# Adaptive Contour Fitting for Pose-Invariant 3D Face Shape Reconstruction

Chengchao Qu<sup>12</sup> chengchao.qu@iosb.fraunhofer.de Eduardo Monari<sup>2</sup> eduardo.monari@iosb.fraunhofer.de Tobias Schuchert<sup>2</sup> tobias.schuchert@iosb.fraunhofer.de Jürgen Beyerer<sup>21</sup> juergen.beyerer@iosb.fraunhofer.de

- <sup>1</sup> Vision and Fusion Laboratory Karlsruhe Institute of Technology Karlsruhe, Germany
- <sup>2</sup> Fraunhofer IOSB Karlsruhe, Germany

#### Abstract

Direct reconstruction of 3D face shape—solely based on a sparse set of 2D feature points localized by a facial landmark detector—offers an automatic, efficient and illumination-invariant alternative to the conventional analysis-by-synthesis 3D Morphable Model (3DMM) fitting. In this paper, we propose a novel algorithm that addresses the inconsistent correspondence of 2D and 3D landmarks at the facial contour due to head pose and localization ambiguity along the edge. To facilitate dynamic correspondence while fitting, a small subset of 3D vertices that serves as the contour candidates is annotated offline. During the fitting process, we employ the Levenberg-Marquardt Iterative Closest Point (LM-ICP) algorithm in combination with Distance Transform (DT) within the constrained domain, which allows for fast convergence and robust estimation of 3D face shape against pose variation. Superior evaluation results reported on ground truth 3D face scans over the state-of-the-art demonstrate the efficacy of the proposed method.

### **1** Introduction

Since the introduction of the 3D Morphable Model (3DMM) by Blanz and Vetter [ $\square$ ], 3D model fitting for facial analysis has seen broad applications in face recognition [ $\square$ ,  $\square$ ], computer animation [ $\square$ ,  $\square$ ] and face hallucination [ $\square$ ,  $\square$ ], *etc.* In [ $\square$ ], a complete 3DMM fitting framework with regard to shape, texture and illumination is given at the cost of extremely high computational expense. Hence, in quest of efficiency, especially with the surge of mobile entertainment systems and social media applications in recent years, a new strategy that only recovers 3D shape of the face using a few facial landmarks emerges. Our method also falls into this category.

The 2D facial landmarks are either labeled manually or automatically detected by face alignment algorithms [23]. Afterwards, 3D motion and 3DMM shape parameter can be estimated by the correspondence of 2D and 3D landmarks, for which most existing approaches [10, 11, 11, 11, 11], 12] assume a fixed mapping. Until recently, an evident flaw of this mapping



Figure 1: Overview of the proposed 3D face shape reconstruction framework.

scheme is identified in  $[\square X]$ ,  $[\square X]$ , in which it is shown that in the less visible face half, the 2D contour landmarks deviate greatly from the true 3D locations because of self-occlusion in non-frontal poses. In this work, we go a step further to account for ambiguous landmark positions along the facial contour for both halves of the face.

We distinguish between fixed and flexible (contour) landmarks in the course of shape reconstruction. Instead of directly minimizing the distance of the corresponding landmarks, Distance Transform (DT) is first applied to the line segments bounded by 2D contour landmarks. At the same time, the proper 3D vertices can be chosen from a small candidate set. Subsequently, together with the fixed points, the projected distance is minimized by the Levenberg-Marquardt Iterative Closest Point (LM-ICP) algorithm and the 3DMM shape coefficients as the optimization parameters are obtained within only a few iterations. Comparison with prevailing approaches justifies the robustness of the proposed 3D shape reconstruction algorithm against pose variation. The framework is illustrated in Fig. 1 and our main contributions are summarized as follows:

- i. We argue that not only the self-occluded contour landmarks, but also the visible ones are susceptible to 2D–3D correspondence discrepancy.
- ii. By formulating the 3D shape reconstruction as a general-purpose LM-ICP optimization problem incorporated with DT, a robust unified solution for fixed and flexible landmark mapping is found without loss of efficiency.
- iii. A fast method is presented to estimate the 3D silhouette vertices in LM-ICP iterations.

The remainder of this paper is organized as follows. A brief introduction to the previous work in 3D face reconstruction is given in §2. §3 first recalls 3D shape recovery using 3DMM and analyzes the encountered problem as our motivation before we elaborate on the proposed framework. Quantitative results of our fitting method are compared to existing approaches in §4. Finally, our work is concluded in §5.

### 2 Related Work

2

The merit of a pose, expression and illumination invariant description of 3D faces has attracted considerable attention and research effort over the past decades. Hindered by the high cost and practical difficulties during enrollment, 3D cameras are still limited from being deployed outside of the lab environment [III]. Other techniques such as Shape-from-Shading (SFS) [III] are also beyond the scope of this paper. Hence, in this section, we give a short overview on model-based 3D face reconstruction from a single image.

The pioneering work of 3DMM [1] establishes the fundamental idea of describing human faces as linear shape and texture subspaces of aligned 3D training scans. When fitting 3DMM to 2D images, both shape and photometric parameters, *e.g.*, camera calibration, illumination and shading, are simultaneously estimated. This analysis-by-synthesis framework, although extended by Romdhani and Vetter w.r.t. fitting strategy [24] and with the inclusion of auxiliary image features [22], is extremely time-consuming considering the enormous parameter space.

Fortunately, by leveraging a few fiducial facial landmarks, it is viable to dramatically reduce the dimensionality by leaving out the entire motion, albedo and illumination parameters, as only the 3DMM shape coefficients need to be reconstructed. Moreover, the shrinkage from tens of thousands of dense vertices to merely dozens of sparse ones also contributes to a huge boost in runtime. Following this direction, Blanz et al. [2] first propose to use manually annotated sparse feature points to infer the 3D shape within the span of 3DMM based on the correlation of features. The camera projection is vectorized as extra parameters such that a closed solution for arbitrary poses exists. As an extension of [1], Faggian *et al.* [1] first involve facial landmark detection with the help of the Active Appearance Model (AAM) towards fully automatic 3D face shape reconstruction. By incorporating combined optimization of multiple images, robustness across data is enhanced [1]. Aldrian and Smith [1] loose the assumption in [1] that observations of all landmarks are subject to uncorrelated Gaussian noise with a global variance, as they learn the individual generalization errors by projecting out-of-sample data onto the 3DMM subspace. Without the explicit training of a 3DMM, Rara et al. [23] avail of 3D sample faces directly and exploit Principal Component Regression (PCR) to reconstruct the 3D shape as a linear combination of the samples instead of 3DMM eigenfaces as in [0, 6]. Practically the same evaluation result is reported. As an application to face recognition, Jiang et al. [11] are able to reconstruct frontal probe faces for synthesizing images in different poses, expressions and illuminations.

The empirical assumption of a fixed mapping between 2D and 3D landmarks gives rise to a major drawback of the aforementioned methods. Since the contour landmarks of 2D AAM are originally defined as the jawline that is easily occluded even with small head pose, the points on the face boundary in the image plane are detected instead, which have a considerable distance to their true locations. To mitigate the negative impact of the erroneous observation, Lee *et al.* [II] propose to discard these self-occluded landmarks while reconstructing non-frontal faces. Qu et al. [22] further incorporate a multi-frame least squares approach [12] for both image and video data of unconstrained poses. Experimental results show that this straightforward idea appears to be surprisingly effective. Asthana *et al.*  $[\Box]$  are also aware of this issue when normalizing poses for face recognition. Unlike our automatic silhouette detection algorithm, they manually label the 2D-3D correspondence of 199 poses with yaw angles from  $-45^{\circ}$  to  $+45^{\circ}$  and pitch angles from  $-30^{\circ}$  to  $+30^{\circ}$  and produce realistic frontalization and excellent recognition scores. Lately, Dou et al. [] circumvent this problem by learning a coupled regression subspace of 2D and 3D sparse landmarks, and a coupled dictionary of 3D sparse-dense shapes. While preparing this manuscript, a novel approach [11] emerges, which, similar to this work, dynamically finds the 3D correspondence of 2D contour landmarks. However, flexible fitting along the facial contour is not possible with the employed conventional fixed 2D-3D mapping.

### **3** Proposed Method

This section details our fully automatic framework for dense 3D face shape reconstruction using dozens of feature points localized by off-the-shelf landmark detectors. To start with, introduction to 3DMM and the basic linear shape reconstruction method is given. We argue the necessity of facilitating flexible 2D–3D mapping of landmarks and propose a novel reconstruction algorithm that can effectively handle self-occlusion and inaccurate landmark localization at the facial contour.

#### 3.1 Basics: 3DMM and 3D Shape Recovery

A morphable face model [**G**], usually built from 3D laser scans of several hundred subjects, represents human faces with 3D geometry  $\mathbf{s} = [x_1, y_1, z_1, \dots, x_P, y_P, z_P]^\top \in \mathbb{R}^{3P}$  and albedo  $\mathbf{t} = [r_1, g_1, b_1, \dots, r_P, g_P, b_P]^\top \in \mathbb{R}^{3P}$ , where the *P* vertices are put into dense correspondence across the data. Applying Principal Component Analysis (PCA) to all scans yields

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{S}\operatorname{diag}(\sigma)\alpha$$
 and  $\mathbf{t} = \bar{\mathbf{t}} + \mathbf{T}\operatorname{diag}(\tau)\beta$ , (1)

where  $\mathbf{S} \in \mathbb{R}^{3P \times M}$  and  $\mathbf{T} \in \mathbb{R}^{3P \times M}$  are composed by the *M* eigenvectors, while  $\sigma \in \mathbb{R}^{M}$  and  $\tau \in \mathbb{R}^{M}$  are the respective eigenvalues spanned by the PCA subspace. Together with the mean shape  $\bar{\mathbf{s}}$  and mean texture  $\bar{\mathbf{t}}$ , the main components of 3DMM are representable as  $\{\bar{\mathbf{s}}, \mathbf{S}, \sigma, \bar{\mathbf{t}}, \mathbf{T}, \tau\}$ , of which only the shape related part  $\{\bar{\mathbf{s}}, \mathbf{S}, \sigma\}$  is of interest in this work.

When recovering the 3D shape from a sparse set of feature points, an estimate of the 3DMM shape coefficients  $\alpha$  alone suffices according to Eq. (1). Assuming that  $F \ll P$  facial landmarks are manually or automatically localized, their 2D image coordinates  $\mathbf{r} \in \mathbb{R}^{2F}$  can be expressed as a projection of 3D shapes

$$\mathbf{r} = \Pi \Phi(\bar{\mathbf{s}} + \mathbf{S} \operatorname{diag}(\sigma) \alpha), \qquad (2)$$

where  $\Pi$  denotes the affine camera projection matrix that models scaling, rotation and translation.  $\Phi$  selectively maps a subset of *P* vertices to the sparse *F* landmarks. In order to obtain  $\Pi$ , the Gold Standard Algorithm [II] using least squares minimization of the normalized 2D–3D correspondences can be applied. Alternatively, Blanz *et al.* [I] linearize the camera projection parameters to obtain a closed solution. In our own experiments, we find that both methods produce satisfactory and comparable results in 2~3 iterations. In the presence of 3DMM generalization error, global zero-mean Gaussian white noise is assumed [I] and a Maximum a Posteriori (MAP) formulation of the cost function reveals

$$\hat{\boldsymbol{\alpha}} = \underset{\alpha}{\arg\min} E(\boldsymbol{\alpha}) = \underset{\alpha}{\arg\min} \| \Pi \Phi \mathbf{S} \operatorname{diag}(\boldsymbol{\sigma}) \boldsymbol{\alpha} - (\mathbf{r} - \Pi \Phi \bar{\mathbf{s}}) \|_{2}^{2} + \eta \| \boldsymbol{\alpha} \|_{2}^{2}$$

$$= \underset{\alpha}{\arg\min} \| \mathbf{Q} \boldsymbol{\alpha} - \mathbf{y} \|_{2}^{2} + \eta \| \boldsymbol{\alpha} \|_{2}^{2}.$$
(3)

Here simplification is made by substituting  $\mathbf{Q} = \Pi \Phi \mathbf{S} \operatorname{diag}(\sigma)$  and  $\mathbf{y} = \mathbf{r} - \Pi \Phi \mathbf{\bar{s}}$ . The regularization term  $\eta \|\alpha\|_2^2$  prevents overfitting when minimizing the absolute 2D projection error, and thus regulates the plausibility of fitting [**L**]. Finally, by setting

$$\nabla_{\alpha} E = \mathbf{0},\tag{4}$$

a straightforward regularized least squares solution exists

$$\hat{\boldsymbol{\alpha}} = \left( \mathbf{Q}^{\top} \mathbf{Q} + \eta \mathbf{I} \right)^{-1} \mathbf{Q}^{\top} \mathbf{y}.$$
 (5)



Figure 2: Correspondence errors of 2D (red) and 3D (green) facial contour landmarks w.r.t. yaw angles of (a)  $0^{\circ}$ , (b)  $10^{\circ}$ , (c)  $20^{\circ}$  and (d)  $30^{\circ}$ .

#### 3.2 Self-Occlusion and Ambiguity of Contour Landmarks

In the previous efforts towards automated 3D shape reconstruction by means of facial landmark detectors, most authors underestimate the difference in representation power between 2D face alignment and 3DMM. Fig. 2 demonstrates the automatically detected 2D facial contour landmarks (red) and the respective 3D ground truth vertices (green) at different yaw angles of a sample 3D face. Obviously, with increasing yaw angle, remarkable deviation in the self-occluded face half can be observed. The reason behind is that face alignment methods determine the landmark positions on the basis of the whole face texture or the image patches around the sparse landmarks, where these 2D texture features by nature cannot infer the hidden 3D structure [22]. Therefore, only the detection of the face silhouette is practical, whereas the real invisible jawline of non-frontal faces turns out to be intractable, even with a 2.5D extension [12]. On the contrary, 3DMM provides a much denser 3D representation. Both geometry and albedo information is tightly coupled into the 3D vertices, which always correspond to roughly the same place on the face independent of pose variation. As a consequence, it is safe to claim that a fixed 2D–3D mapping is inappropriate.

Nevertheless, after careful reconsideration, we discover that the issue related to the facial contour is in fact more than just self-occlusion. Fig. 2 reveals that the visible contour landmarks are also affected, which is again attributed to 2D landmark detection. While detecting contour landmarks, change of gradient across the jawline or the silhouette offers helpful information for localizing the overall curve profile. However, unlike for the inner facial components, it lacks distinct image features to determine the absolute positions along the contour. That means even in frontal view, a tight correspondence of the contour landmarks cannot be necessarily guaranteed (see Fig. 2a). Unfortunately, the authors in [12, 12] are only aware of self-occlusion. Manual correction of the hidden 2D contour landmarks compared to the 3D ground truth is depicted in the figures to emphasize this problem and the visible ones are regarded as irrelevant.

#### 3.3 Fast Detection of Silhouette Vertices

The question now arises as to how to alleviate these two issues effectively and efficiently, as discarding the occluded 2D landmarks during fitting [13, 22] is not considered here due to loss of valuable information in the first place. Furthermore, we cannot ignore the visible landmarks in our case, either. Otherwise the facial form would be totally unconstrained.

Recall that the 2D landmarks are always located at the boundary of the rotated faces,

6



Figure 3: Fast detection of silhouette vertices using (a) a few annotated candidates. (b) and (c) show our result in different views compared to (d) the direct approach  $[\Box A]$ .

which varies w.r.t. shape and pose variations. A straightforward approach is to compute the boundary vertices on the fly. Mathematically, the tangent planes of those vertices are perpendicular to the image plane, meaning that their normal vectors projected onto the z-axis are close to zero. It is then an intuitive idea to treat those with the absolute z-projection values of the normals  $|\mathbf{n}_z| < t$  as silhouette points [ $\square$ ]. By carefully choosing an appropriate threshold t and the face region constraint, an example detection is shown in Fig. 3d. At first sight, this method seems to give legitimate results. However, a universally valid threshold for all cases is hard to find, which leaves the number of the selected vertices unstable. Secondly, the spatial distribution is uncontrollable, too. Both nuisance factors make it extremely challenging to derive a robust closed-form solution. Last but no least, the high computational effort of densely evaluating the normals rules out the possibility of online calculation for iterative methods [ $\square$ ], e.g., LM-ICP in our case.

We present a fast approach free of the aforementioned drawbacks to specify the closest 3D silhouette vertices to the 2D landmarks. First, starting from each original contour landmark mapping on the 3D model, a maximum of 20 extended vertices towards the center of the face are labeled offline. During the fitting process, the ones with the smallest  $|\mathbf{n}_z|$  on each horizontal line are chosen. Despite following the principle of the direct approach [ $\Box$ ], the additional geometric constraints reduce the number of evaluated vertices for normal calculation by two orders of magnitude to approximately 100. Moreover, the same number of 3D silhouette vertices as 2D landmarks with a uniform spatial distribution is guaranteed. An overview of the proposed silhouette detection method is illustrated in Fig. 3. Note that the vertices, now like those in the visible face half, are still subject to positional uncertainty along the path of 2D contour landmarks. Our adaptive fitting solution addresses this issue.

#### 3.4 Adaptive Contour Fitting for 3D Shape Reconstruction

In consequence of the apparently non-isotropic Gaussian uncertainty (see Fig. 2), 3D shape recovery by separately modeling the noise variances for each landmark [I] is not applicable. Since deviation of the 3D vertices detected in §3.3 from the 2D contour landmarks should not be penalized, as long as those 3D vertices stay on the curve formed by the 2D landmarks, it makes sense to exploit the continuous curve instead of discrete landmarks when reconstructing the shape. As a side effect, though, the coupled correspondence of 2D–3D contour landmarks is lost, as all 2D coordinates on the curve are now eligible (see Fig. 4a).

We start from scratch and revisit the basic reconstruction formulation in Eq. (3) to seek a



Figure 4: (a) Improved 2D features with connected contour lines, (b) their Distance Transform and its derivatives in (c) x and (d) y directions.

solution. Assuming that the 2D–3D correspondence of the inner facial landmarks is detected plausibly by virtue of their informative image features, separating all landmarks into two disjoint subsets of fixed and contour ones in Eq. (3) leads to

$$E(\alpha) = \left\| \begin{bmatrix} \mathbf{Q}_{\text{contour}} \\ \mathbf{Q}_{\text{fixed}} \end{bmatrix} \alpha - \begin{bmatrix} \mathbf{y}_{\text{contour}} \\ \mathbf{y}_{\text{fixed}} \end{bmatrix} \right\|_{2}^{2} + \eta \|\alpha\|_{2}^{2}.$$
(6)

7

An unknown mapping denoted  $\phi(i) = j$  which selects, for each 3D contour vertex *i*, the corresponding 2D contour pixel *j*, is now also a part of the minimization process

$$E(\boldsymbol{\alpha}, \boldsymbol{\phi}) = \sum_{i} \left\| \mathbf{Q}_{i} \boldsymbol{\alpha} - \mathbf{y}_{\boldsymbol{\phi}(i)} \right\|_{2}^{2} + \left\| \mathbf{Q}_{\text{fixed}} \boldsymbol{\alpha} - \mathbf{y}_{\text{fixed}} \right\|_{2}^{2} + \eta \left\| \boldsymbol{\alpha} \right\|_{2}^{2}$$

$$E(\boldsymbol{\alpha}) = \sum_{i} \min_{j} \left\| \mathbf{Q}_{i} \boldsymbol{\alpha} - \mathbf{y}_{j} \right\|_{2}^{2} + \left\| \mathbf{Q}_{\text{fixed}} \boldsymbol{\alpha} - \mathbf{y}_{\text{fixed}} \right\|_{2}^{2} + \eta \left\| \boldsymbol{\alpha} \right\|_{2}^{2}.$$
(7)

As a result, estimation of the shape parameter  $\alpha$  is formulated as a "minimization of minimization" problem

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \sum_{i} \min_{j} \left\| \mathbf{Q}_{i} \boldsymbol{\alpha} - \mathbf{y}_{j} \right\|_{2}^{2} + \left\| \mathbf{Q}_{\text{fixed}} \boldsymbol{\alpha} - \mathbf{y}_{\text{fixed}} \right\|_{2}^{2} + \eta \left\| \boldsymbol{\alpha} \right\|_{2}^{2}.$$
(8)

A common practice for solving the correspondence problem is the Iterative Closest Point (ICP) algorithm, which computes  $\phi$  given fixed  $\alpha$  and updates  $\alpha$  on the basis of  $\phi$  in a suboptimal alternating manner. Fitzgibbon [13] addresses the deficiency with the Levenberg-Marquardt Iterative Closest Point (LM-ICP) algorithm, which tolerates a larger basin of convergence and allows for a closed solution and speedup. The Levenberg-Marquardt (LM) optimization procedure is particularly suited to our cost function  $E(\alpha)$  in Eq. (7) which is a sum of squared residuals. However, like conjugate gradient and Gauss-Newton, the requirement for first derivatives seems intractable for the discrete minimization over *j* within summation. The trick to circumvent this difficulty is to apply Distance Transform (DT) to the discrete 2D features. On the 2D image raster **x**, DT assigns each image pixel with the distance to its closest point on the contour lines

$$D(\mathbf{x}) = \min_{j} \left\| \mathbf{x} - \mathbf{y}_{j} \right\|_{2}.$$
(9)

In particular, we make use of the Bresenham's algorithm  $[\square]$  to first mark all contour pixels on a binary image. Once DT is then efficiently computed (see Fig. 4b), it is reusable for the

# LM-ICP iterations	1	3	5	10	20	30	40	50
Shape error (mm)	5.490	5.472	5.487	5.485	5.447	5.466	5.469	5.472
Normal direction error (°)	9.320	9.272	9.276	9.277	9.245	9.246	9.245	9.245

Table 1: Influence of the number of LM-ICP iterations on the 3D reconstruction errors.

entire reconstruction procedure by virtue of its independence of the model parameter  $\alpha$ . The merit of DT lies in that the mapping function  $\phi(i)$ , or the minimization over *j* in the contour cost of Eq. (7), then vanishes and is thereby simply replaced with

$$D(\mathbf{Q}_i \alpha) = \min_j \|\mathbf{Q}_i \alpha - \mathbf{y}_j\|_2.$$
(10)

Integrating Eq. (10) into Eq. (7) and vectorizing DT over all contour vertices *i* yields

$$E(\boldsymbol{\alpha}) = \|\mathbf{D}(\mathbf{Q}_{\text{contour}}\boldsymbol{\alpha})\|_{2}^{2} + \|\mathbf{Q}_{\text{fixed}}\boldsymbol{\alpha} - \mathbf{y}_{\text{fixed}}\|_{2}^{2} + \eta \|\boldsymbol{\alpha}\|_{2}^{2}.$$
 (11)

Rather than the sum of squares  $E(\alpha)$ , LM-ICP demands the vector of residuals

$$\mathbf{e}(\alpha) = \begin{bmatrix} \mathbf{D}(\mathbf{Q}_{\text{contour}}\alpha) \\ \mathbf{Q}_{\text{fixed}}\alpha - \mathbf{y}_{\text{fixed}} \\ \sqrt{\eta}\alpha \end{bmatrix}.$$
 (12)

Differentiating Eq. (12) analytically subject to the shape parameter  $\alpha$  is possible when the chain rule  $\frac{\partial D_i}{\partial \alpha_j} = \frac{\partial D}{\partial x} \mathbf{f}_i \cdot \frac{\partial \mathbf{f}_i^x}{\partial \alpha_j} + \frac{\partial D}{\partial y} \mathbf{f}_i \cdot \frac{\partial \mathbf{f}_i^y}{\partial \alpha_j}$  and the precomputed gradient images in x and y directions (Figs. 4c and 4d) are applied to calculate the discrete derivatives of the contour  $\cot \nabla_{\alpha} \mathbf{D}(\mathbf{Q}_{\text{contour}}\alpha)$ . The target Jacobian matrix  $\mathbf{J}_{ij} = \frac{\partial \mathbf{e}_i}{\partial \alpha_i}$  is then

$$\mathbf{J} = \begin{bmatrix} \mathbf{D}_{x}(\mathbf{Q}_{\text{contour}}\alpha) \cdot \mathbf{Q}_{\text{contour}}^{x} + \mathbf{D}_{y}(\mathbf{Q}_{\text{contour}}\alpha) \cdot \mathbf{Q}_{\text{contour}}^{y} \\ \mathbf{Q}_{\text{fixed}} \\ \sqrt{\eta}\mathbf{I} \end{bmatrix},$$
(13)

which dramatically reduces the reconstruction time to around one second from over one minute using finite difference approximation.

**Discussion** Romdhani and Vetter [ $\Box$ ] also employ LM-ICP [ $\Box$ ] to simultaneously find the 2D–3D correspondence and minimize the error function. Major differences that distinguish our contribution from theirs are: (i) The contour is an indispensable 2D feature for our 3D shape reconstruction, while in [ $\Box$ ], it is merely one of the several supplementary features, *e.g.*, textured edges and specular highlights, to the analysis-by-synthesis framework [ $\Box$ ]; (ii) We exploit the 2D contour landmarks in both face halves, whereas [ $\Box$ ] detects the silhouette edges in the occluded face half; (iii) The proposed fast detection of silhouette vertices ideally facilitates online update within LM-ICP iterations. By comparison, direct global estimation [ $\Box$ ] can be done only once on the initial shape owing to performance reasons.

### 4 **Experiments**

8

Our experiments are systematically conducted on the Basel Face Model (BFM) [21], which includes a 3DMM consisting of 53,490 vertices trained with 100 male and 100 female subjects, as well as 10 out-of-sample faces for evaluation purposes. Each face is rendered in



Figure 5: Mean 3D reconstruction errors of shape and normal direction in given poses, averaged over 10 BFM sample faces.

7 poses of yaw rotation from  $-30^{\circ}$  to  $30^{\circ}$  with  $10^{\circ}$  interval. 68 fiducial facial landmarks of the rendered images from a recent detector [23] and manual offline annotation of the corresponding inner and contour vertices are available. We report quantitative errors on a laptop with 2.6 GHz CPU and 8 GB RAM using two criteria, *i.e.*, mean Euclidean shape error and mean normal error w.r.t. the ground truth shape, as they reflect on the reconstruction quality in the vertex and facet perspectives, respectively.

The convergence property of LM-ICP is first evaluated and the results are listed in Tab. 1. The errors are averaged over all 70 renderings of the tested BFM subjects and poses. Interestingly, even with a single iteration of LM-ICP optimization, the performance is almost on par with that of the best option. Fast convergence and stability of LM-ICP [I] as well as the efficacy of pose estimation are demonstrated. With an increasing number of iterations, the least errors in both shape and normal are reached after 20 iterations. Hence, despite minimal difference in performance, we fix this number for further experiments, since shape reconstruction in approximately one second is still orders of magnitude faster than the full 3DMM fitting algorithms [I], [I], [I], [I]], which need over one minute to fit an image.

We now compare our work against state-of-the-art approaches on BFM, namely the basic 3D shape reconstruction Blanz04 [**D**], self-occlusion handling with visible contour landmarks Qu14 [**D**] and linear modeling of generalization error Aldrian10 [**D**]. Additionally, integration of Qu14 into Aldrian10 makes it possible to build a strong baseline, referred to as Aldrian10+Qu14. On the basis of the error curves w.r.t. yaw angles plotted in Fig. 5, clearly "U"-shaped curves are seen in the cases of Blanz04 and Aldrian10, which do not take into account the correspondence problem of contour landmarks at all and fail at large angles as expected. By employing simple occlusion handling, Qu14 shows improved performance for these cases, which is again outperformed by Aldrian10+Qu14, which is the best on BFM among all compared methods. By comparison, the proposed algorithm adaptively models all contour features and achieves pose-invariant 3D face shape recovery with the lowest and most stable errors across all tested poses, even for frontal faces thanks to our flexible contour fitting approach.

To help understand the curves in Fig. 5, qualitative fitting results are illustrated in Fig. 6. 3D shape error is rendered as skin texture using heat maps on the reconstructed faces respectively. Blanz04 fails to recover the facial form starting from already  $10^{\circ}$  of yaw rotation.



Figure 6: Renderings of BFM sample face No. 4 and the heat maps of reconstruction error using Blanz04, Qu14 and the proposed method from top to bottom in the respective rows.

Qu14 undergoes less performance degradation with increasing head poses and plausible shapes are generated at 30°. Nevertheless, reconstruction quality of both the outer area and the inner structure is still heavily limited by not leveraging the valuable self-occluded information and the flawed fixed correspondence on the facial contour. In contrast, superior and constant performance invariant to pose changes is achieved by the proposed method, which conforms to the quantitative evaluation in Fig. 5.

## 5 Conclusions

This paper revisits the general framework of automatic 3D face shape reconstruction from 2D facial features and argues the importance of properly modeling the entire contour landmarks. Instead of using individual landmark positions, we exploit the connected curve feature leveraging DT and LM-ICP, rendering our fitting algorithm flexible to tolerate discrepancy of 2D–3D correspondence, yet constrained enough to achieve robustness along the facial contour independent of pose variation. On the other hand, fast detection of silhouette vertices allows us to keep the computational cost of the complex optimization process at a very low level. Promising reconstruction results outperforming state-of-the-art approaches justifies its theoretical and practical advances. In future work, we will explore more constraints and texture features to improve the landmark-based 3D shape reconstruction framework.

## Acknowledgment

This study was partially supported by the MisPel project, co-funded by the German Federal Ministry of Education and Research (BMBF) under grant 13N12063, and by the MobilePass project, co-funded by the European Union under FP7 grant 608016.

### References

- O. Aldrian and W. A. P. Smith. A linear approach of 3D face shape and texture recovery using a 3D morphable model. In *BMVC*, pages 75.1–75.10, 2010.
- [2] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith. Fully automatic pose-invariant face recognition via 3D pose normalization. In *ICCV*, pages 937–944, 2011.
- [3] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In SIG-GRAPH, pages 187–194, 1999.
- [4] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1063–1074, 2003.
- [5] V. Blanz, C. Basso, T. Poggio, and T. Vetter. Reanimating faces in images and video. *Comput. Gr. Forum*, 22(3):641–650, 2003.
- [6] V. Blanz, A. Mehl, T. Vetter, and H.-P. Seidel. A statistical method for robust 3D surface reconstruction from sparse data. In *3DPVT*, pages 293–300, 2004.
- [7] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Syst. J.*, 4 (1):25–30, 1965.
- [8] B. Chu, S. Romdhani, and L. Chen. 3D-aided face recognition robust to expression and pose variations. In CVPR, pages 1907–1914, 2014.
- [9] P. Dou, Y. Wu, S. K. Shah, and I. A. Kakadiaris. Robust 3D face shape reconstruction from single images via two-fold coupled structure learning. In *BMVC*, 2014.
- [10] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama. 3D face recognition under expressions, occlusions, and pose variations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(9):2270–2283, 2013.
- [11] N. Faggian, A. P. Paplinski, and J. Sherrah. Active appearance models for automatic fitting of 3D morphable models. In AVSS, page 90, 2006.
- [12] N. Faggian, A. P. Paplinski, and J. Sherrah. 3D morphable model fitting from multiple views. In FGR, pages 1–6, 2008.
- [13] A. W. Fitzgibbon. Robust registration of 2D and 3D point sets. In *BMVC*, pages 411–420, 2001.
- [14] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2nd edition, 2004. ISBN 0521540518.
- [15] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- [16] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao. Efficient 3D reconstruction for face recognition. *Pattern Recognition*, 38(6):787–798, 2005.

- [17] I. Kemelmacher-Shlizerman and R. Basri. 3D face reconstruction from a single image using a single reference face shape. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(2): 394–405, 2011.
- [18] Y. J. Lee, S. J. Lee, K. R. Park, J. Jo, and J. Kim. Single view-based 3D face reconstruction robust to self-occlusion. *EURASIP J. Adv. Signal. Proces.*, 2012(1):176, 2012.
- [19] I. Matthews, J. Xiao, and S. Baker. 2D vs. 3D deformable face models: Representational power, construction, and real-time fitting. *Int. J. Comput. Vis.*, 75(1):93–113, 2007.
- [20] P. Mortazavian, J. Kittler, and W. Christmas. 3D-assisted facial texture superresolution. In *BMVC*, pages 119.1–119.11, 2009.
- [21] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3D face model for pose and illumination invariant face recognition. In AVSS, pages 296–301, 2009.
- [22] C. Qu, E. Monari, T. Schuchert, and J. Beyerer. Fast, robust and automatic 3D face model reconstruction from videos. In AVSS, pages 113–118, 2014.
- [23] C. Qu, H. Gao, E. Monari, J. Beyerer, and J.-P. Thiran. Towards robust cascaded regression for face alignment in the wild. In *CVPRW*, 2015.
- [24] C. Qu, C. Herrmann, E. Monari, T. Schuchert, and J. Beyerer. 3D vs. 2D: On the importance of registration for hallucinating faces under unconstrained poses. In *CRV*, pages 139–146, 2015.
- [25] H. M. Rara, A. A. Farag, and T. Davis. Model-based 3D shape recovery from single images of unknown pose and illumination using a small number of feature points. In *IJCB*, pages 1–7, 2011.
- [26] S. Romdhani and T. Vetter. Efficient, robust and accurate fitting of a 3D morphable model. In CVPR, pages 59–66, 2003.
- [27] S. Romdhani and T. Vetter. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In *CVPR*, volume 2, pages 986–993, 2005.
- [28] D. Vlasic, M. Brand, H. Pfister, and J. Popović. Face transfer with multilinear models. In SIGGRAPH, pages 426–433, 2005.
- [29] N. Wang, X. Gao, D. Tao, and X. Li. Facial feature point detection: A comprehensive survey. arXiv:1410.1037 [cs.CV], 2014.
- [30] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li. High-fidelity pose and expression normalization for face recognition in the wild. In *CVPR*, pages 787–796, 2015.