is proposed to build 3D face surface with individual geometry and physical features reservation through wild face images directly and without camera calibration. Low rank and B-Spline are employed to estimate the aligned 2D structure, to calculate the depth information with SSIM, and to reconstruct the 3D face surface from control points and their space transformation. Furthermore, LFW and Bosphorus datasets, as well as Young-to-Aged samples, are introduced to verify the proposed approach and the experimental results demonstrate the feasibility and effectiveness even with different poses, expressions and age-variety.



## 1. Introduction

Face reconstruction is usually based on a scene that the subject is fixed and scanned with special equipment under a lab condition. And other techniques can be based on video sequences or multi-view photographs. The reconstructed face model will be of greate help for many other tasks relative to face, such as recognition [19], animation [4], and tracking [33-35]. However, a more challenging work is to reconstruct face structure based on photos in the wild, with unknown camera calibration.

*Wild* means that the images are captured under different environments, resulting in their variety in expression, illumination, poses, and even various ages. Affected by these variations, the *shape* of a person's face is difficult to be defined. Currently, there exists two main kinds of approaches used to create a 3D face model. The first kind uses special 3D scanner device to capture the shapes of face, e.g. [1]. The second one reconstructs face based on 2D images, such as video sequences [2], or multi-view photographs, e.g. [3]. Several relative techniques will be introduced.

Most state of the art techniques [4,5,1] implement high quality face reconstruction using 3D scanning under highly-controlled lab condition, with special equipment such as laser, stereo and structural lighting. And Multi-view stereo approaches [5–7] rely on synchronized data from multiple high resolution cameras with known cameras calibration. Structured light [4] and light stages [8] based reconstructions need multiple synchronized and calibrated lights. However, it is inconvenient to distribute the devices for wide application. Therefore, it is significant to research reconstruction without calibration.

*Single-view method*: Kemelmacher et al. [9] uses a single image for reconstruction by shape-from-shading (SFS) approach [10], but it also relies on a template face as a prior. And Suwajanakorn et al. [11] reconstruct highly detailed 3D shape using dense 3D flow and SFS for each video frame, but it needs an individual model. Due to the fact that the single-view problem is ill-posed [12], there is a uncertain uniqueness of shape recovering via the reflection model. Thus, solution with SFS is still an intractable problem [13]. A template in [9] and an individual model in [11] are applied to yield good looking results. Sometimes, multiple photos are more reliable.

Structure-from-motion (SFM) uses multiple frames of image sequences to recover 3D shape of an object [14,15], estimating the rigid 3D structure of the feature points with 2D observation [16,17]. The incremental SFM approach is proposed in [18] to build a 3D generic face model for non-rigid face. Although it gives a good generalization performance with respect to expression and identify by incorporating prior 3D shape information, the incremental SFM is not applicable for the multi-view wild images.

Most related work are the spatial-transformation approach [19], and reconstruction based on multiple wild photos [20]. Similarity transformation is used to optimize the 3D face structure with a set of face images under different poses in [19]. During the procedure, a frontal constraint is necessary to yield good result,

* Corresponding author.
*E-mail address:* chaoxu@tju.edu.cn (C. Xu).

http://dx.doi.org/10.1016/j.neucom.2015.11.090
0925-2312/

but at last it cannot generate a dense model. Comparing with Bundle adjustment in [14], our method fits the facial structure problem well although both of them is essentially based on minimizing reprojection error. It is because that face is nonrigid and varies in different wild images in our problem but bundle adjustment always fits the large scale rigid object reconstruction. Our method also differs from the 3DMM [21–23] that learns basis-face representation based on a 3D database, because it directly optimizes without learning. What is important, we solve the frontal optimization by low rank, reducing the difficulty of structure optimization, and also generate the spline surface of face. Kemelmacher [20] assumes that there exists a subset of photos with local consistent shade for each local region, and build face model based on local optimization. Its good reconstruction always relies on a constraint that the image set must be "identified" as one person by some algorithm, to ensure the local area with global consistency. Then it is invalid without identification when the photos are various at ages, poses, and so on. Roth et al. [24] perform reconstruction by landmark driven mesh deformation based on photometric normal. It assumes that the face surface is $C^0$-continuous at every vertex of mesh model, and it cannot be approximated and guaranteed for surface smoothness and true shape in arbitrary precision. In contrast, we focus on optimizing 3D structure according to the face 2D structure cues in the photos, and try to obtain the consistent $C^2$ shape of the whole face via optimization. Significantly, the $C^2$ face is built by optimizing based on low rank and B spline, which is an innovative work of combining statistic method and continuous geometric model.

In our work, at first we obtain the 3D face structure by leveraging the structure cues from multiple images with low rank and SSIM based depth optimization. Then face can be reconstructed based on 3D structure transformation and B-spline. The main contributions of this paper are:

(1) *3D face structure optimization*: frontal face structure optimization is considered as a sparse and low rank decomposition, and depth estimation as nonlinear programming based on constraints of multi-substructure;
(2) Face surface construction with B-spline control grid deforming guided by 3D structure transformation;
(3) 3D reconstruction solution based on wild photos, instead of calibration.

The remainder of the paper is organized as follows. Section 2 presents our method in detail. Section 3 shows the experimental results and related discussions.

## 2. Methodology

In detail, we optimize firstly the objective 3D structure from a reference one; Secondly, B-Spline control grid of reference surface is guided to be deformed according to structure transformation from reference to objective; At last, objective surface is constructed by the deformed control points. This procedure is illustrated in Fig. 1. Particularly, the 3D face structure refers to the CANDIDE 3D model [25] for coding face structure.

Since it is difficult to optimize 3D structure with 2D images, we re-arrange the optimization as: (1) frontal 2D structure optimization by matrix rank minimum in image coordinates for generating $x$ and $y$ components of 3D structure, and (2) depth optimization for generating $z$ component for frontal 2D structure scaled into 3D coordinate. All the 3D structures are unified to a standard one with 50 unit long in distance between two eye centers (seen in the left of Fig. 1). Therefore, the optimized 3D face structure can guide construction of B-Spline face surface.

### 2.1. Frontal face structure optimization by matrix rank minimization

In this section, we discuss why rank is a natural measurement of face structure similarity, and how to abstract the frontal face structure. And then 2D structure optimization is formulated to achieve a set of transformations by minimizing the rank of the images.

*Modeling Assumption*: Given N face images from an individual, note $np$ facial points representing 2D structures for each image as

$$S_i = \begin{bmatrix} u_{i,1} & u_{i,2} & \dots & u_{i,np} \\ v_{i,1} & v_{i,2} & \dots & v_{i,np} \end{bmatrix},$$

where $i \in 1, 2, \dots, N$. These facial points in 2D structure can be marked manually or by detection algorithm [26] that has a similar high performance to human. In fact, these 2D structures are projections from the same 3D facial structure $T$ from 3D space, and

$$T = \begin{bmatrix} x_1 & x_2 & \dots & x_{np} \\ y_1 & y_2 & \dots & y_{np} \\ z_1 & z_2 & \dots & z_{np} \end{bmatrix}.$$

The differences among these structures mainly result from face

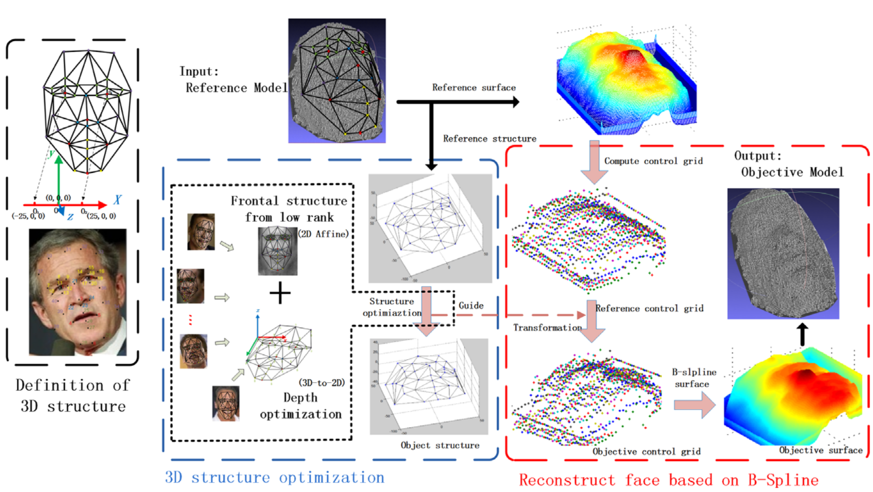


**Fig. 1.** Framework: 3D face reconstruction based on 3D structure optimization and B-Spline. The left is the definition of 3D structure including 40 points: the bottom shows that the structure topology has point $O_1(-25, 0, 0)$ and point $O_2$ (25, 0, 0) locating eye centers in 3D space, and it looks like a frontal 2D face structure in direction of normal (0, 0, 1); the up shows the point positions on the face.