创 新 工 场
人工智能工程院
SINOVATION VENTURES AI INSTITUTE

A survey on 3D face reconstruction

Zhixiong Zuo

2019.4.4

*"**3D reconstruction** is the process of capturing the **geometry** and **appearance** of real objects."*

--from Wikipedia

Geometry Reconstruction

- Active methods

Structured light (Kinect, iPhone X)

Laser range finder

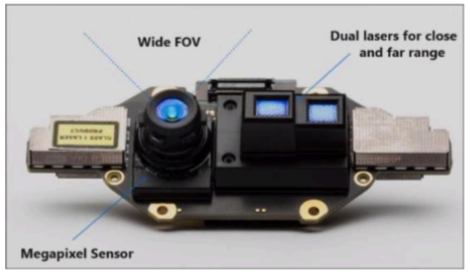
- Passive methods

Multi-view Stereo

Monocular cues methods

Data-driven

Kinect 2.0



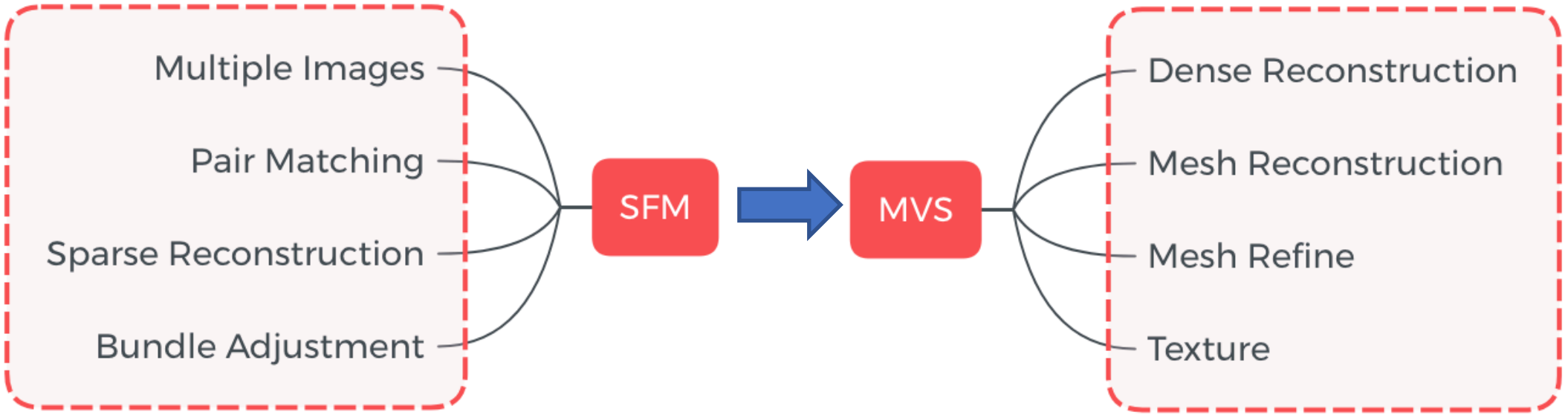
ToF of Microsoft

iPhone X

Outline

- Multi-view reconstruction
- Structured light reconstruction
- Model-based 3D face reconstruction
- End-to-end 3D face reconstruction

Multi-view reconstruction



Multi-view reconstruction

Open source

- OpenMVG+MVS
- VisualSFM+MVS
- Colmap
- Meshroom

Disadvantage

- GPU
- Time consuming

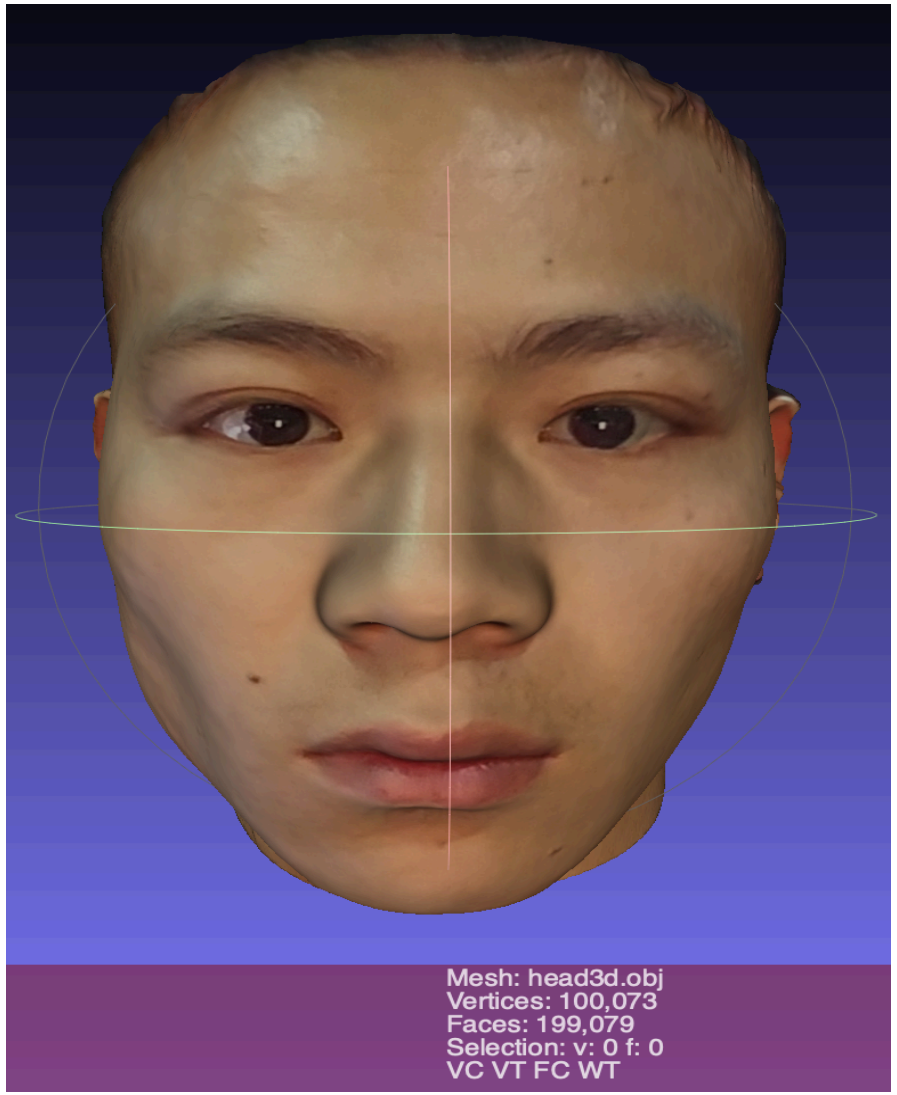
https://zhuanlan.zhihu.com/p/34256509

40 picturres+960*540+face segmentation=16582vertices 5min

Structured light reconstruction

Bellus3D Face Camera

100000+vertices，20s



Model-based 3D face reconstruction

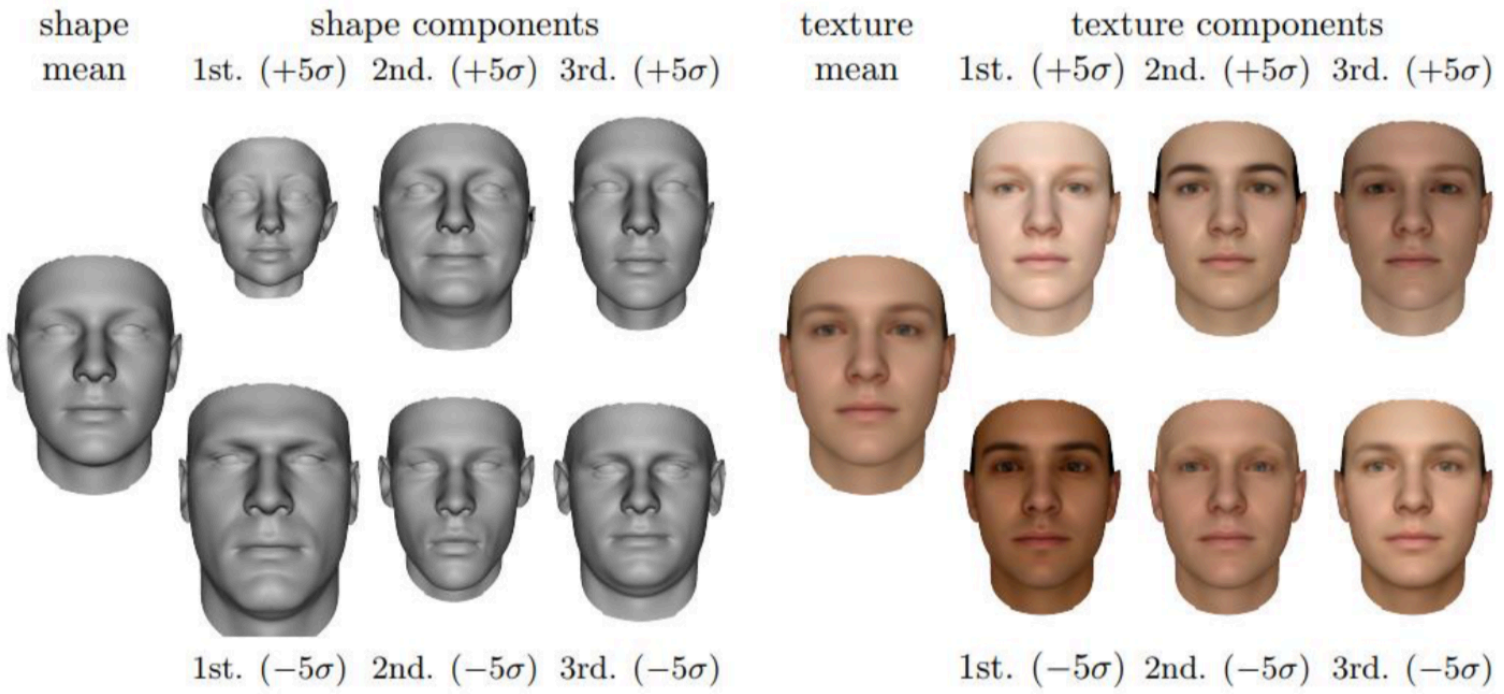
3D Morphable Model

- Introduction
- Generation
- fitting

3DMM

$$G = \bar{G} + U_{id}\alpha_{id} + U_{exp}\alpha_{exp}$$

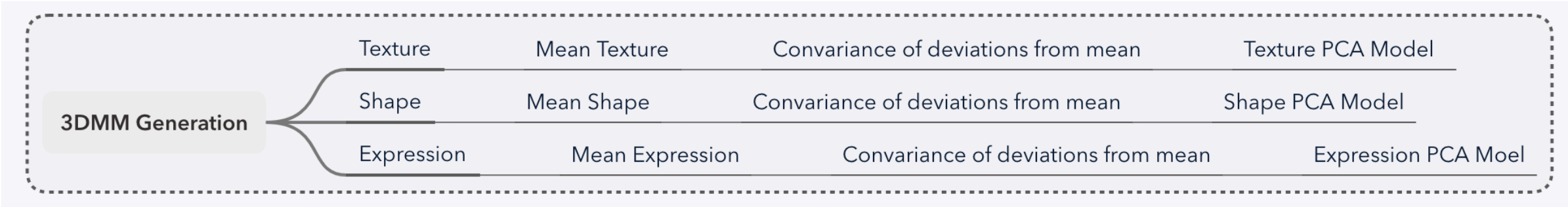
$$T = \bar{T} + U_{tex}\alpha_{tex}$$



Paysan P, Knothe R, Amberg B, et al. A 3D face model for pose and illumination invariant face recognition[C]//2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. Ieee, 2009: 296-301.

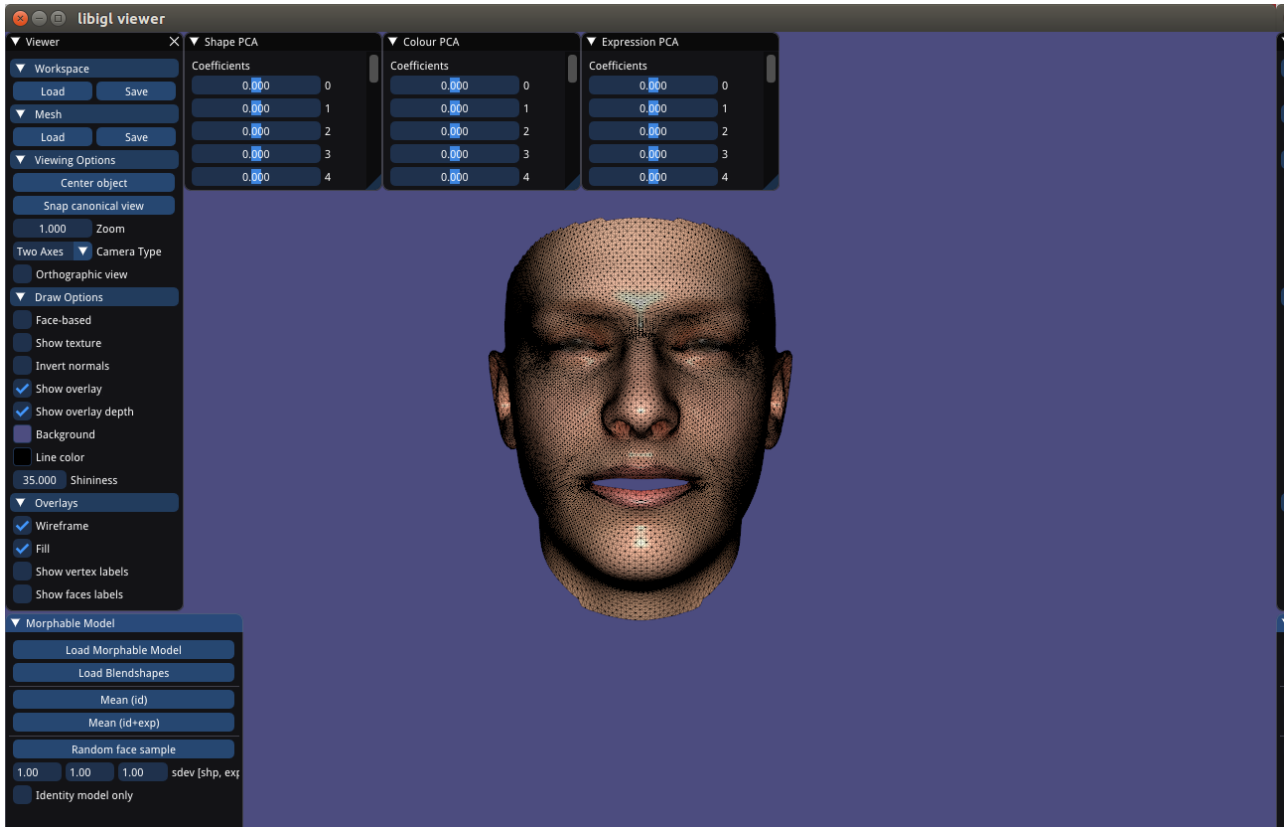
3DMM Generation

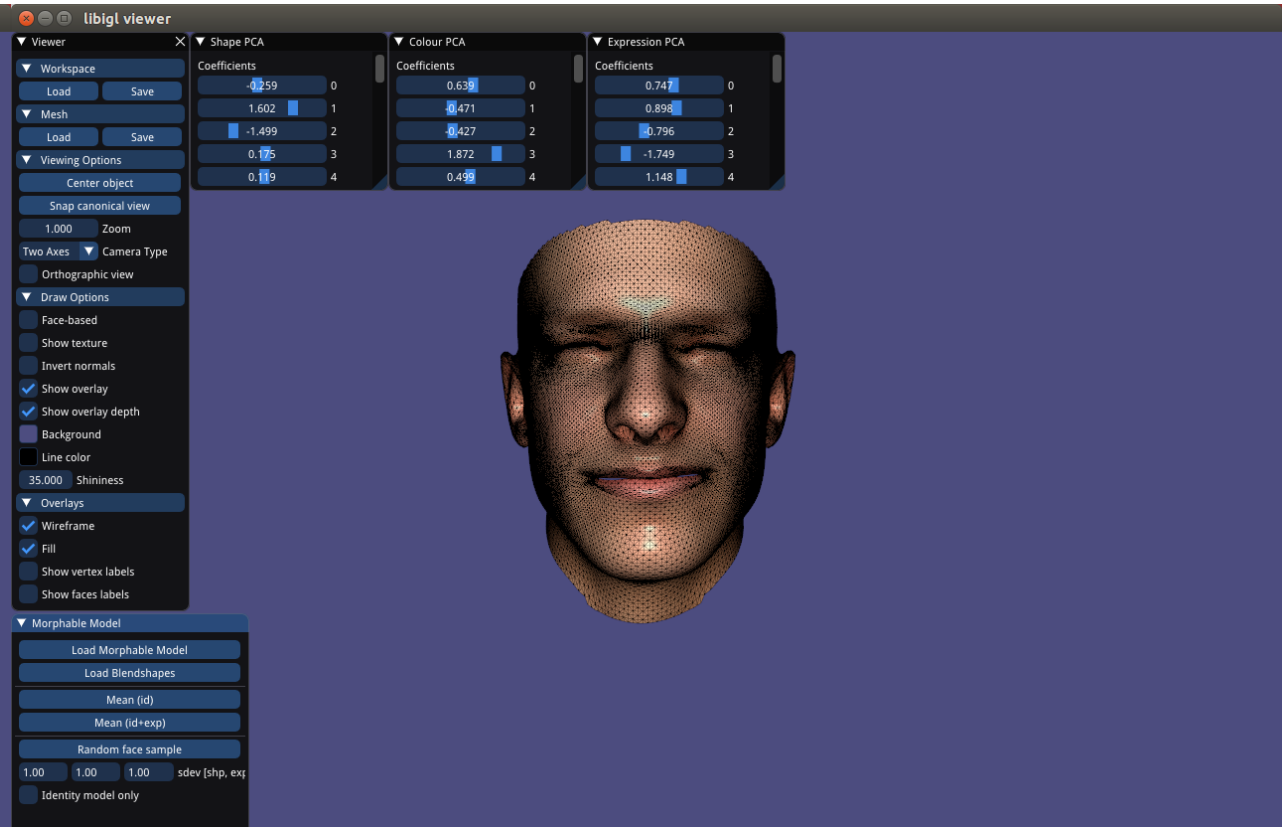
1. The target scan is denoised using Gaussian and median filtering if spikes and noise are present.
2. Perform a global mapping from the generic model to the target scan using facial landmarks, smoothly deforming the generic model.
3. Do a local matching on the current resolution level based on the distances between reference and target vertices. If a particular vertex cannot be matched, its mirrored counterpart is used (and if that fails as well, the algorithm interpolates us- ing the neighbouring matches).
4. The final set of matches guides an energy min- imisation process that conforms the model to the target scan. Steps 3 and 4 are iterated.
5. The generic face model is subdivided using the 4- 8 mesh subdivision algorithm (Velho and Zorin, 2001).
6. Steps 3 to 5 are repeated until the desired highest mesh resolution is achieved.



Huber P, Hu G, Tena R, et al. A multiresolution 3d morphable face model and fitting framework[C]//Proceedings of the 11th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. 2016.

3DMM Generation





https://github.com/patrikhuber/eos-model-viewer

3DMM

- Basel Face Model
  - BFM2009 (53490vertices)
  - BFM2017 (53149vertices)
- Surrey Face Model
  - SFM1724
  - SFM3448
  - SFM16759 (No texture)
  - SFM29587 (No texture)
- Large scale Face Model

SFM

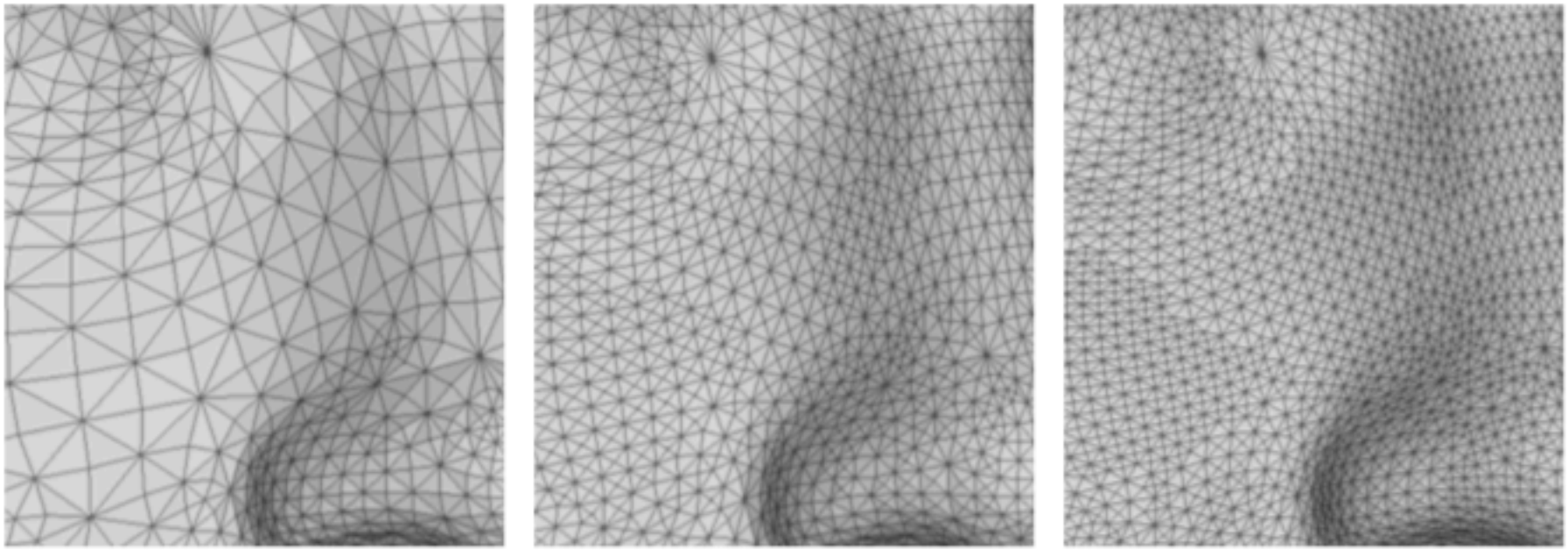


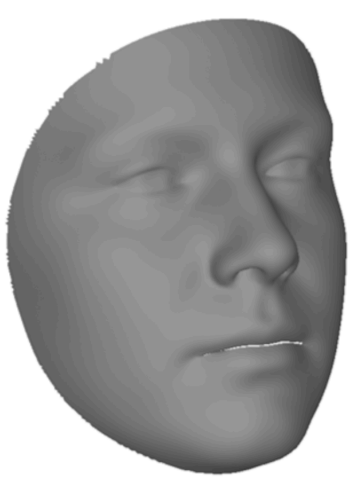
Figure 3: Close-up of the different mesh resolutions of the Surrey Face Model. *(left)*: The low-resolution model with 3448 vertices. *(middle)*: Medium-resolution model (16759 vertices) *(right)*: The full resolution model (29587 vertices).

A Multiresolution 3D Morphable Face Model and Fitting Framework, P. Huber, G. Hu, R. Tena, P. Mortazavian, W. Koppen, W. Christmas, M. Rätsch, J. Kittler, International Conference on Computer Vision Theory and Applications (VISAPP) 2016, Rome, Italy

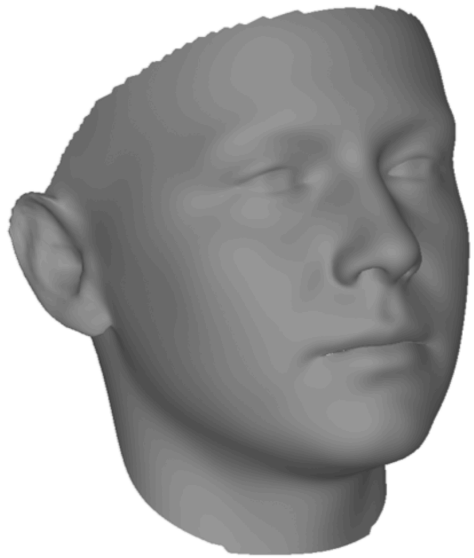
LSFM

- The filename always takes the form `GENDER_ETHNICITY_AGE{_cropped}.mat`
- Our most powerful, generic model, is therefore `all_all_all.mat`
- The number of components retained is always the minimal sufficient to retain **99.7% of the variance** in the model's training set
- Among all possible combinations of age, gender and ethnicity, we retain only those ones for which sufficient training data exists in our database to build a useful model

Filename	Gender	Ethnicity	Age	Region	Subjects	Components
`all_all_18-50.mat`	All	All	18 to 50	Full	5106	183
`all_all_18-50_cropped.mat`	All	All	18 to 50	Cropped	5106	280
`all_all_7-18.mat`	All	All	7 to 18	Full	2029	132
`all_all_7-18_cropped.mat`	All	All	7 to 18	Cropped	2029	194



Cropped facial region



Full facial region

Facial Region	No. Vertices	No. Triangles
Full	53215	105840
Cropped	28431	56272

https://github.com/menpo/lsfm/blob/master/MODELS.md

3DMM Fitting

1. The shape and texture parameters of the reconstructed 3D face are initialized to 0.
2. According to the shape and texture parameters to generate a 3D face.
3. Project 3D face into 2D image.
4. Calculate 2D face similarity obtained from 2D image face and 3D face projection.
5. Optimize shape and texture parameters with optimization algorithm.
6. Repeat steps 2 to 5 until the most similar face is obtained by the projection of 2D face and 3D face.

3DMM Fitting

Real-time 3D Face Super-resolution From Monocular In-the-wild Videos

SIGGRAPH 2016 poster id 220

Patrik Huber[1], William Christmas[1], Matthias Rätsch[2], Adrian Hilton[1], Josef Kittler[1]

[1]Centre for Vision, Speech and Signal Processing
University of Surrey
United Kingdom

[2]Image Understanding and Interactive Robotics
Reutlingen University
Germany

Contact: www.patrikhuber.ch

eos

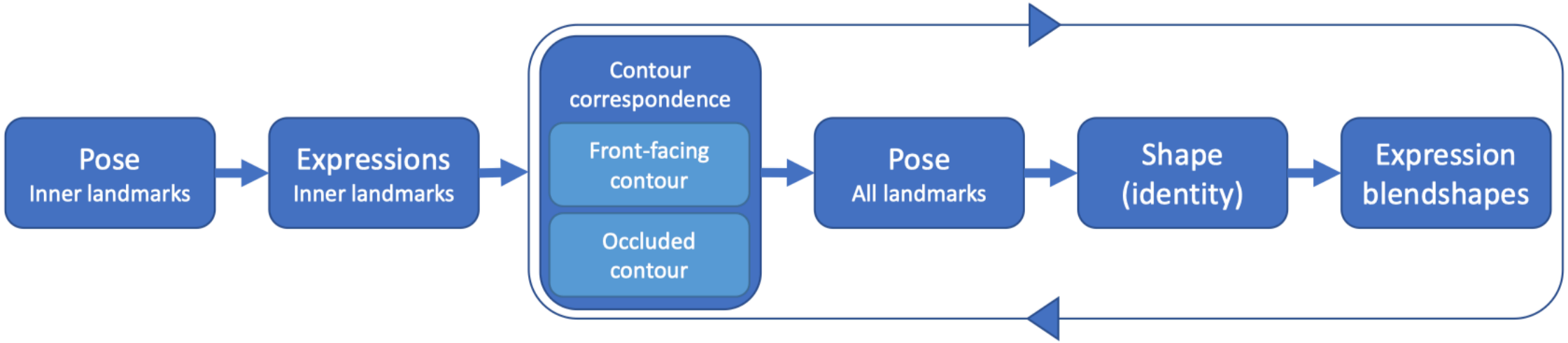


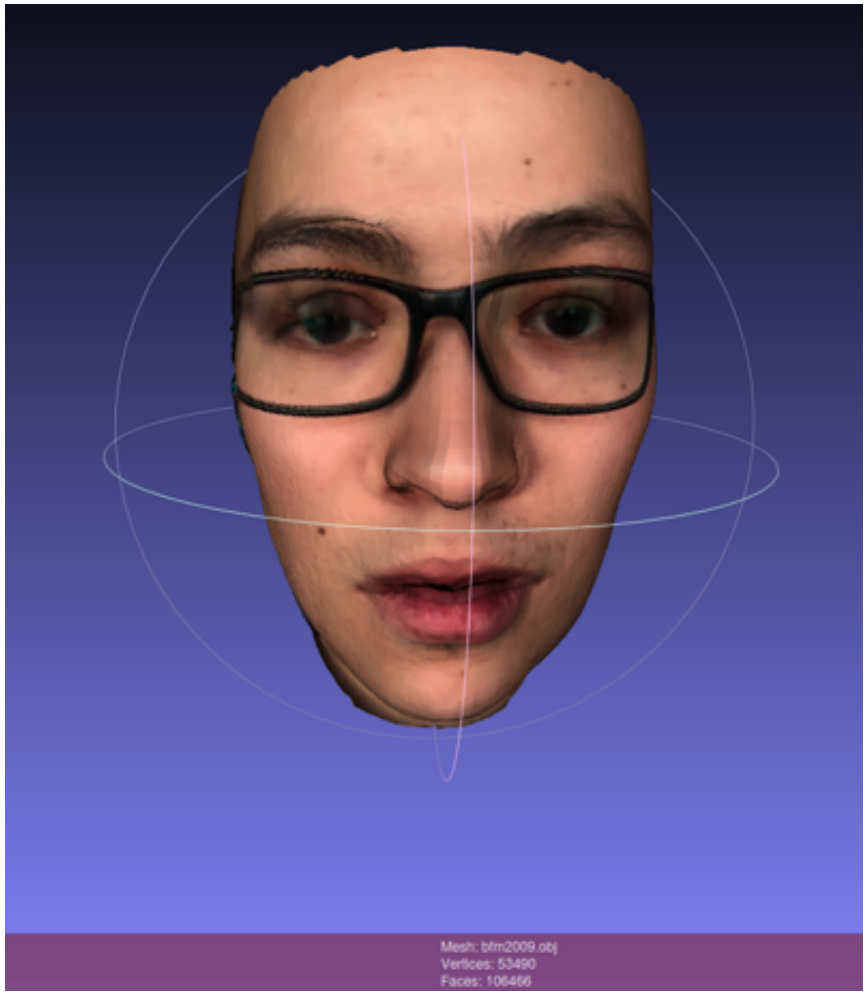
Figure 4.7: Flowchart of the iterative linear fitting algorithm. The fitting is initialised with a rough pose and expression fit, and then proceeds alternating contour, pose, identity and expression fitting.

https://github.com/patrikhuber/eos

Huber, Patrik (2017) *Real-time 3D morphable shape model fitting to monocular in-the-wild videos.* Doctoral thesis, University of Surrey.

eos

Multi-reconstruction
5 frames with SFM3448



1 frame with BFM2009

4dface

End-to-end 3D face reconstruction

- 3D DENSE FACE ALIGNMENT (3DDFA)
- Position Map Regression Network
- Nonlinear 3D Face Morphable Model
- Face Model Learning from Videos（CVPR2019 oral）

3DDFA

$$\mathbf{S} = \overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp},$$

where $\mathbf{S}$ is a 3D face, $\overline{\mathbf{S}}$ is the mean shape, $\mathbf{A}_{id}$ is the principle axes trained on the 3D face scans with neutral expression and $\boldsymbol{\alpha}_{id}$ is the shape parameter, $\mathbf{A}_{exp}$ is the principle axes trained on the offsets between expression scans and neutral scans and $\boldsymbol{\alpha}_{exp}$ is the expression parameter. In this work, the $\mathbf{A}_{id}$ and $\mathbf{A}_{exp}$ come from BFM [53] and FaceWarehouse [54] respectively. After the

Zhu X, Liu X, Lei Z, et al. Face alignment in full pose range: A 3D total solution[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 41(1): 78-92.

3DDFA

$$V(\mathbf{p}) = f * \mathbf{Pr} * \mathbf{R} * (\overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp}) + \mathbf{t}_{2d}$$

where $V(\mathbf{p})$ is the model construction and projection function, leading to the 2D positions of model vertices, f is the scale factor, $\mathbf{Pr}$ is the orthographic projection matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, $\mathbf{R}$ is the rotation matrix and $\mathbf{t}_{2d}$ is the translation vector. The collection of all the model parameters is $\mathbf{p} = [f, \mathbf{R}, \mathbf{t}_{2d}, \boldsymbol{\alpha}_{id}, \boldsymbol{\alpha}_{exp}]^{\mathrm{T}}$.

Zhu X, Liu X, Lei Z, et al. Face alignment in full pose range: A 3D total solution[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 41(1): 78-92.

3DDFA

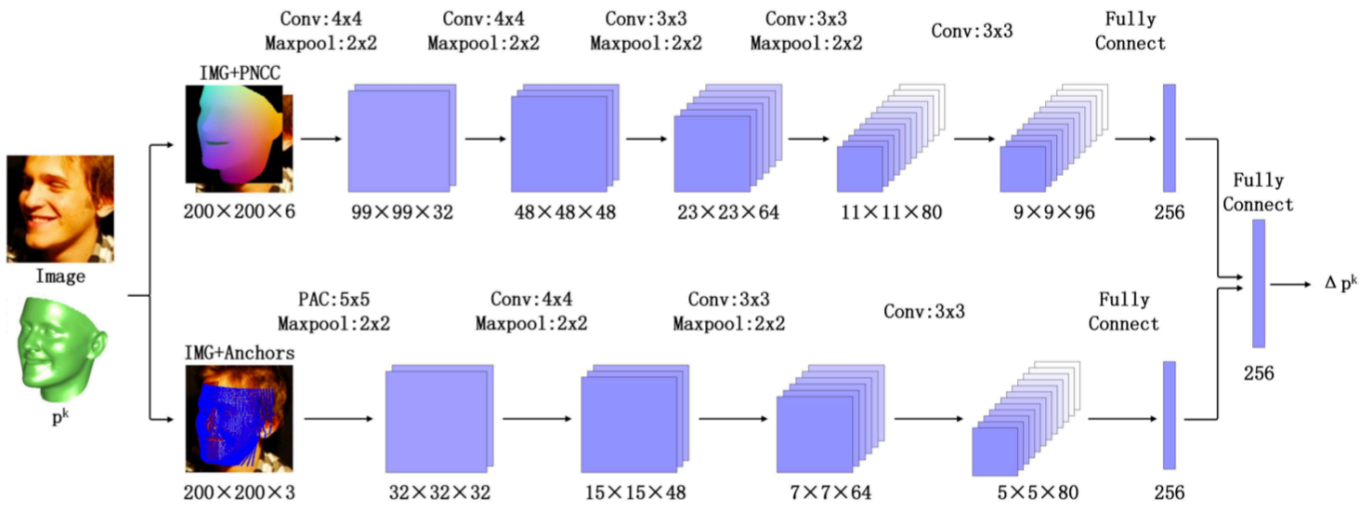
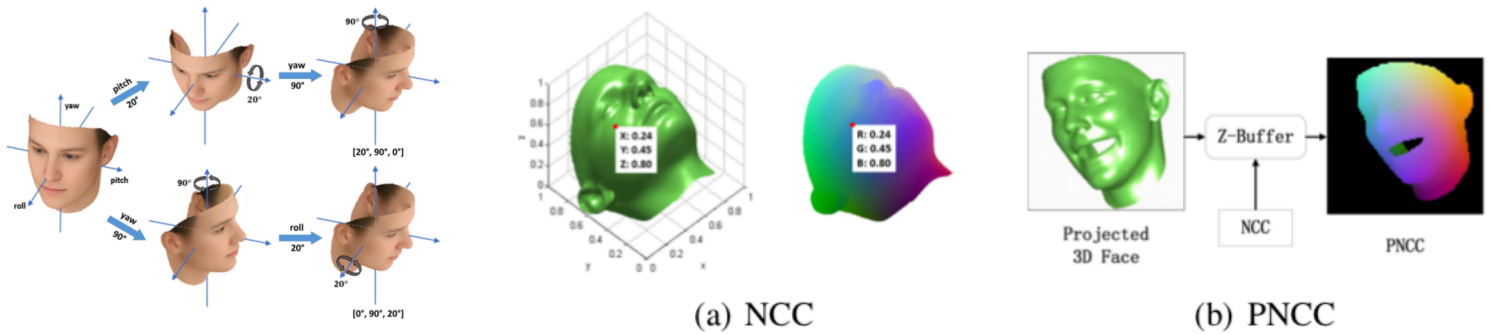
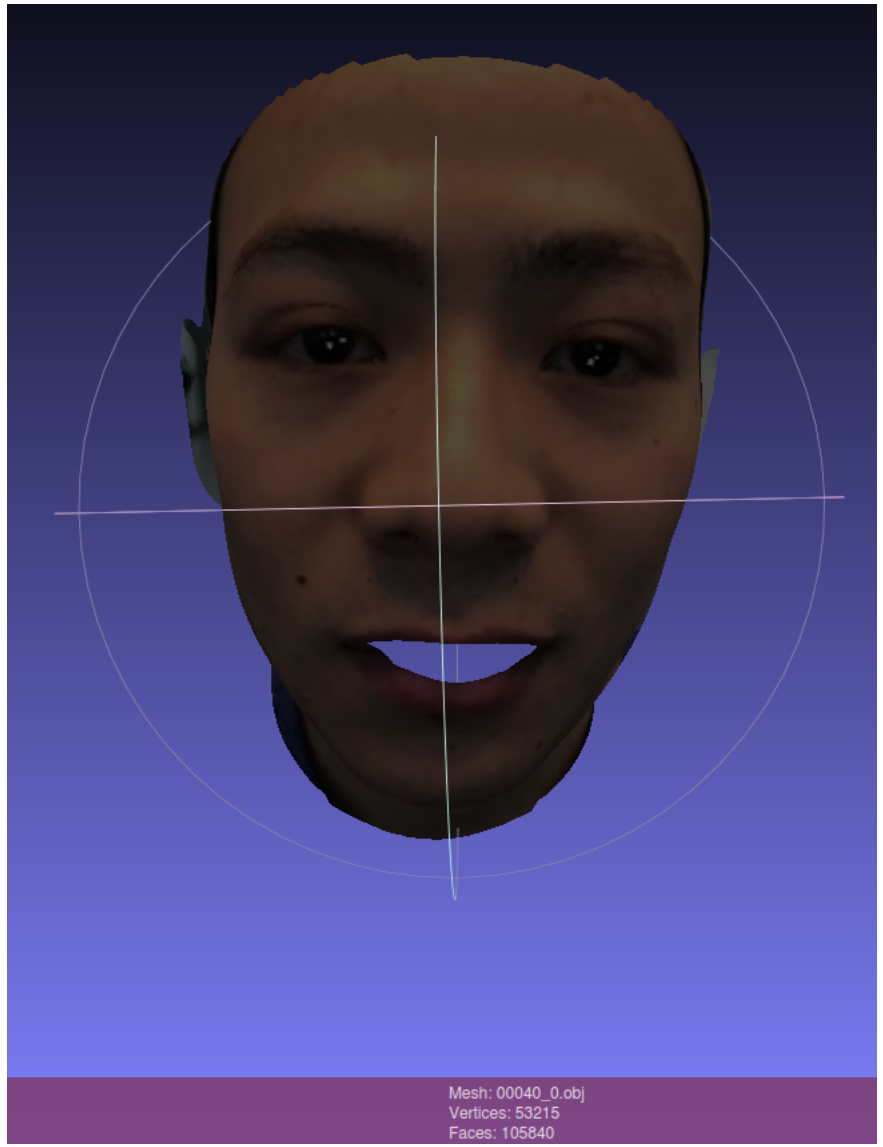


Fig. 2. An overview of the two-stream network in 3DDFA. With an intermediate parameter $\mathbf{p}^k$, in the first stream we construct a novel Projected Normalized Coordinate Code (PNCC), which is stacked with the input image and sent to the CNN. In the second stream, we get some feature anchors with consistent semantics and conduct Pose Adaptive Convolution (PAC) on them. The outputs of the two streams are merged with an additional fully connected layer to predict the parameter update $\Delta\mathbf{p}^k$.





Zhu X, Liu X, Lei Z, et al. Face alignment in full pose range: A 3D total solution[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 41(1): 78-92.
https://github.com/cleardusk/3DDFA

PRNet

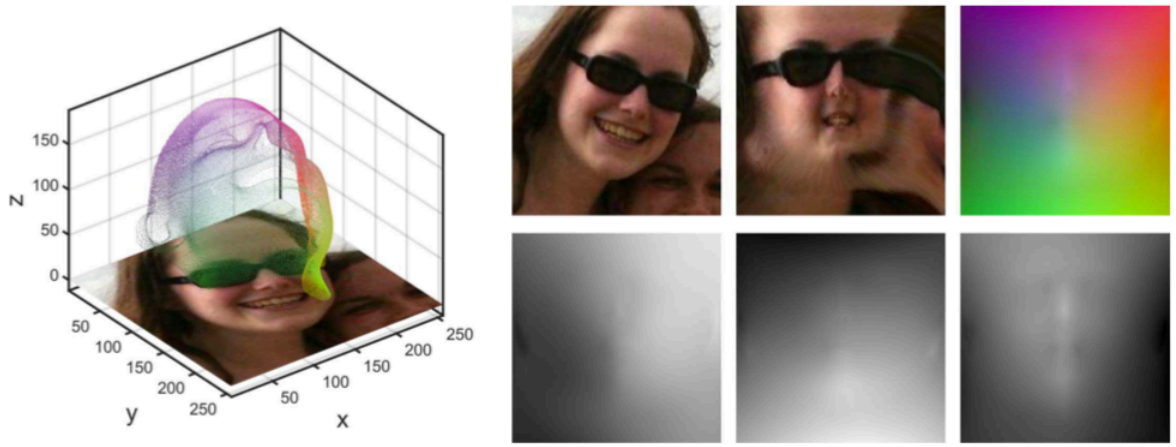


Fig. 2: The illustration of UV position map. Left: 3D plot of input image and its corresponding aligned 3D point cloud(as ground truth). Right: The first row is the input 2D image, extracted UV texture map and corresponding UV position map. The second row is the x, y, z channel of the UV position map.

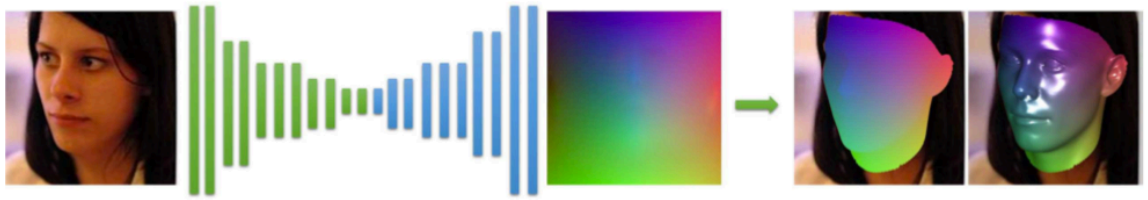
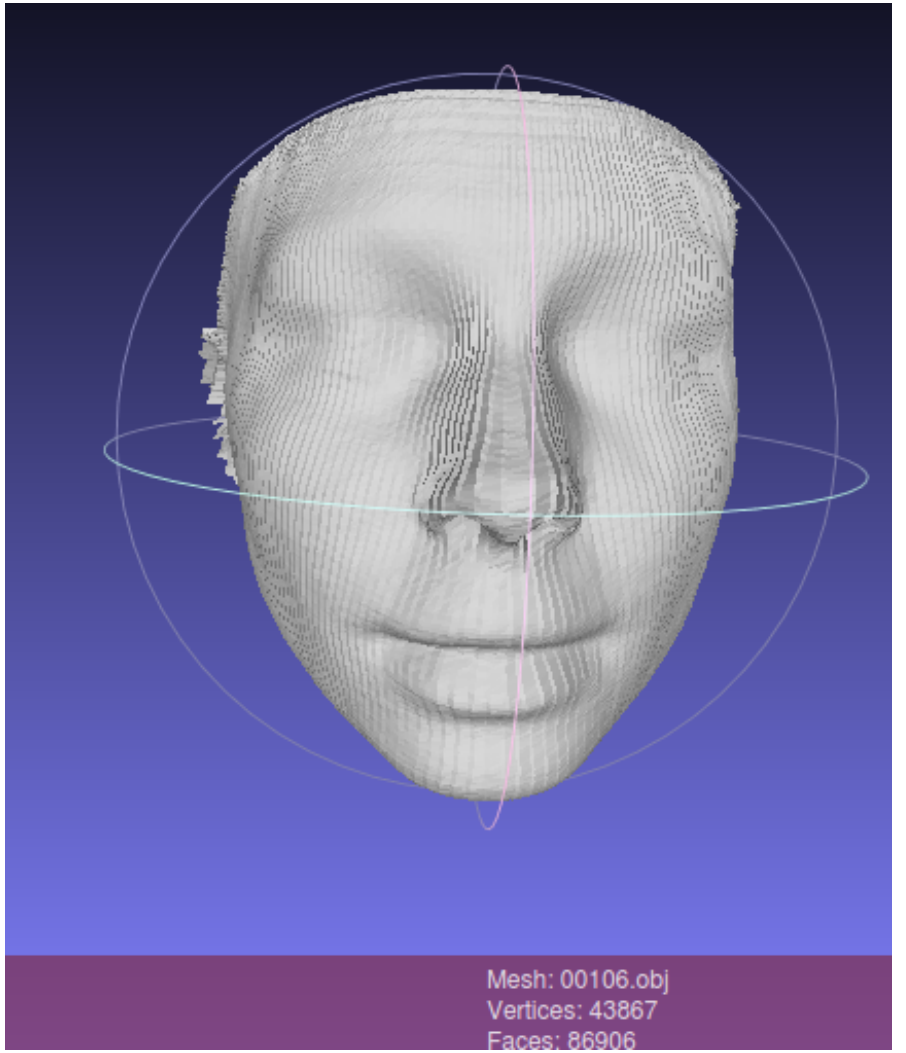


Fig. 3: The architecture of PRN. The Green rectangles represent the residual blocks, and the blue ones represent the transposed convolutional layers.



https://github.com/YadiraF/PRNet

Feng Y, Wu F, Shao X, et al. Joint 3d face reconstruction and dense alignment with position map regression network[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 534-551.

Nonlinear 3D Face Morphable Model

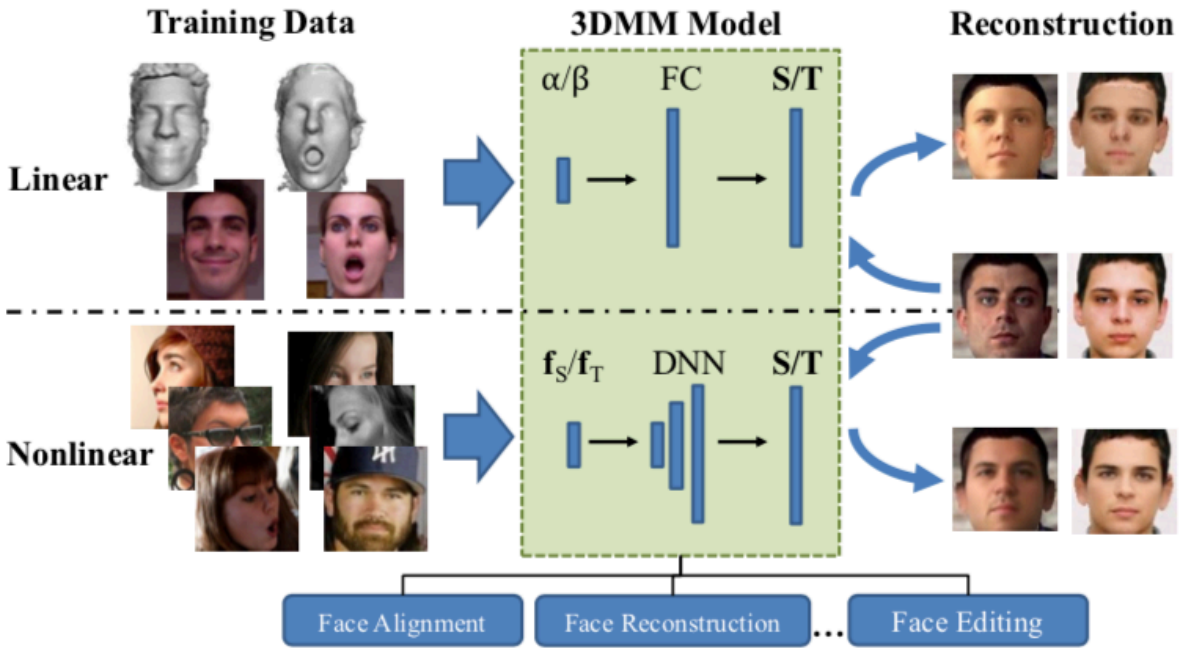


Figure 1: Conventional 3DMM employs linear bases models for shape/texture, which are trained with 3D face scans and associated controlled 2D images. We propose a nonlinear 3DMM to model shape/texture via deep neural networks (DNNs). It can be trained from in-the-wild face images without 3D scans, and also better reconstructs the original images due to the inherent nonlinearity.

https://github.com/tranluan/Nonlinear_Face_3DMM

Tran L, Liu X. Nonlinear 3D face morphable model[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7346-7355.

Face Model Learning from Videos

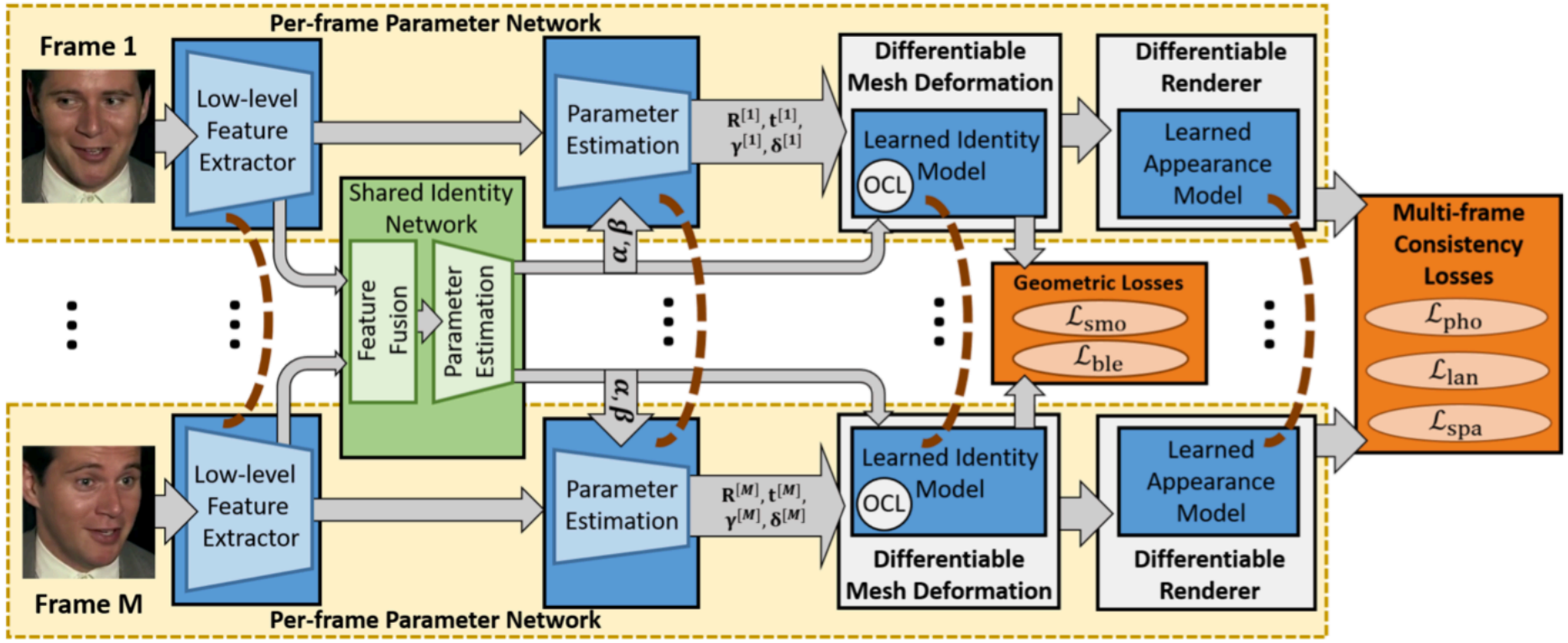


Figure 2. Pipeline overview. Given multi-frame input that shows a person under different facial expression, head pose, and illumination, our approach first estimates these parameters per frame. In addition, it jointly obtains the shared identity parameters that control facial shape and appearance, while at the same time learning a graph-based geometry and a per-vertex appearance model. We use a differentiable mesh deformation layer in combination with a differentiable face renderer to implement a model-based face autoencoder.

https://www.youtube.com/watch?v=4ylQQfLP4cQ

Thank You