

Feature-Preserving Detailed 3D Face Reconstruction from a Single Image

Yue Li

University of Pennsylvania
yueli.cg@gmail.com

Haoqiang Fan

Megvii Inc. (Face++)

ABSTRACT

Dense 3D face reconstruction plays a fundamental role in visual media production involving digital actors. We improve upon high fidelity reconstruction from a single 2D photo with a reconstruction framework that is robust to large variations in expressions, poses and illumination. We provide a global optimization step improving the alignment of 3D facial geometry to tracked 2D landmarks with 3D Laplacian deformation. Face detail is improved through, extending Shape from Shading reconstruction with fitted albedo prior masks, together with a fast proportionality constraint between depth and image gradients consistent with local self-occlusion behavior. Together these measures better preserve the crucial facial features that define an actor's identity, and we illustrate this through a variety of comparisons with related works.

CCS CONCEPTS

- [I.3.7 Computer Graphics]: Three-Dimensional Graphics and Realism → Color, shading, shadowing, and texture;
- [I.4.8 Image Processing and Computer Vision] Scene Analysis → Shading;

KEYWORDS

Feature-Preserving, 3D Face Reconstruction, Optimization

ACM Reference Format:

Yue Li, Liqian Ma, Haoqiang Fan, and Kenny Mitchell. 2018. Feature-Preserving Detailed 3D Face Reconstruction from a Single Image. In *CVMP '18: European Conference on Visual Media Production (CVMP '18), December 13–14, 2018, London, United Kingdom*, Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 10 pages. <https://doi.org/10.1145/3278471.3278473>

1 INTRODUCTION

The problem of 3D face reconstruction from stills and video is a hot research topic across Computer Vision and Computer Graphics

This work was done when Yue Li was an intern at Megvii Research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CVMP '18, December 13–14, 2018, London, United Kingdom

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-6058-6/18/12...\$15.00
<https://doi.org/10.1145/3278471.3278473>

Liqian Ma

Megvii Inc. (Face++)

Kenny Mitchell

Edinburgh Napier University

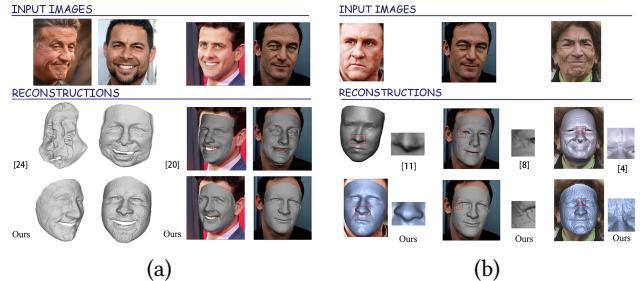


Figure 1: Our single image to 3D reconstruction results compared with related referenced works. Robustness and feature-preservation are two major components of 3D reconstruction. (a) demonstrates robustness under a varied poses, and (b) retaining detail of the actors' facial identity.

with foundations in face recognition [2, 30] and face alignment [13, 17, 36], as well as core applications in face reenactment [29], Animoji [8], 3D relighting, avatar [11, 16], and augmented reality [4], etc.. The aim of face reconstruction is to reproduce facial geometry features that give strong depiction of one's identity. Although reconstruction from video and multiple views typically have higher reconstruction accuracy [5], single frame methods are still of great importance in image based applications [12, 14, 20].

As 3D face reconstruction from a single image is an ill-posed problem, the key ingredient in a reconstruction algorithm is the face shape prior model. However, this entails a difficult trade-off between robustness and detail preservation, especially in the presence of complex real-world illuminations, large pose variations, expressions or occlusions.

In our experiments, we find current state-of-the-art works are either insufficiently robust [25], fail to capture fine facial features [9, 21], or overly smooth thin geometry details [5] (See Fig. 1)). In this paper, we propose a redesigned 3D face reconstruction framework that enhances robustness and improves the ability to preserve fine detailed face identity features. Specifically, the main contributions of this paper are:

1. A global shape correction stage enhancing the alignment between the coarse geometry and the 2D facial landmark features through 3D Laplacian deformation.
2. Improving beyond Shape-From-Shading approaches, we provide an iterative optimization procedure, introducing a masked albedo prior to improve detail recovery that would otherwise be lost to albedo estimation.

3. A faster method incorporating explicit consideration of self-occlusion for fine grained geometry reconstruction.

4. The resulting system comprises a robust reconstruction pipeline that handles variations in large expressions, poses and illuminations.

2 RELATED WORKS

2.1 Face Prior

Earlier research provides reliable priors of face shape, expression and albedo. Blanz and Vetter [2] presents 3D Morphable Model(3DMM) as a low-dimensional representation of the geometry and albedo of faces, which is obtained via principle component analysis (PCA) on registered scanned real 3D faces. FaceWarehouse [6] builds a bi-linear face model that encodes 150 individual identities and 47 expressions. Booth et al [3] create a large-scale 3DMM from 10,000 scanned faces from a huge variety of the human population, and further creates domain-specific 3DMM. The application of this representation includes but not limited to: 3D face reconstruction [9, 21], face recognition [2, 30], face alignment [13, 17, 36] and face reenactment [29]. These priors are generally helpful to obtain rough face geometry and albedo, but it cannot model the crucial face features to describe the refined identity of a person, for example, the shape of eyes, the nose, or the mouth, and face details, like wrinkles and folds.

2.2 3D Face Reconstruction

3D Face reconstruction focusing on single image [12, 21, 25, 26, 28, 31] or image sequence [5, 8, 9, 22, 23] is not a new topic. Many are based on 3DMM face priors [5, 8, 9, 12, 21, 25], whilst Trigeorgis et al [33] directly solve face normals with fully convolutional networks. Other recent CNN approaches [9] perform reconstruction in real-time sacrificing some identity detail. Although some unsupervised methods exist [21, 28], a CNN is often trained in a supervised way, except where synthetic data [9, 20] are employed. While deep learning approaches require careful training design and appropriate datasets whether real or synthetic, they can be thought of as more generalistic perhaps being lossy to a person's uniqueness and identity, alternatively analytic methods can be applied generally without training data dependence and directly perform to observed principles, and therefore fall into the focus for this work.

Current state-of-the-art 3D face reconstruction results remain susceptible to artifacts. The most prevalent work flow (fit 3DMM coefficients, and then add displacement map or bump map) [9] produces large silhouettes and component matching errors (Fig. 3), since low-rank representation of faces only produce the rough geometry. [12] embed a medium layer for shape correction, but the solution is not perfect (Fig. 8). Some works tend to capture the most visibly apparent fine-scale details only [5], and some algorithms are less robust under various lighting conditions or occlusions [25].

In this paper, we redesign the 3D face reconstruction framework with a focus on practical facial identity recovery. Our algorithm is feature preserving, and robust under complex lighting environment as well as a large range face 3D poses.

In wider context beyond the scope of our work, there are other literatures focusing on 3D reconstruction of hair [10], glasses [18], or other reconstruction algorithms which are robust to occlusion [32].

3 OVERVIEW

Our framework is illustrated in the Fig. 2 pipeline diagram flowing from left to right. 3DMM shape, expression and albedo coefficients, camera, and lighting parameters are solved in the coarse level referred as *initialization*, as discussed in Section 5.1. Given the coarse geometry of the initialized 3DMM shape, we employ Laplacian deformation to achieve optimal positions for all vertices of the rough mesh while fixing landmark points as soft constraints. We refer to this step as *correction* in Section 5.2 since it corrects the geometry to share more resemblance of the input image rather than limited to the low-level subspace. The final detail is reconstructed by firstly performing intrinsic imgae decomposition to the input which results in albedo, residual and shading map. Together with the medium level depth map, we use shape-from-shading technique to solve the final detailed depth map in Section 5.3.

4 BACKGROUND

4.1 3D Face Morphable Model

3DMM [2] is a widely-used face model comprising a low-dimension statistical representation of rough face geometry and albedo. Formally, 3D face geometry and albedo can be expressed as:

$$S = S_\mu + P_{id}\alpha + P_{exp}\gamma, \quad (1)$$

$$T = T_\mu + P_{alb}\beta \quad (2)$$

where S and T are 3D face geometry and albedo, S_μ and T_μ are shape and albedo matrices of the average 3D face respectively, P_{id} , P_{exp} and P_{alb} are the principle axes trained on a set of scanned 3D Faces by PCA. α , β and γ are the corresponding coefficient vectors which characterize the identity/albedo/expression information of a certain person.

4.2 Perspective Projection

We use a weak perspective projection to project 3D Face onto the image plane:

$$V_{3D} = M_{4 \times 4} * S, \quad (3)$$

$$V_{2D} = f * Pr * V_{3D} + t_{2D} \quad (4)$$

$M_{4 \times 4}$ is the model-view matrix that transforms S in world space, which can be expressed by 3D translation vector, uniform scale, and rotation angles yaw, pitch and roll. V_{3D} is the transformed face in world space. f , Pr , and t_{2D} are the projection scaling factor, orthographic projection matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, and 2D translation vector, respectively. V_{2D} denotes the 2D position of the shape on image plane.

4.3 Rendering

Following [9], we assume the faces have Lambertian reflectance, and approximate the global illumination by second-order spherical harmonics basis functions, then the final rendering color can be written as:

$$P = L \cdot \phi(N_{3D}) \cdot \rho \quad (5)$$

where N_{3D} and ρ are the per-pixel mesh normal and albedo after rasterization. L is the spherical harmonics coefficients, $\phi(N_{3D})$ is

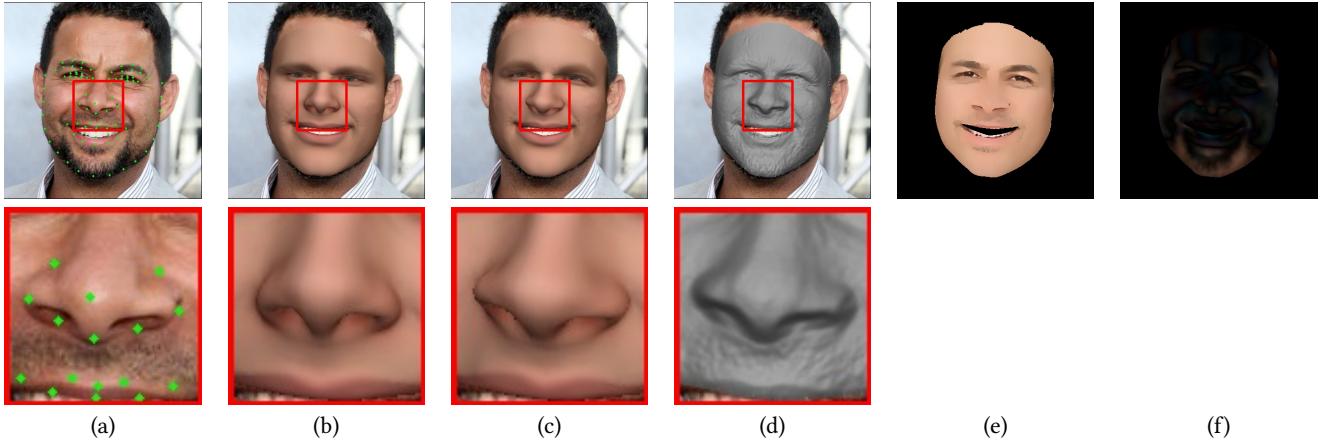


Figure 2: Algorithm pipeline. (a) is the input image, green dots indicate face landmarks from 2D detectors. (b) shows the face shape and albedo after initialization step (3DMM fitting). (c) shows the face shape after the correction step, salient features are preserved (nose shape, etc.) (d) is the resulting face after the detail reconstruction step. (e) and (f) are resulting face albedo and residual map, respectively. Detailed geometry information is faithfully preserved (wrinkles, etc.).

the spherical harmonics basis for N_{3D} . $L \cdot \phi(N_{3D})$ is also called the shading term.

In particular, we use the famous Basel Face Model (BFM [19]) for $S_\mu, P_{id}, T_\mu, P_{alb}$, and FaceWarehouse [6] for P_{exp} . From above, all the parameters that characterize a 2D face image are $\{\alpha, \beta, \gamma, M_{4x4}, f, Pr, t_{2D}, L\}$.

5 FEATURE-PRESERVING FACE INVERSE RENDERING

5.1 Initialization

We fit $\{\alpha, \beta, \gamma, M_{4x4}, f, Pr, t_{2D}, L\}$ from 2D face image as the initialization step of the whole algorithm. The problem is well addressed in many works [9, 36]. In this paper, the objective function is based on [9]:

$$E = E_{consistency} + w_{landmark}E_{landmark} + w_{reg}E_{reg}. \quad (6)$$

Here $E_{consistency}$ is the per-pixel L2 distance between source image and the rendered image, $E_{landmark}$ is the 2D distance in image space between face landmarks detected from source image and landmarks calculated from the current parameters, E_{reg} is the regularization term which restricts the final 3D face around the 3DMM mean face, $w_{landmark}$ and w_{reg} are the weighting parameters. We simply inherit all the parameter settings from [9]. The optimization is solved via Gauss-Newton approach.

Note that to produce all the results in our paper, the face landmark points from the source image are automatically detected, and mesh landmark vertices are beforehand manually annotated on the BFM surface. To make the optimization robust under large pose variations, we update vertex indices of the face contour landmarks on the BFM surface after each optimization iteration, since these landmarks should be restricted to the face silhouette, as in most 2D image landmark detectors.

5.2 Face Shape Correction

3DMM fitting restricts resulting the 3D face to the PCA linear subspace and is often over-regularized to get visual-acceptable results, however, a real person's face shape has its unique shape features in nose, mouth, and contour areas, which are very important for face and expression recognition. As before, we note these fail to be described sufficiently by current 3DMMs.

The resulting landmark positions from 2D detectors and from 3DMM fitting parameters do not align very well, shown in Fig. 3. Silhouettes matching is a very hard problem in 3DMM fitting. In our experiments, we find that 2D image landmark detectors can generate much more accurate landmark positions (especially along silhouettes) than subspace shape modeling. To solve silhouette matching and shape preservation then, we propose a simple algorithm that explicitly uses 2D landmark information from detectors and successfully corrects the face shape obtained from 3DMM fitting.

We define 2D landmark positions from detectors as V_{2D}^M . After 3DMM fitting, S, V_{3D}, V_{2D} of landmark vertices can be calculated via Equ.1,3 and 4. The shape correction problem can be formulated as a 3D mesh deformation problem with all the landmark vertices defined as control points. The deformation should transform these 3D control points from V_{3D} to V_{3D}^M , and V_{3D}^M must satisfy:

$$V_{2D}^M = f * Pr * V_{3D}^M + t_{2D}. \quad (7)$$

However V_{3D}^M is still ambiguous for solving since it is only a 2D constraint, thus we must also constrain its distance to the camera (z-coordinate in camera space). Heuristically, we find the following simple constraint is enough to produce good results:

$$V_{3D} \cdot z = V_{3D}^M \cdot z \quad (8)$$

Here $\cdot z$ means fetching z component from a 3D vector. That means, for landmark vertices, the deformation changes the 2D

projection position on image plane, but preserves its distance to the camera. After calculating V_{3D}^M , we apply Laplacian Deformation [27] to get the deformed mesh, S_M .

Experimental results (Fig. 3) show that this step greatly reduces silhouettes and landmark matching errors, while capturing global shape features.

The formulation of our deformation approach is similar to [34], but we are addressing two different issues. They try to move the tracked mesh based on constrained vertices, but not too much along the scene flow to avoid destroying the integrity. On the contrary, we want the landmarks to move exactly to their target positions while maintaining a reasonable global mesh structure. During our evaluations we find that despite the errors of 2D landmark detectors, we still provide better results. 2D landmark detector error is the lower bound of fitting error, thus, it is rather meaningful to minimize such errors utilizing deformation. The error caused by deformation from inaccurate landmarks is much less than those from over-regulated 3DMM coefficients both qualitatively (Fig. 6) and quantitatively (Table. 1).

5.3 Face Detail Reconstruction

Similar to [9], our face detail reconstruction algorithm is based on a Shape from Shading framework (SFS [7]). Given face image, I , we estimate a depth map, D , that captures fine-scale surface details such as wrinkles, an albedo map, A , that captures albedo details and a residual map, B , that captures smooth highlights and smooth shadows which our lighting model cannot explain. Hence, I , A , B and D should satisfy:

$$I = A \cdot S + B, \quad (9)$$

$$S = L \cdot Depth2Normal(D) \quad (10)$$

where S indicates the shading map, and $Depth2Normal(D)$ transforms depth map to normal map (see [7]). Given rasterization from 3DMM fitting and our face shape correction results, we can get an initial albedo map A^* , shading map S^* and depth map D^* .

Previous algorithms cannot model real fine-scale surface details well, as demonstrated in Fig. 6. Some algorithms tend to capture only visibly apparent fine-scale features [9], while some algorithms tend to create too much fine-scale geometry [25]. Generally, each of these algorithms are insufficient to capture real face identity features perfectly. We carefully redesign each component of the framework, and propose a novel, identity feature-preserving solution.

5.3.1 Solving Albedo Map. Unlike previous works, we utilize an albedo prior map, A^* , to inform our albedo estimation, A . We assume albedo is almost constant across the skin region, whilst we can confidently capture high-frequency albedo details in some sensitive regions like eyes, nose and mouth.

To reflect this, we manually prepare a smooth mask in a face texture space map indicating regions of high confidence of uniform skin albedo (Fig. 4). After rendering the texture on fitted 3DMM using fitted camera matrix, we get conforming skin mask, M . Following [7], we solve albedo map A and residual beta map B by minimizing the following equations via iteration:

$$\min_A ||AS^* + B - I||^2 + q_s^A \sum_{j \in \mathcal{L}_i} \omega_j^i (A_i - A_j) + q_p^A M ||A - A^*||^2; \quad (11)$$

$$\min_B ||AS^* + B - I||^2 + q_s^B \sum_{j \in \mathcal{L}_i} \omega_j^i (B_i - B_j). \quad (12)$$

The first term (data term) of Equ.11 and Equ.12 ensures the correctness of such decomposition, the second term (smooth term) of Equ.11 and Equ.12 makes resulting albedo map and beta map smooth, here ω_j^i is the color affinity between pixel i and pixel j (in pixel i 's neighborhood, notated as \mathcal{L}_i), which is given more weights when neighboring pixels are similar. The third term of Equ.11 restricts the resulting albedo map A to albedo prior map A^* , and is only valid when current pixel is within the masked skin region M .

In our implementation, the number of iterations is 3. q_s^A , q_p^A and q_s^B are relative weights and we set $q_s^A=3.0$, $q_s^B=2.0$ for all iterations, and $q_p^A=2.0$, 1.0 , 0.0 in each successive iteration. The iteration gradually transfers parts of the image that our lighting model failed to explicitly explain into the residual map, which results in less artifacts during our final depth reconstruction. The relaxing of the albedo prior constraint gives data term more weight to be optimized in the last iteration, when residual map already consists mostly of smooth highlights and shadows of I .

5.3.2 Solving Depth Map. After calculating A and B , we intend to solve fine-scale depth map D . The objective function is defined as follows:

$$\min_z \{Loss_{data} + Loss_{prior} + Loss_{smooth} + Loss_{occlusion}\}. \quad (13)$$

where the *data* term ensures the correctness of such decomposition:

$$Loss_{data} = ||L \cdot Depth2Normal(D) \cdot A + B - I||^2. \quad (14)$$

the *priori* term restricts resulting depth map D close to initial depth map D^* :

$$Loss_{prior} = q_p^D ||D - D^*||^2. \quad (15)$$

and the *smooth* term controls resulting depth map smoothing behavior:

$$Loss_{smooth} = q_s^D ||\Delta D||^2. \quad (16)$$

lastly, the *occlusion* term explicitly measures self-occlusion, which is a crucial component in appearance, but not accounted for by our spherical harmonics lighting model. As shown in Fig. 11, deep wrinkles are affected heavily by ambient occlusion, which makes the image darker in the wrinkle crevices. Inspired by [1], we define our occlusion term with the following to conform image gradient to depth gradient:

$$Loss_{occlusion} = q_o^D ||\nabla I - r_o^D \nabla D||^2 \quad (17)$$

where ∇ denotes the gradient vector of an image. This is based on the observation that in a very small patch (for example 3px*3px) of I , both albedo and depth can be plausibly assumed to be constant, thus gradients of image color are only due to self-occlusion

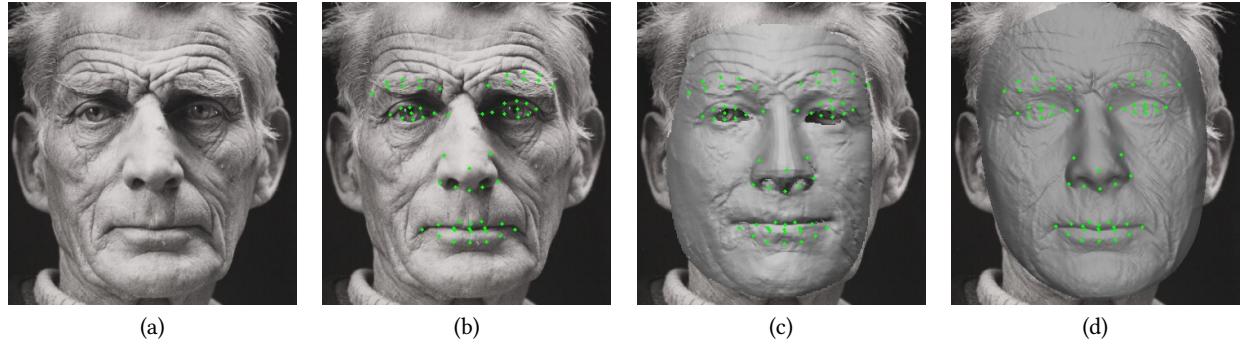


Figure 3: Landmark fitting demonstration. (a) input image, (b) shows the landmark points detected, (c) the result of [9], and (d) is our result. One can observe the necessity of the deformation, given that our reconstruction preserves more identity information of the input image.

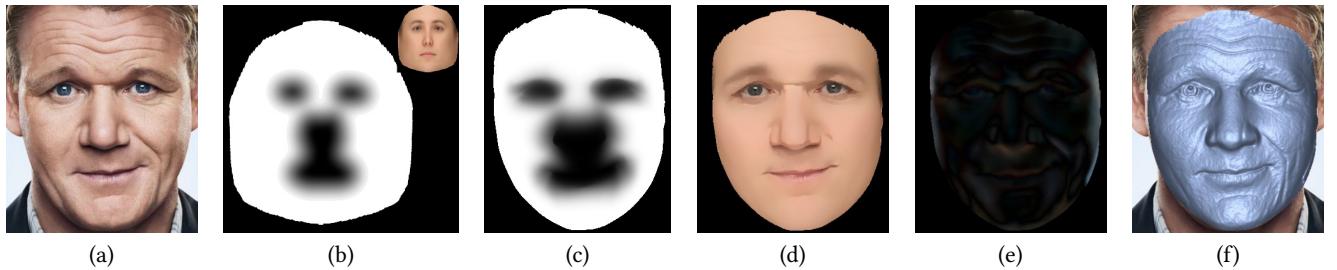


Figure 4: (a) input image, (b) is the mask prior in unwrapped UV space (c) is the skin mask M after fitting, (d) and (e) are our albedo map and residual map after our last iteration of the optimization respectively, and (f) is the final reconstruction

(Fig. 11). Theoretically, self-occlusion is caused by the variance of depth. Here we directly restrict the gradients of D to be proportional to the gradients of I , and in our experiments we find that this simple modification makes fine geometry more promising, albeit not metrically correct [1]. The proportionality coefficient r_o^D can be manually adjusted to control the scale of fine geometry. In our implementation, we set $q_p^D=2.5$, $q_s^D=30$, $q_o^D=1.0$, $r_o^D=1.0$ to generate all the results. A slower, more complex way of handling self-occlusion could be modeling the self-occlusion in a sub-space in advance [24].

6 COMPARISONS AND RESULTS

6.1 Results

We analyze our algorithm with a set of challenging face photos, including large variations in pose, exaggerated expressions, and complex illuminations. For all the images listed below, our results capture the local details at different levels, meanwhile, match closely to the global shape of the input image (Fig. 5).

6.2 Comparisons

6.2.1 Qualitative Comparisons. We compare our results qualitatively with those reconstructed by the state-of-the-art detail preserving methods. In Fig. 6 we compare results of ours, [21], and [9]. As shown in Fig. 6 our reconstructions demonstrate better geometry fitting and facial identity detail preservation.

We also outperform the results of Sela et al [25] (Fig. 7), which do provide impressive geometry details, however, the depth map produced by the neural network is not robust to beards or self-shadowed illumination. Hence, their registration step results in missing or distorted face region geometry.

We then compare our reconstruction with Cao et al [5] (Fig. 8b), showing only recovery of the most visual salient wrinkles, and Jiang et al [12] (Fig. 7e) where severe distortion at the nose and missing major details can be seen.

6.2.2 Quantitative Comparisons. We ran our method, along with [32], [37], and [9] on the neutral faces of 100 subjects from the BU-3DFE dataset [35]. We use dlib [15] to detect landmarks, and compute the mean absolute error to the ground truth mesh on facial pixels. Quantitative results are shown in Table 1.

Table 1: Quantitative comparisons: We demonstrate better reconstructions compared to other state-of-the-art methods

	[32]	[37]	[9]	Ours
Mean Absolute Error (MAE)	19.594	17.397	14.154	12.923

6.3 Analysis

Figure 9 demonstrates the critical importance of the medium level face shape correction. (b) and (e) show that our reconstruction fails



Figure 5: Demonstration of our algorithm under different poses and lighting conditions. Columns (a) and (c) are the input images, (b) and (d) are our reconstruction results. Note that our framework can handle a large range of pose angles, strong highlights and extreme non-symmetrical, non-uniform lighting

to represent the input image faithfully since the coarse level fitting does not match the landmark points perfectly. A notable improvement could be seen in (c) and (f) by solving optimal geometric positions of the coarse level mesh, utilizing Laplacian Deformation.

Fig. 10 indicates the necessity of integrating the albedo prior as a soft constraint through the optimization process. Compared to

(b), (c) presents better detail preserving since wrinkles remain in the albedo map without the leveraging the prior.

Fig. 11 shows that the gradient term is necessary since self-shadows are partially included in the residual map and results in less satisfactory recovery in the final reconstruction.



Figure 6: Column (a) are the input images, (b) are results of [21], (c) are results of [9], (d) our results.

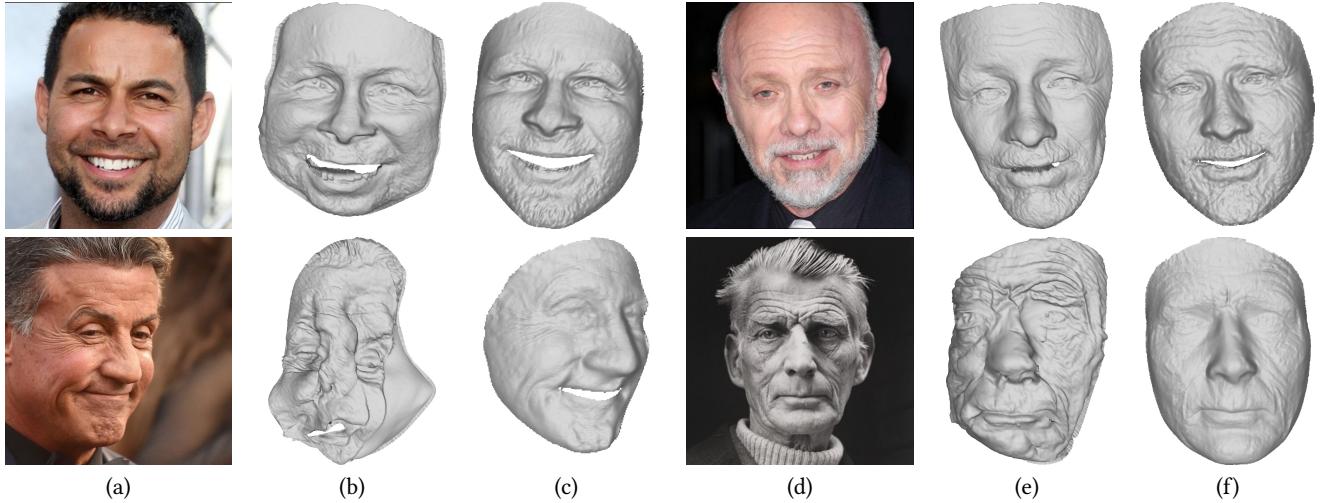


Figure 7: Robustness Comparisons. (a) and (d) are input images, (b) and (e) are results of [25], (c) and (f) are our reconstructions

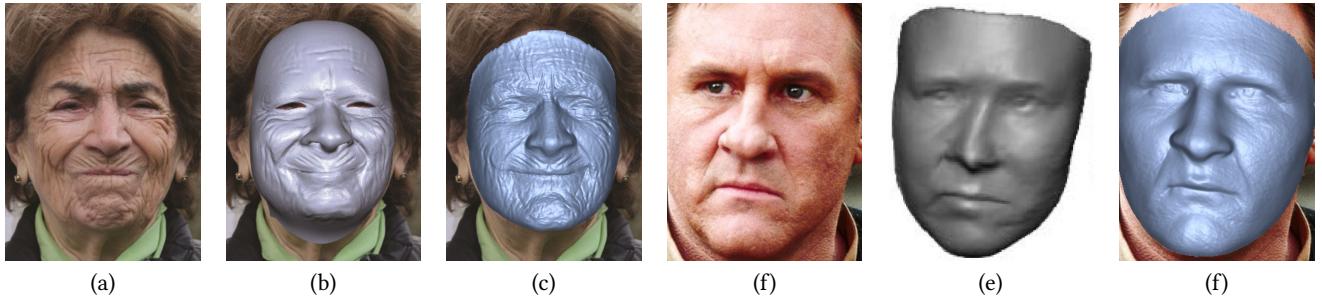


Figure 8: (a) and (f) are input images, (b) is results of [5], (c) is the results of [12] (d) and (e) are our results

6.4 Limitation

We draw the skin mask because eyes, nose and mouth area contains more identity information, which 3DMM priori have large fitting errors. Since it is manually drawn on the priori's UV map, facial hairs are still blended into the shading when dealing with different real images.

7 CONCLUSIONS

In this paper, we propose a robust, facial identity feature preserving, 3D face reconstruction algorithm. Our novel global shape correction step, provides robustness under large face pose variations and expressions, meanwhile, preserving the shape at nose, eyes, and mouth regions as well as face silhouettes. Since we employ an iterative optimization procedure to separate albedo from geometry and consider self-occlusion when optimizing detailed geometry in the fine level, our algorithm can produce high quality reconstruction under complex lighting conditions, and different poses.

8 ACKNOWLEDGEMENTS

We thank the author of [9] for providing the source images and their results, [32, 37] for making public their codes, and the reviewers for their advises.

REFERENCES

- [1] Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. 2010. High-quality single-shot capture of facial geometry. In *ACM Transactions on Graphics (ToG)*, Vol. 29. ACM, 40.
- [2] V. Blanz and T. Vetter. 2003. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 9 (2003), 1063–1074.
- [3] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway. 2016. A 3D Morphable Model Learnt from 10,000 Faces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5543–5552.
- [4] Dan A. Caiyan, Jean-François Lalonde, Paulo Gotardo, Tomas Simon, Iain Matthews, and Kenny Mitchell. 2018. From Faces to Outdoor Light Probes. *Computer Graphics Forum* (2018). <https://doi.org/10.1111/cgf.13341>
- [5] Chen Cao, Derek Bradley, Kun Zhou, and Thabo Beeler. 2015. Real-time High-fidelity Facial Performance Capture. *ACM Trans. Graph.* 34, 4, Article 46 (July 2015), 9 pages.
- [6] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou. 2014. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics* 20, 3 (2014), 413–425.
- [7] R. Or El, G. Rosman, A. Wetzler, R. Kimmel, and A. M. Bruckstein. 2015. RGBD-fusion: Real-time high precision depth recovery. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5407–5416. <https://doi.org/10.1109/CVPR.2015.7299179>
- [8] Pablo Garrido, Michael Zollhöfer, Dan Casas, Levi Valgaerts, Kiran Varanasi, Patrick Pérez, and Christian Theobalt. 2016. Reconstruction of Personalized 3D Face Rigs from Monocular Video. *ACM Trans. Graph.* 35, 3 (May 2016), 28:1–28:15.
- [9] Y. Guo, J. Zhang, J. Cai, B. Jiang, and J. Zheng. 2017. 3DFaceNet: real-time dense face reconstruction via synthesizing photo-realistic face images. *arXiv preprint arXiv:1708.00980* (2017).

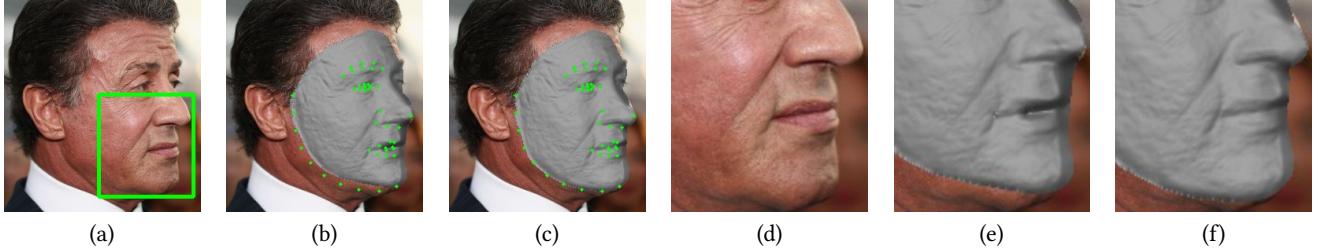


Figure 9: Demonstration of Laplacian deformation. (a) is the input image, (b) our reconstruction without the medium level correction, (c) is our final result with the deformation, (d), (e) and (f) are the zoomed regions.

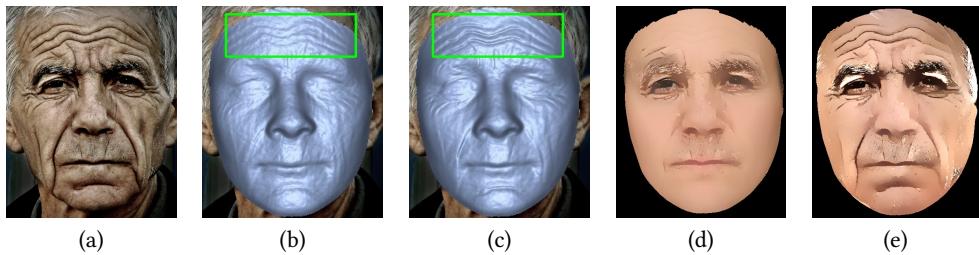


Figure 10: Effect of the fitted albedo prior. (a) is the input image, (b) is the result without prior, (c) is the result incorporating the prior, (d) shows the albedo solved with prior, and (e) shows the solved albedo without prior.

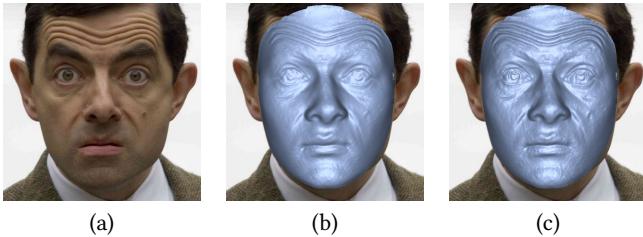


Figure 11: Demonstration of the proportionally constrained image and depth gradients. (a) input, (b) and (c) are reconstruction results with alternative weights on the gradient term.

- [10] Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. 2015. Single-view hair modeling using a hairstyle database. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 125.
- [11] Liwen Hu, Shunsuke Saito, Lingyu Wei, Koki Nagano, Jaewoo Seo, Jens Fursund, Iman Sadeghi, Carrie Sun, Yen-Chun Chen, and Hao Li. 2017. Avatar digitization from a single image for real-time rendering. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 195.
- [12] Luo Jiang, Juyong Zhang, Bailin Deng, Hao Li, and Ligang Liu. 2017. 3D Face Reconstruction with Geometry Details from a Single Image. *arXiv preprint arXiv:1702.05619* (2017).
- [13] A. Jourabloo and X. Liu. 2016. Large-pose face alignment via cnn-based dense 3d model fitting. *IEEE Conference on Computer Vision and Pattern Recognition* (2016), 4188–4196.
- [14] Ira Kemelmacher-Shlizerman and Ronen Basri. 2011. 3D face reconstruction from a single image using a single reference face shape. *IEEE transactions on pattern analysis and machine intelligence* 33, 2 (2011), 394–405.
- [15] Davis E. King. 2009. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10 (2009), 1755–1758.
- [16] Hao Li, Thibaut Weise, and Mark Pauly. 2010. Example-Based Facial Rigging. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2010)* 29, 3 (July 2010).
- [17] F. Liu, D. Zeng, Q. Zhao, and X. Liu. 2016. Joint face alignment and 3d face reconstruction. *European Conference on Computer Vision* (2016), 545–560.

- [18] Fabio Maninchedda, Martin R Oswald, and Marc Pollefeys. 2017. Fast 3D Reconstruction of Faces with Glasses. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 4608–4617.
- [19] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. 2009. A 3d face model for pose and illumination invariant face recognition. *IEEE International Conference on Advanced video and signal based surveillance* (2009), 296–301.
- [20] Elad Richardson, Matan Sela, and Ron Kimmel. 2016. 3D face reconstruction by learning from synthetic data. In *3D Vision (3DV), 2016 Fourth International Conference on*. IEEE, 460–469.
- [21] Elad Richardson, Matan Sela, Roy Or-El, and Ron Kimmel. 2017. Learning Detailed Face Reconstruction from a Single Image. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 5553–5562.
- [22] Joseph Roth, Yiyi Tong, and Xiaoming Liu. 2014. Unconstrained 3D face reconstruction. *Trans. Graph.* 33, 4 (2014), 43.
- [23] Joseph Roth, Yiyi Tong, and Xiaoming Liu. 2016. Adaptive 3D face reconstruction from unconstrained photo collections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4197–4206.
- [24] Andreas Schneider, Sandro Schönborn, Lavrenti Frobeen, Thomas Vetter, and Bernhard Egger. 2017. Efficient global illumination for morphable models. (2017).
- [25] Matan Sela, Elad Richardson, and Ron Kimmel. 2017. Unrestricted facial geometry reconstruction using image-to-image translation. In *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 1585–1594.
- [26] Soumyadip Sengupta, Angjoo Kanazawa, Carlos D Castillo, and David W Jacobs. [n. d.]. SFSNet: Learning Shape, Reflectance and Illuminance of Faces 'in the wild'. [n. d.].
- [27] Olga Sorkine, Daniel Cohen-Or, Yaron Lipman, Marc Alexa, Christian Rössl, and H-P Seidel. 2004. Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*. ACM, 175–184.
- [28] Ayush Tewari, Michael Zollhöfer, Hyeongwoo Kim, Pablo Garrido, Florian Bernard, Patrick Pérez, and Christian Theobalt. 2017. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In *The IEEE International Conference on Computer Vision (ICCV)*, Vol. 2, 5.
- [29] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner. 2016. Face2face: Real-time face capture and reenactment of rgb videos. *IEEE Conference on Computer Vision and Pattern Recognition* (2016), 2387–2395.
- [30] A. T. Tran, T. Hassner, I. Masi, and G. Medioni. 2016. Regressing robust and discriminative 3d morphable models with a very deep neural network. *arXiv preprint arXiv:1612.04904* (2016).
- [31] Anh Tuan Tran, Tal Hassner, Iacopo Masi, and Gérard Medioni. 2017. Regressing robust and discriminative 3D morphable models with a very deep neural network. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 1493–1502.

- [32] Anh Tuan Tran, Tal Hassner, Iacopo Masi, Eran Paz, Yuval Nirkin, and Gerard Medioni. 2017. Extreme 3D Face Reconstruction: Looking Past Occlusions. *arXiv preprint arXiv:1712.05083* (2017).
- [33] George Trigeorgis, Patrick Snape, Iasonas Kokkinos, and Stefanos Zafeiriou. 2017. “Face normals” in-the-wild” using fully convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 38–47.
- [34] Levi Valgaerts, Chenglei Wu, Andrés Bruhn, Hans-Peter Seidel, and Christian Theobalt. 2012. Lightweight binocular facial performance capture under uncontrolled lighting. *ACM Trans. Graph.* 31, 6 (2012), 187–1.
- [35] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J Rosato. 2006. A 3D facial expression database for facial behavior research. In *Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on*. IEEE, 211–216.
- [36] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Li. 2016. Face alignment across large poses: A 3d solution. *IEEE Conference on Computer Vision and Pattern Recognition* (2016), 146â€“155.
- [37] Xiangyu Zhu, Zhen Lei, Junjie Yan, Dong Yi, and Stan Z Li. 2015. High-fidelity pose and expression normalization for face recognition in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 787–796.