3D Reconstruction and Appearance Modeling for Human Face

Rebirth in Deep Neural Network

Tencent Youtu

Chen Li

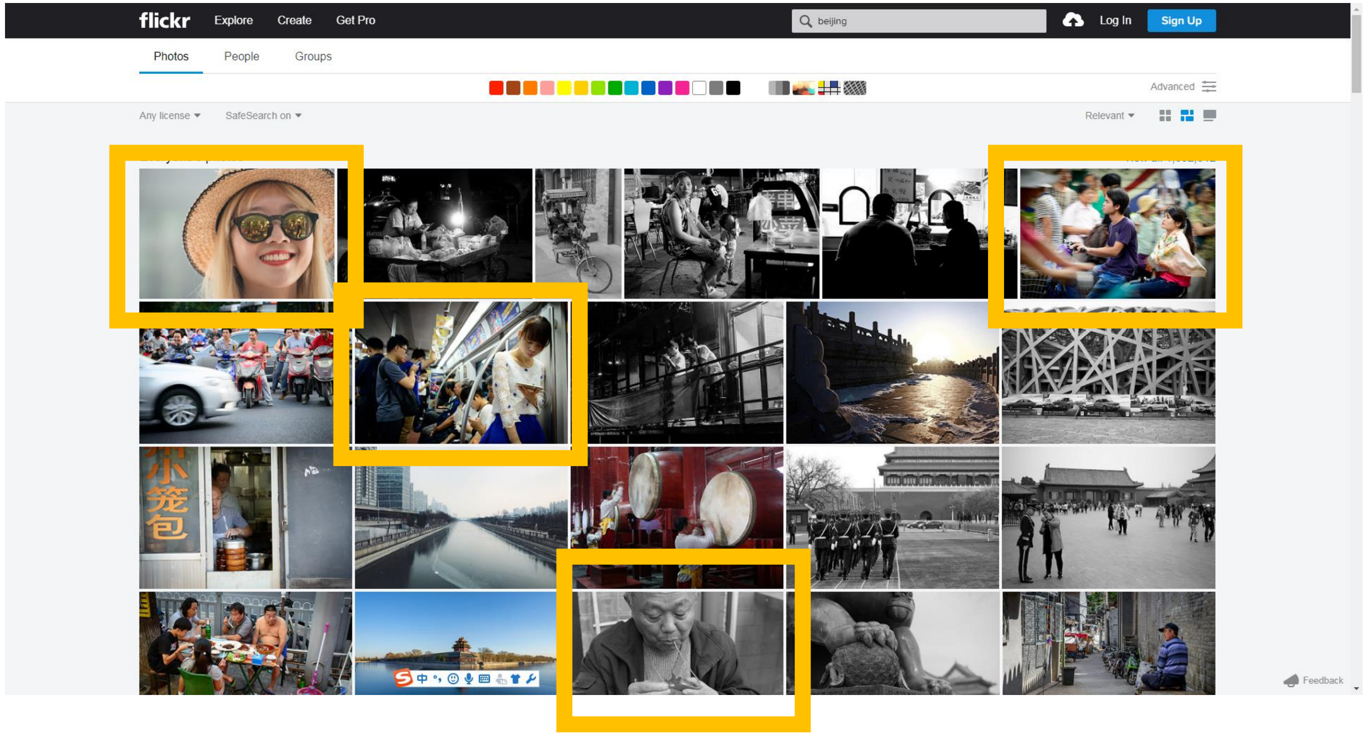
*“**3D reconstruction** is the process of capturing the **geometry** and **appearance** of real objects.”*

--from Wikipedia

Outline

- Why we focus on facial images ?
- Solutions in Industry
- Solutions for single image
  - Geometry
  - Appearance
- Rebirth in Deep Neural Network

Why we focus on facial images ?



*“**Faces** exist in more than 60% of camera phone images.”*

--International Imaging Industry Association

The Curious Case Of

BENJAMIN BUTTON


融合前
融合后


为世界杯打CALL
一键生成球迷彩绘 为你喜欢的球队加油!
21th
WORLD CUP


智识别 人脸识别技术，实时"颜值"诊断
LANCÔME


Animoji
Welcome to Animoji


美容美妆
一键打造精致裸妆
超智能五官精准定位、粉底、唇彩、腮红、
鼻翼高光、修眉、染发


3D FACE MATRIX
鼻部分析
您的鼻子基础良好，微调一下会更完美，使五官更立体，整体比例更好；端正、直而挺、山根丰隆、鼻翼饱满、准头圆润发亮、鼻孔不外露，多半都是很有贵气，为人圆滑，乐于助人，财运好，有夫人命。鼻子色泽明润，没有斑点或痣，一生财运亨通，夫运更好。


新华社 | Sogou搜狗
AI虚拟主播
我的名字叫作"虚拟主播"

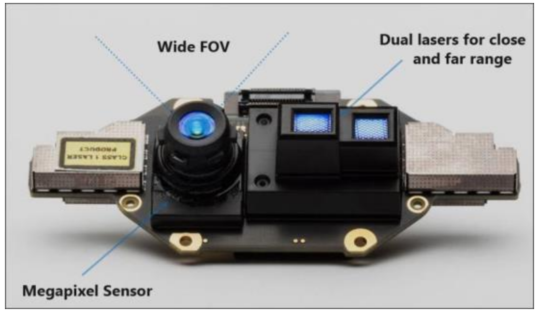
Solutions in Industry

Geometry Reconstruction

- Active methods
  - Structured light (Kinect, iPhone X, *etc.*)
  - Laser range finder
  - Time-of-flight

- Passive methods
  - Multi-view Stereo
  - Shape-from-X (shading, focus/defocus, etc.)
  - Data-driven

Kinect 2.0

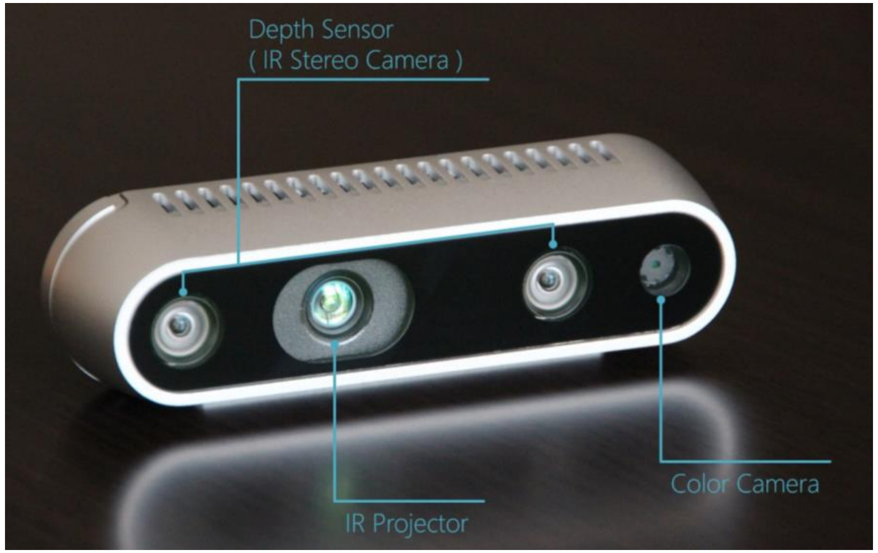


ToF of Microsoft

iPhone X

Geometry Reconstruction

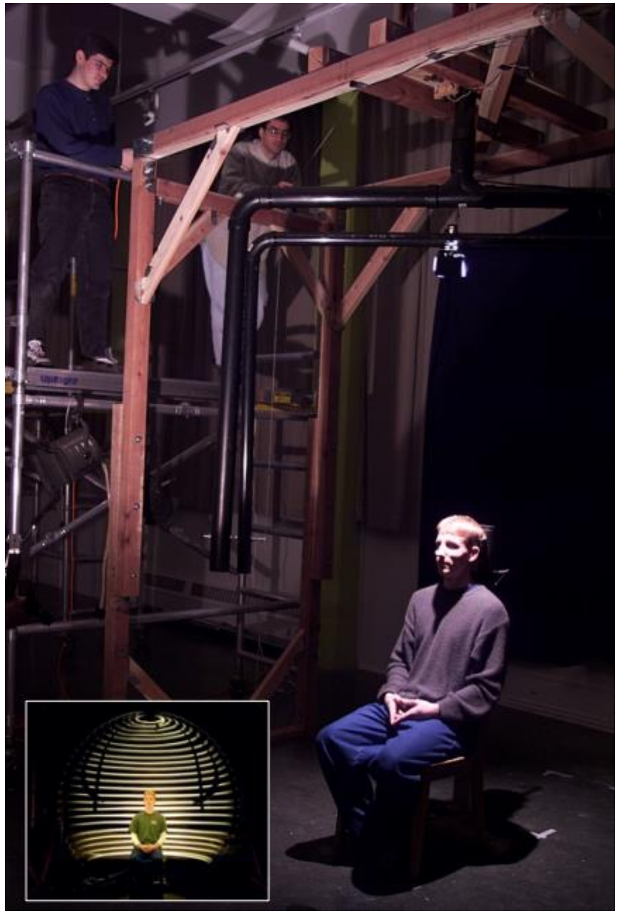
- Active methods
  - Structured light (Kinect, iPhone X, *etc.*)
  - Laser range finder
  - Time-of-flight
- Passive methods
  - Multi-view Stereo
  - Shape-from-X (shading, focus/defocus, silhouettes)
  - Data-driven

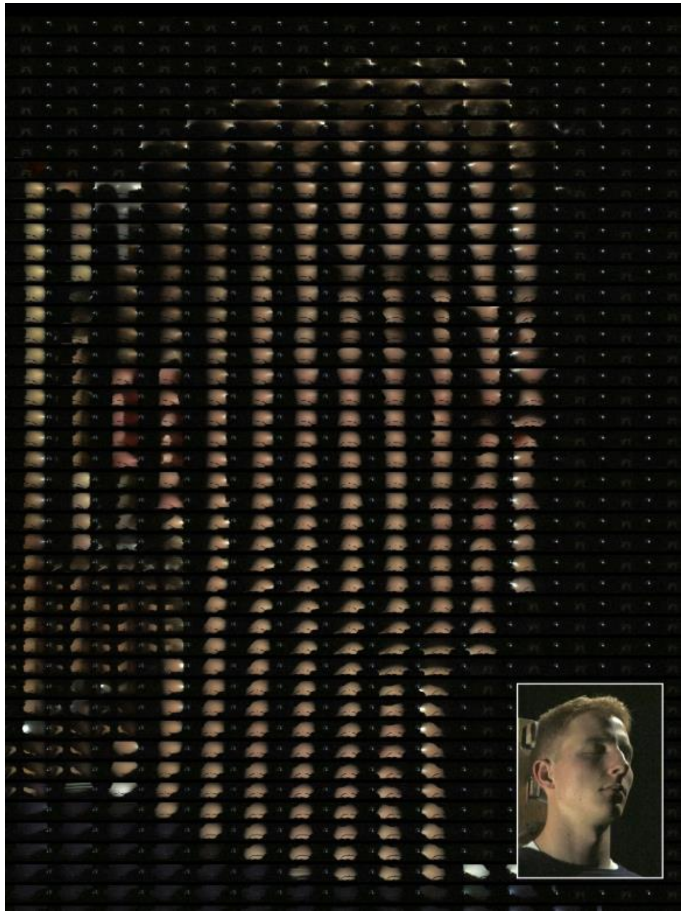


Intel RealSence

USC ICT Light Stage Series

- Light Stage 1
  - Built in 2000
  - One spotlight
  - Takes 1 minute



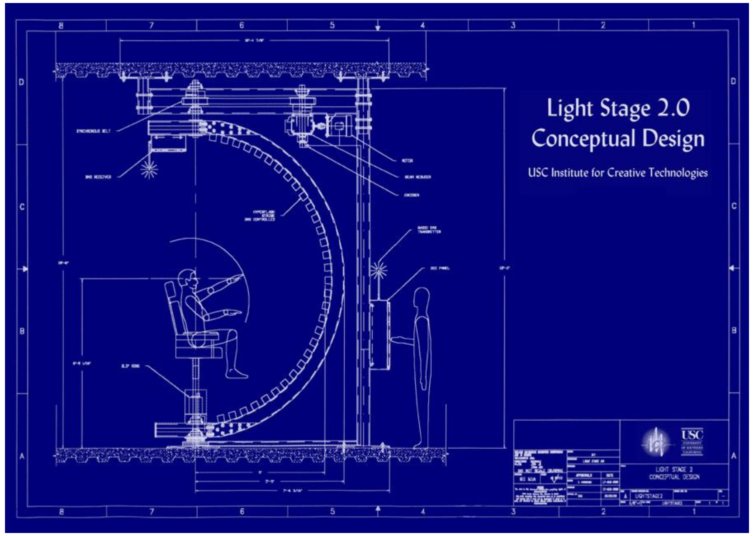


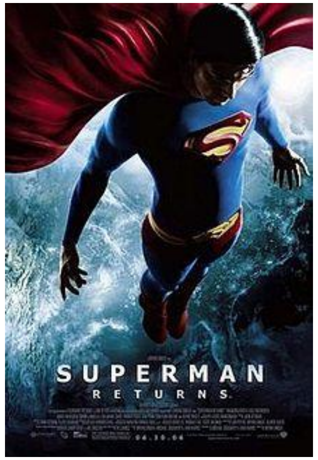
Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. 2000. ***Acquiring the reflectance field of a human face****. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH '00).*

USC ICT Light Stage Series

- Light Stage 1
  - Built in 2000
  - One spotlight
  - Takes 1 minute

- Light Stage 2
  - Built in 2002
  - 30 strobe lights
  - 8 seconds

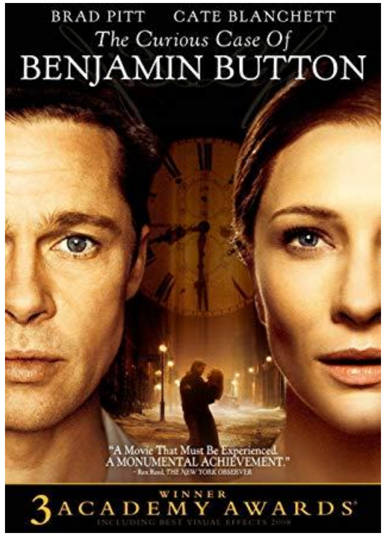




Tim Hawkins, Andreas Wenger, Chris Tchou, Andrew Gardner, Fredrik Göransson, and Paul Debevec. 2004. Animatable facial reflectance fields. In Proceedings of the Fifteenth Eurographics conference on Rendering Techniques (EGSR'04).

USC ICT Light Stage Series

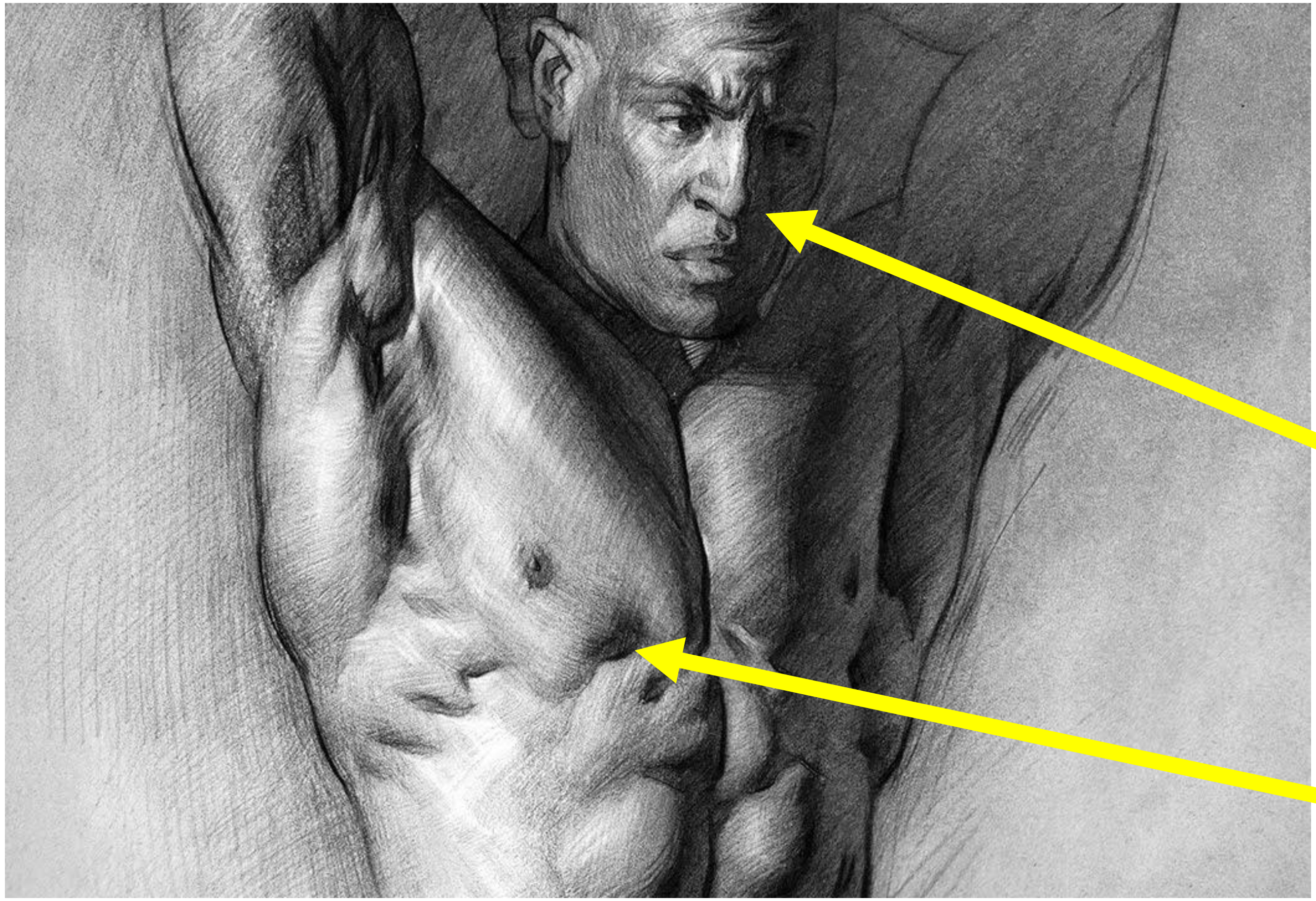
- Light Stage 5
  - Built in 2005
  - 152 lights
  - 24 shot / second



Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. ACM Trans. Graph. 24, 3 (July 2005), 756-764.

Solutions for single image

Analysis by synthesis



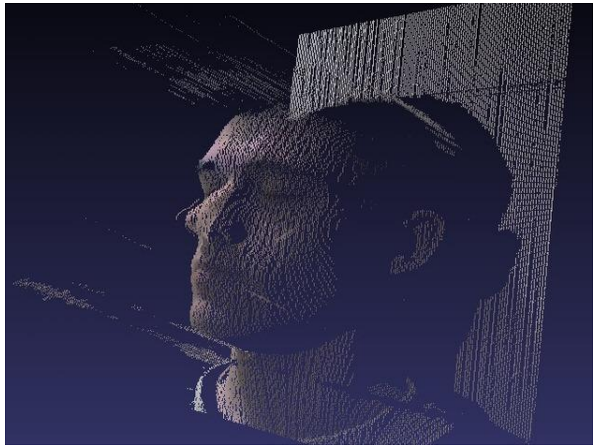
Synthesizing the estimation, *analyzing the accurate.*

Facial landmarks

Shading

Problem formulation

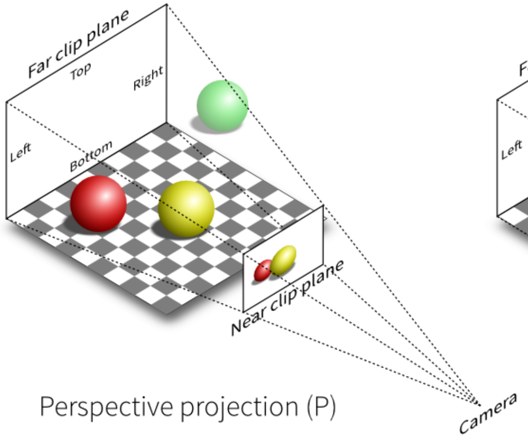
- Geometry representation
  - Point Cloud
  - Depth map/normal map
  - Parametric model



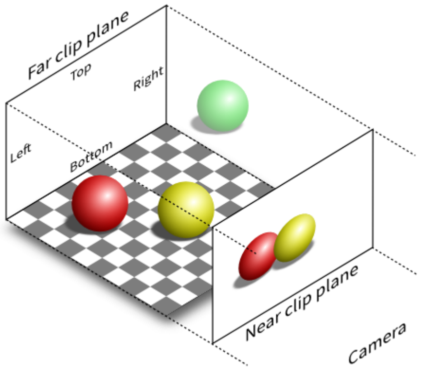
Point Cloud

Normal Map

- 3D-to-2D projection
  - Orthographic projection
  - Weakly perspective projection
  - Perspective projection



Perspective projection (P)



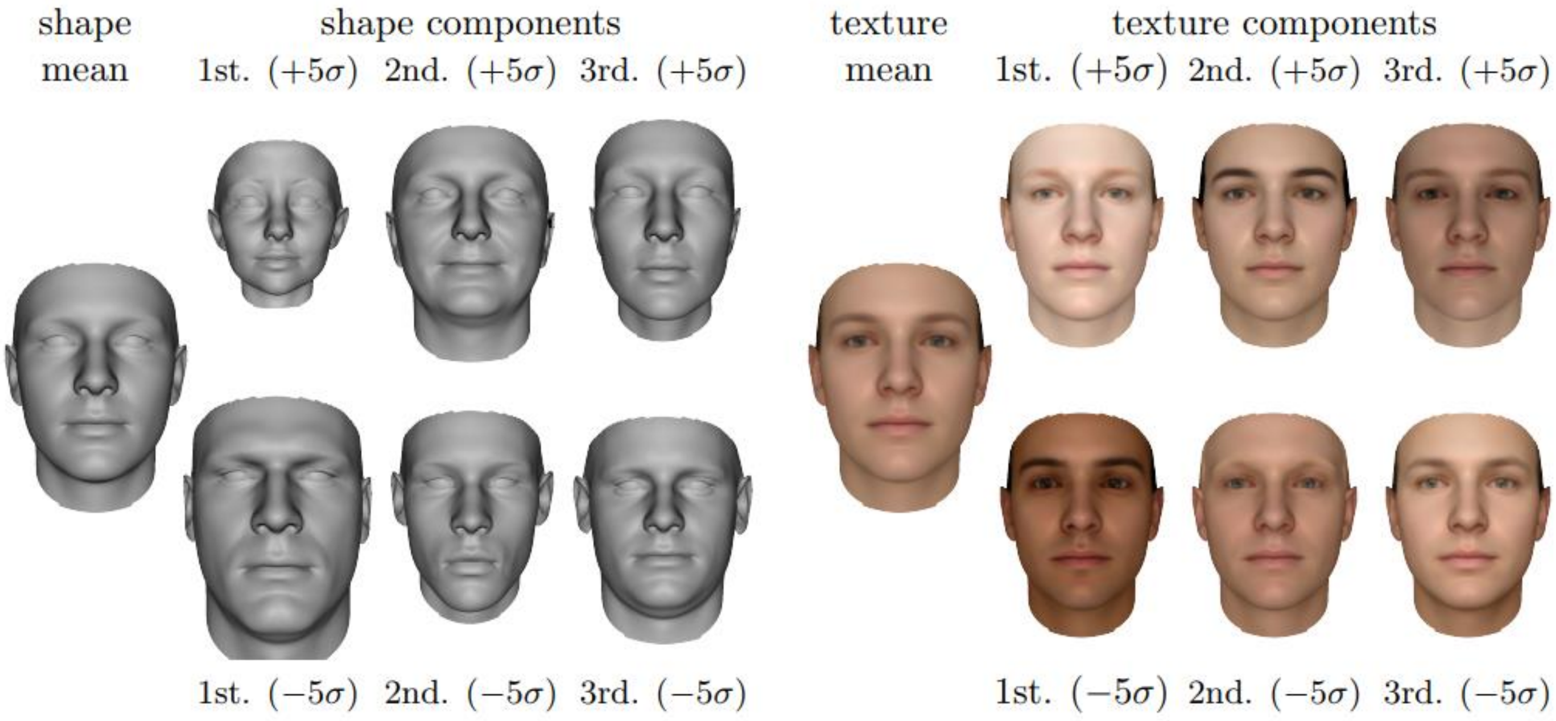
Orthographic projection (O)

3D Morphable Model

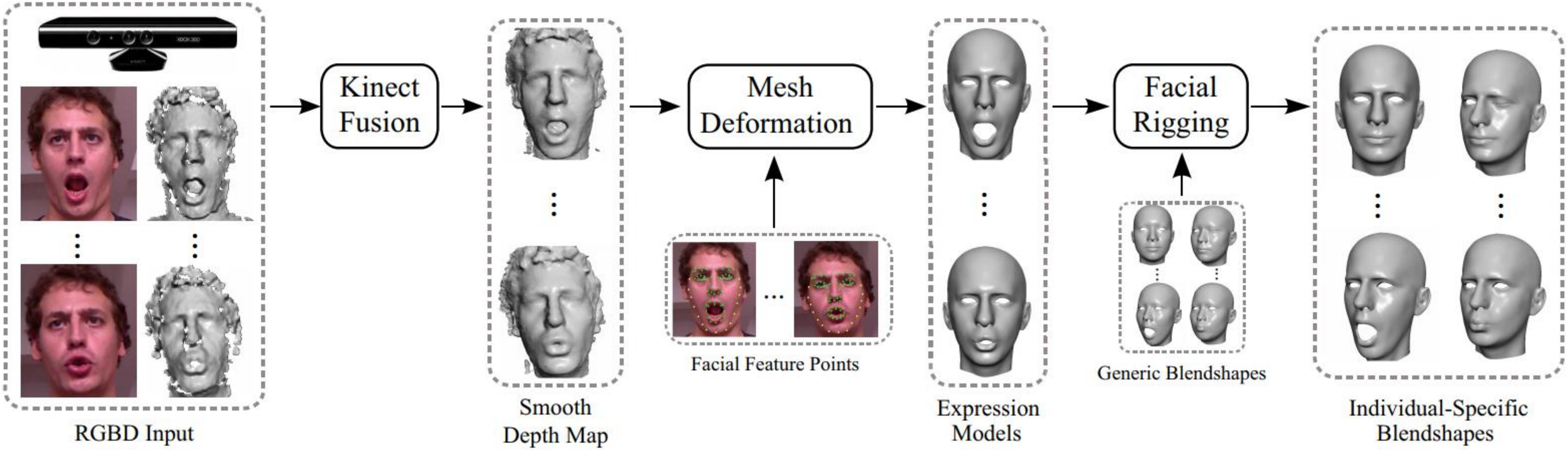
Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter IN:Proceedings of the 6th IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS) for Security, Safety and Monitoring in Smart Environments Genova (Italy) - September 2-4, 2009

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

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



Data Capture



Cao Chen, Yanlin Weng, Shun Zhou, Yiying Tong, Kun Zhou: "FaceWarehouse: a 3D Facial Expression Database for Visual Computing", IEEE Transactions on Visualization and Computer Graphics, 20(3): 413-425, 2014,

Optimization

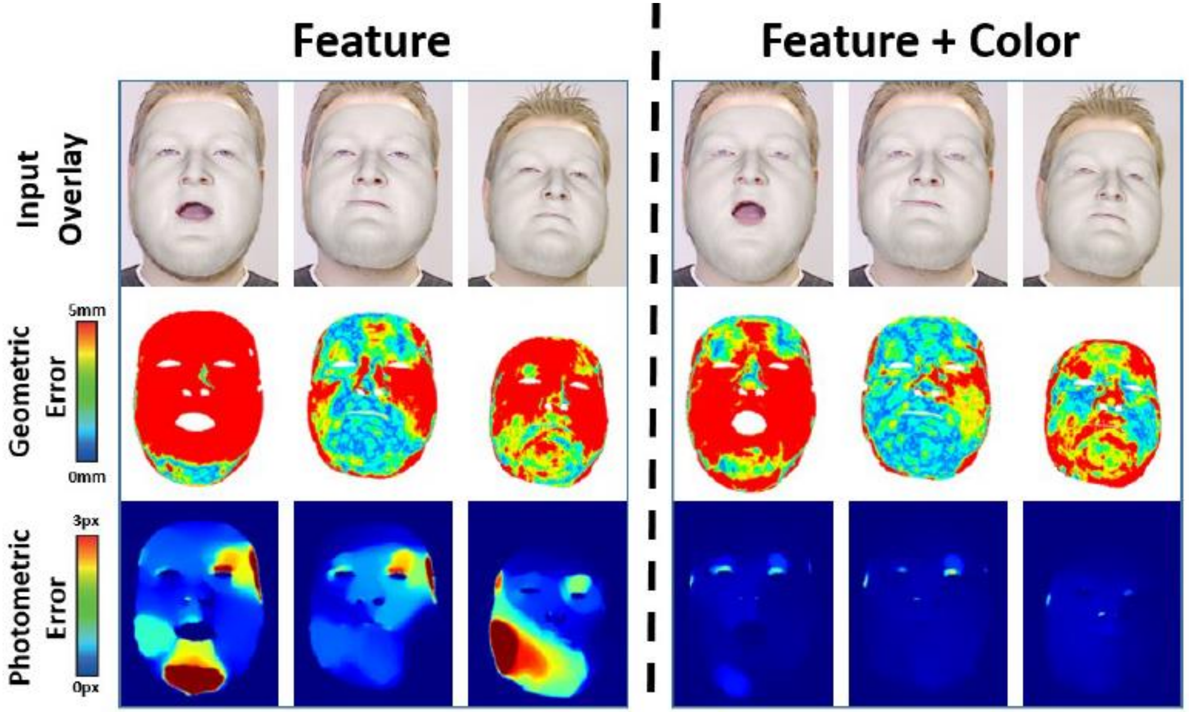
- Target

$$\mathcal{P} = \{\Pi, \alpha_{id}, \alpha_{exp}, \alpha_{tex}\}$$

- Constraints

$$E_{lan} = \sum_{f} ||\Pi\left(G(\alpha_{id}, \alpha_{exp})\right) - f_p||_2^2$$

$$E_{color} = \sum_{p} ||\boldsymbol{T}(\boldsymbol{\alpha}_{tex}) - I(\Pi\left(G(\alpha_{id}, \alpha_{exp})\right))||_2^2$$

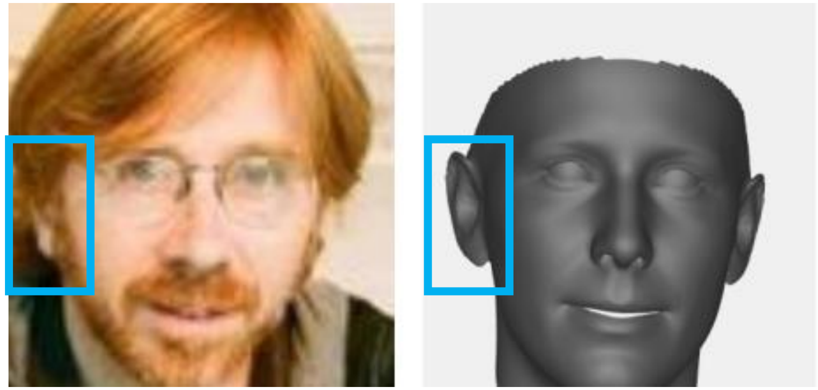


Zollhofer, M. and Thies, J. and Garriod, P. and Bradely, D. and Beeler, T. and Perez, P. and Stamminger, M. and Niesner, M. and Theobalt, C.. 2018. State of the Art on Monocular 3D Face Reconstruction, Tracking, and Applications. Computer Graphics Forum (Eurographics State of the Art Reports 2018).

3D-to-2D Project

Weakly perspective projection

$$\boldsymbol{p} = \boldsymbol{f} * \boldsymbol{R} * \boldsymbol{v} + \boldsymbol{T}_{\boldsymbol{xy}}$$



Tran, et.al. 2018

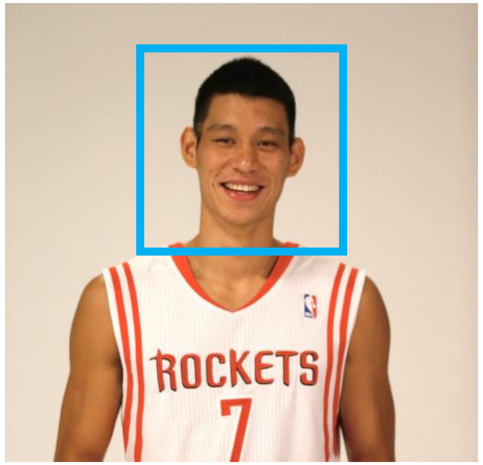
Perspective projection

$$\boldsymbol{v}' = \boldsymbol{R} * \boldsymbol{v} + \boldsymbol{T}$$

$$\boldsymbol{v}'' = \left(\frac{\boldsymbol{v}'_x}{\boldsymbol{v}'_z}, \frac{\boldsymbol{v}'_y}{\boldsymbol{v}'_z}, \mathbf{1}\right)$$

$$\boldsymbol{p} = \begin{bmatrix} \boldsymbol{f}_x & \mathbf{0} & \boldsymbol{c}_x \\ \mathbf{0} & \boldsymbol{f}_y & \boldsymbol{c}_y \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \boldsymbol{v}''$$

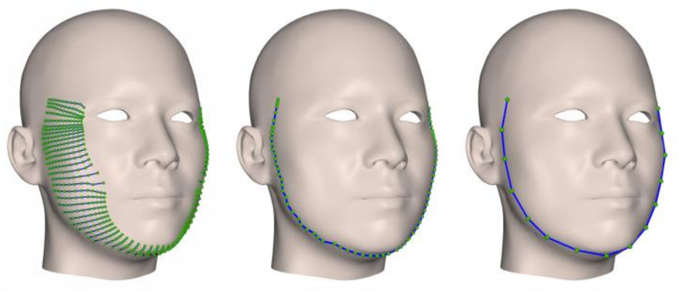
Tewari, et.al. 2018



Face Contour Handling

Dense contour hypothesis

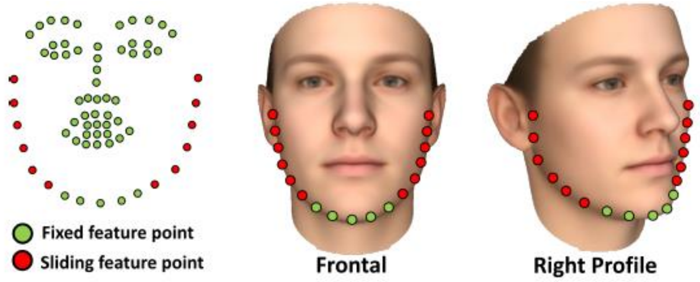
- Pre-defined contour lines
- The computation of visibility is costly



Cao, et.al. 2014

Sliding feature points

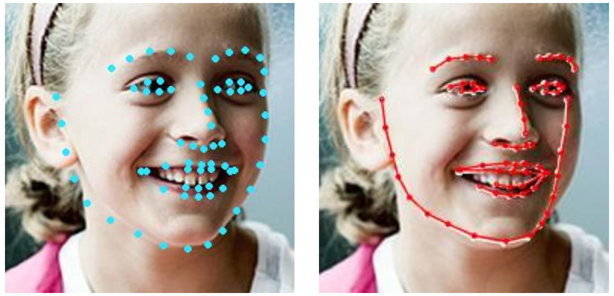
- Iterative updated sliding feature points



Tewari, et.al. 2018

3D landmarks

- Training data is limited
- Hard to capture complex expression



Bulat, et.al. 2017

How about the facial details?

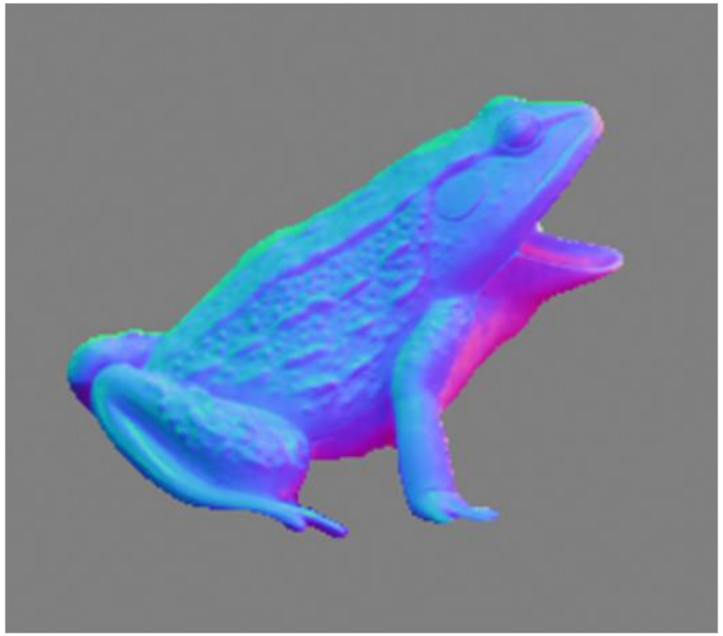
Shape-from-shading

- An intuition

Calibrated illumination

Shading image

Estimated normal map



M. K. Johnson and E. H. Adelson, "Shape estimation in natural illumination," CVPR 2011, Colorado Springs, CO, USA, 2011, pp. 2553-2560.

Formulation

- Distant light

$$\mathrm{I} = A \cdot \mathrm{N} \cdot \mathrm{L}$$

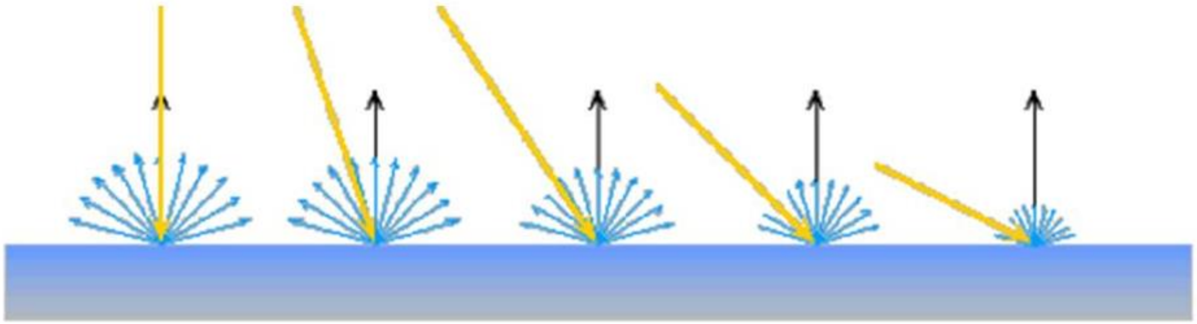
- Parametric illumination

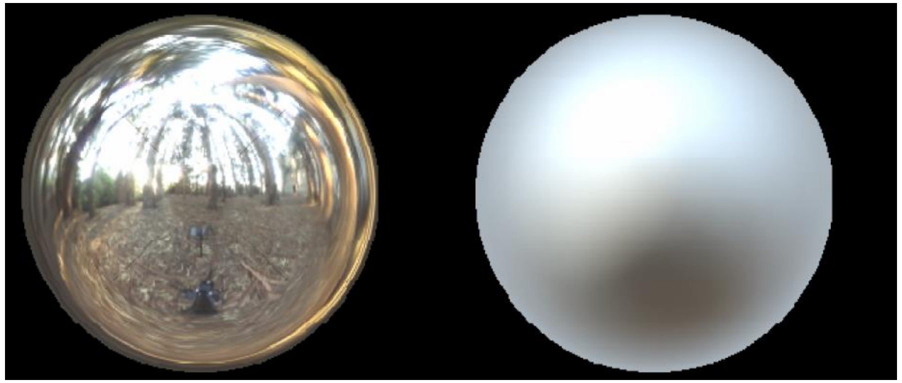
$$\mathrm{L}(\theta, \phi) = \sum_{l,m} L_{lm} Y_{lm}(\theta, \phi)$$

$$\mathrm{I} = A \cdot \mathrm{N}^{\mathrm{T}} \cdot \mathrm{M} \cdot N$$

Lambert cosine law

The radiant intensity observed from an ideal diffuse diffusely reflecting surface is directly proportional to the cosine of the angle between the direction of the incident light and the surface normal.

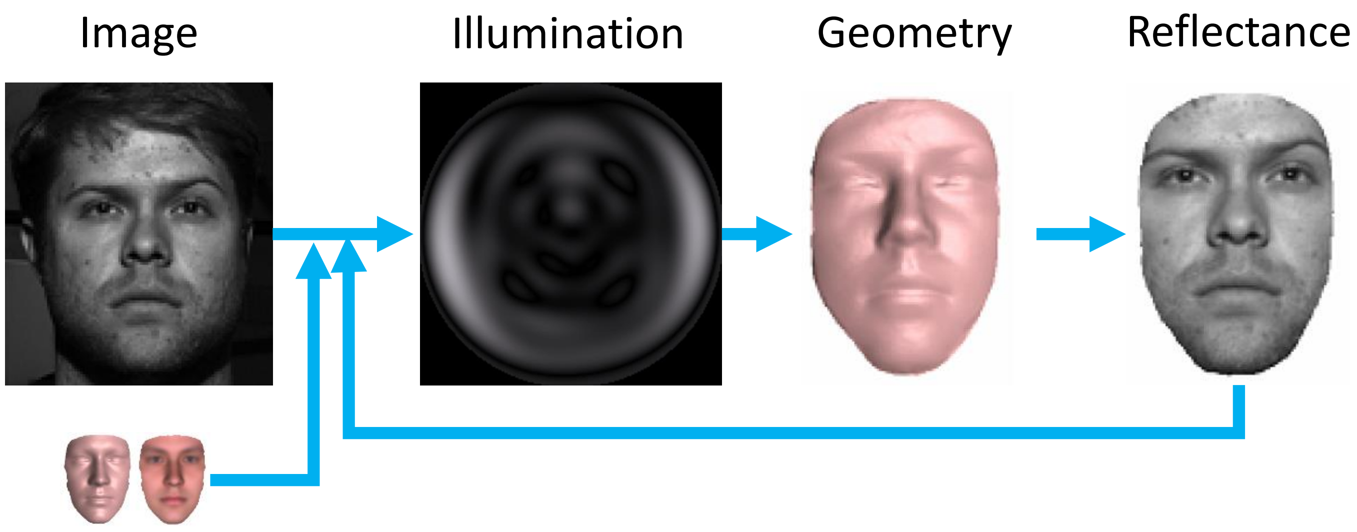


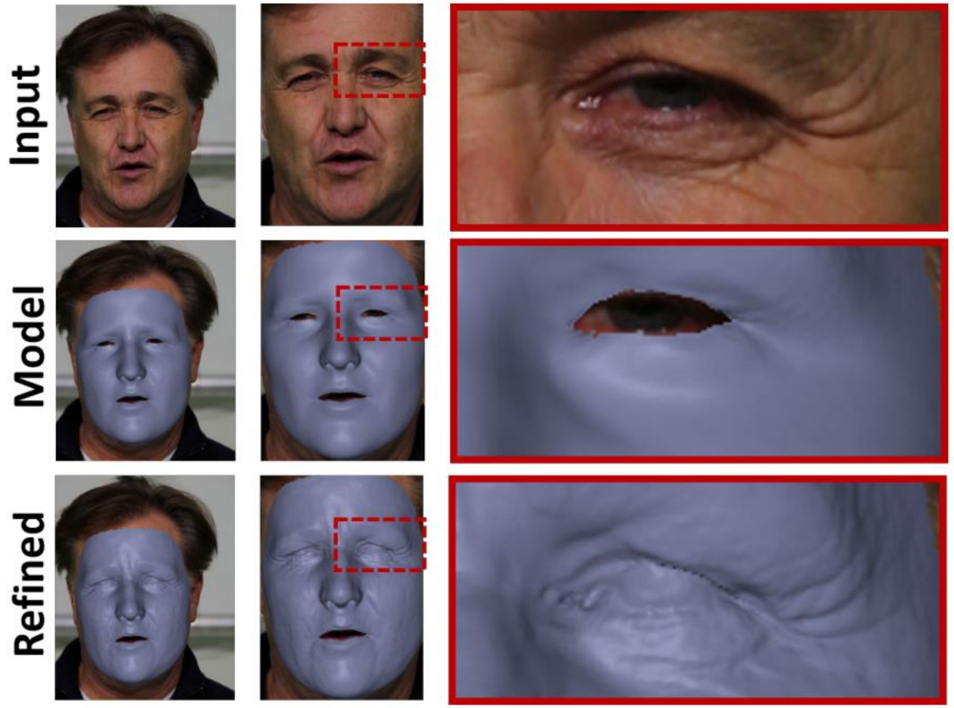


Ravi Ramamoorthi and Pat Hanrahan. 2001. An efficient representation for irradiance environment maps. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH '01).

A solution for face

$$\mathrm{I} = A \cdot \mathrm{N}^{\mathrm{T}} \cdot \mathrm{M} \cdot N$$



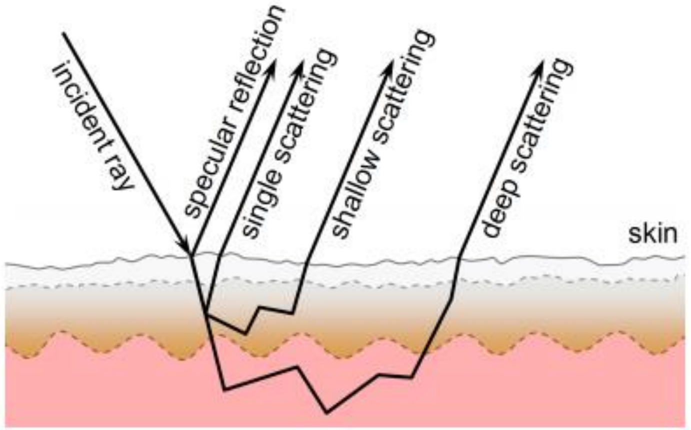


Garrido, et.al. 2016

I. Kemelmacher-Shlizerman and R. Basri, "3D Face Reconstruction from a Single Image Using a Single Reference Face Shape," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 2, pp. 394-405, Feb. 2011.

Is face really Lambertian?

Face Appearance Modeling

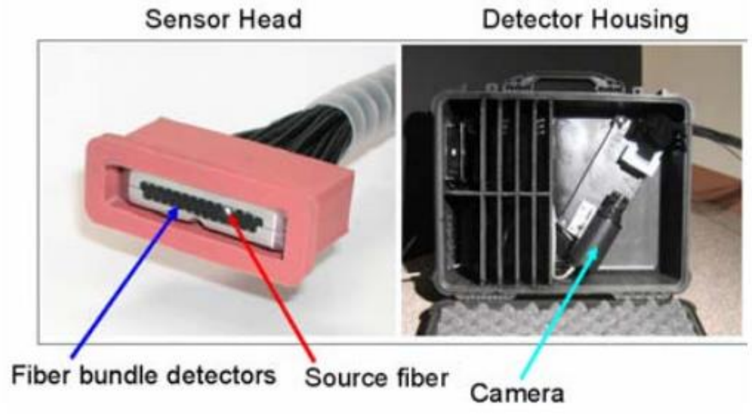


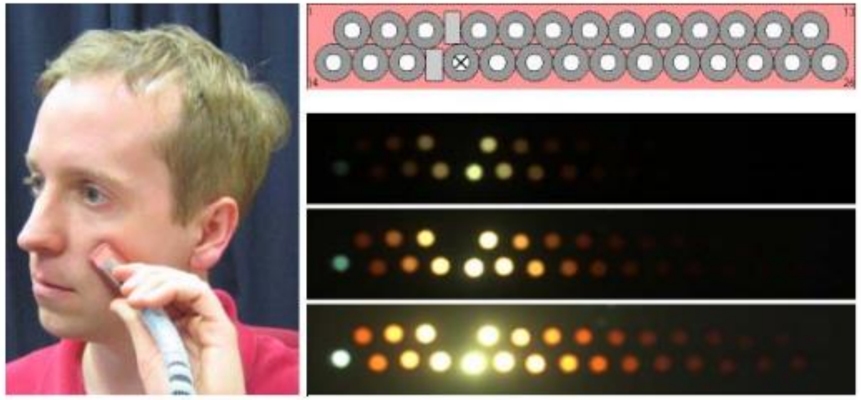
Specular Reflection | Single Scattering | Shallow Scattering | Deep Scattering | Rendering | Photograph

Abhijeet Ghosh, Tim Hawkins, Pieter Peers, Sune Frederiksen, and Paul Debevec. 2008. Practical modeling and acquisition of layered facial reflectance. In ACM SIGGRAPH Asia 2008 papers (SIGGRAPH Asia '08).

Face Appearance Capture

https://docs.sharktacos.com/photography/xpol.html





Weyrich, et.al. 2006

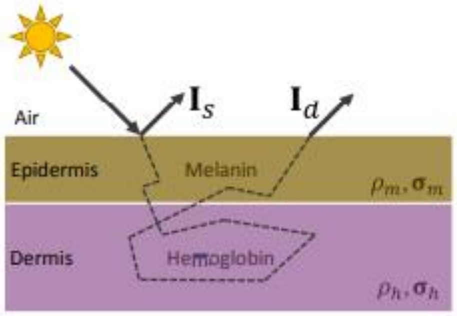
Specular separation

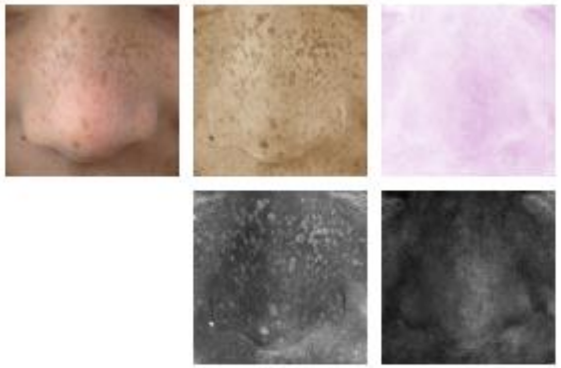
$$\mathrm{I} = A \cdot \mathrm{N}^{\mathrm{T}} \cdot \mathrm{M} \cdot N$$

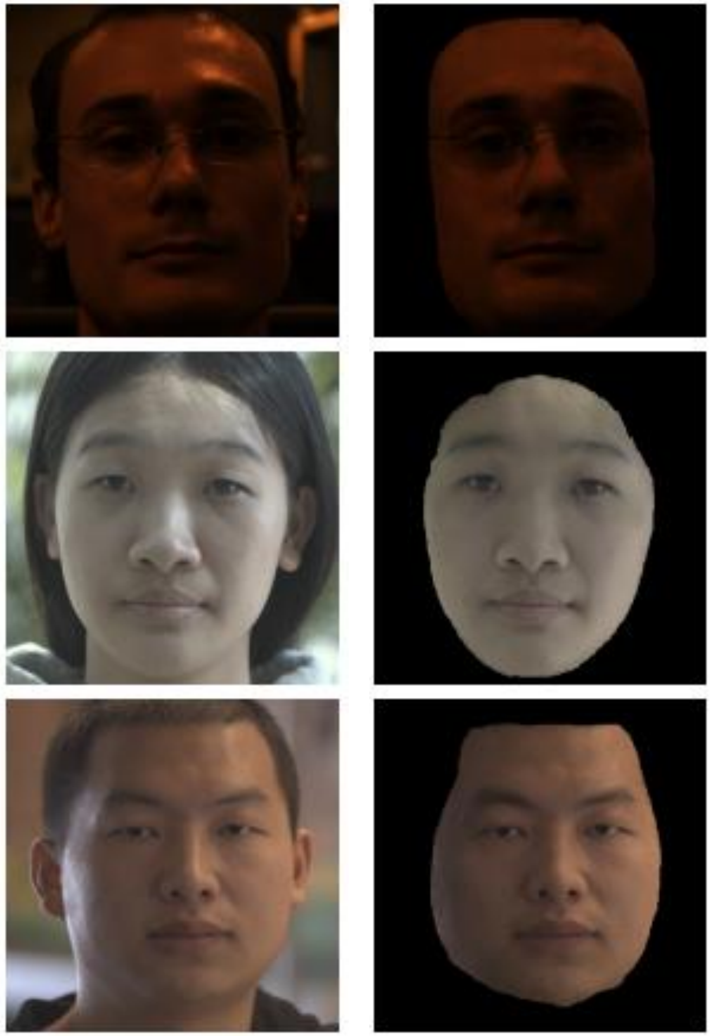
$$\mathrm{I} = A \cdot \mathrm{N}^{\mathrm{T}} \cdot \mathrm{M} \cdot N + S$$

$$A = \sigma_m^{\rho_m} \sigma_h^{\rho_h}$$

Low-frequency illumination ⇠ ⇢ High-frequency illumination







C. Li, S. Lin, K. Zhou and K. Ikeuchi, "Specular Highlight Removal in Facial Images," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 2780-2789.

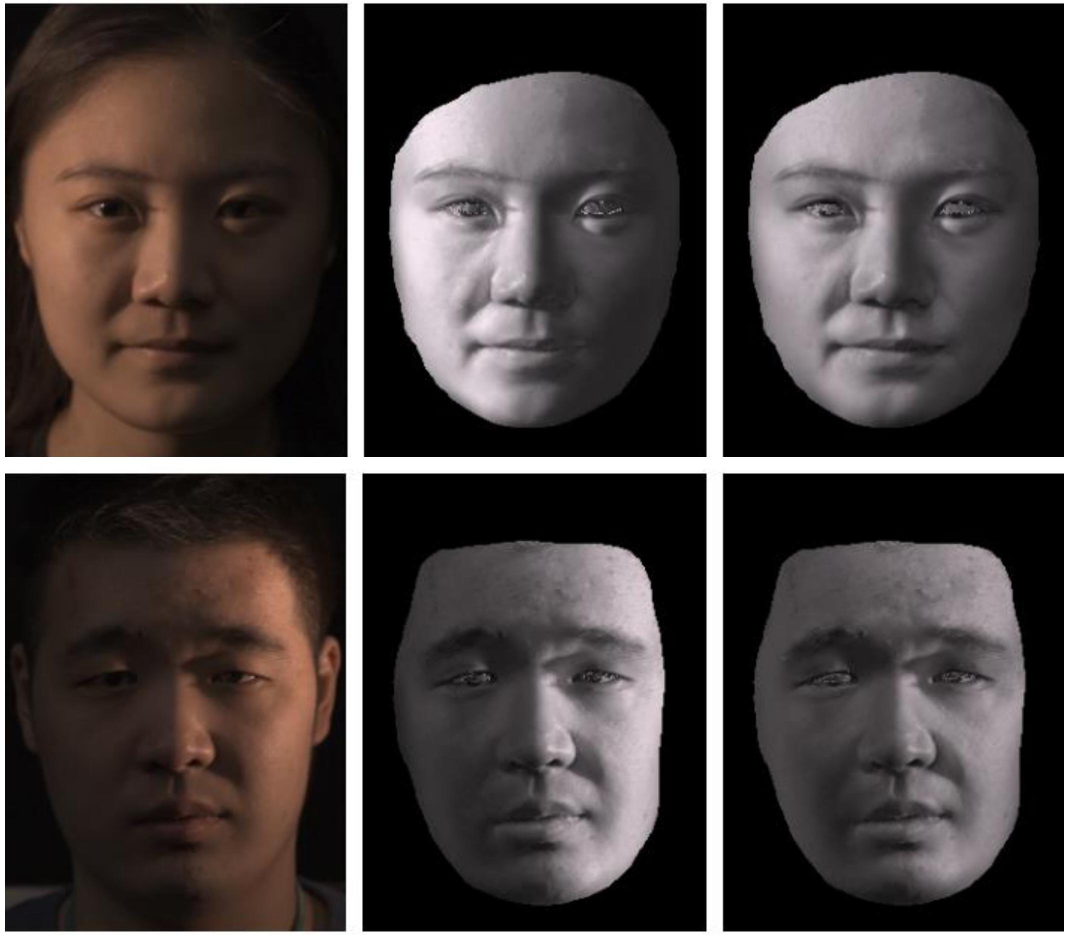
Subsurface scattering

$$I = A \cdot N^{T} \cdot M \cdot N + S$$

$$I = BSSRDF + S$$

Lambertian BSSRDF

Lambertian BSSRDF



Li C., Zhou K., Lin S. (2014) Intrinsic Face Image Decomposition with Human Face Priors. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham

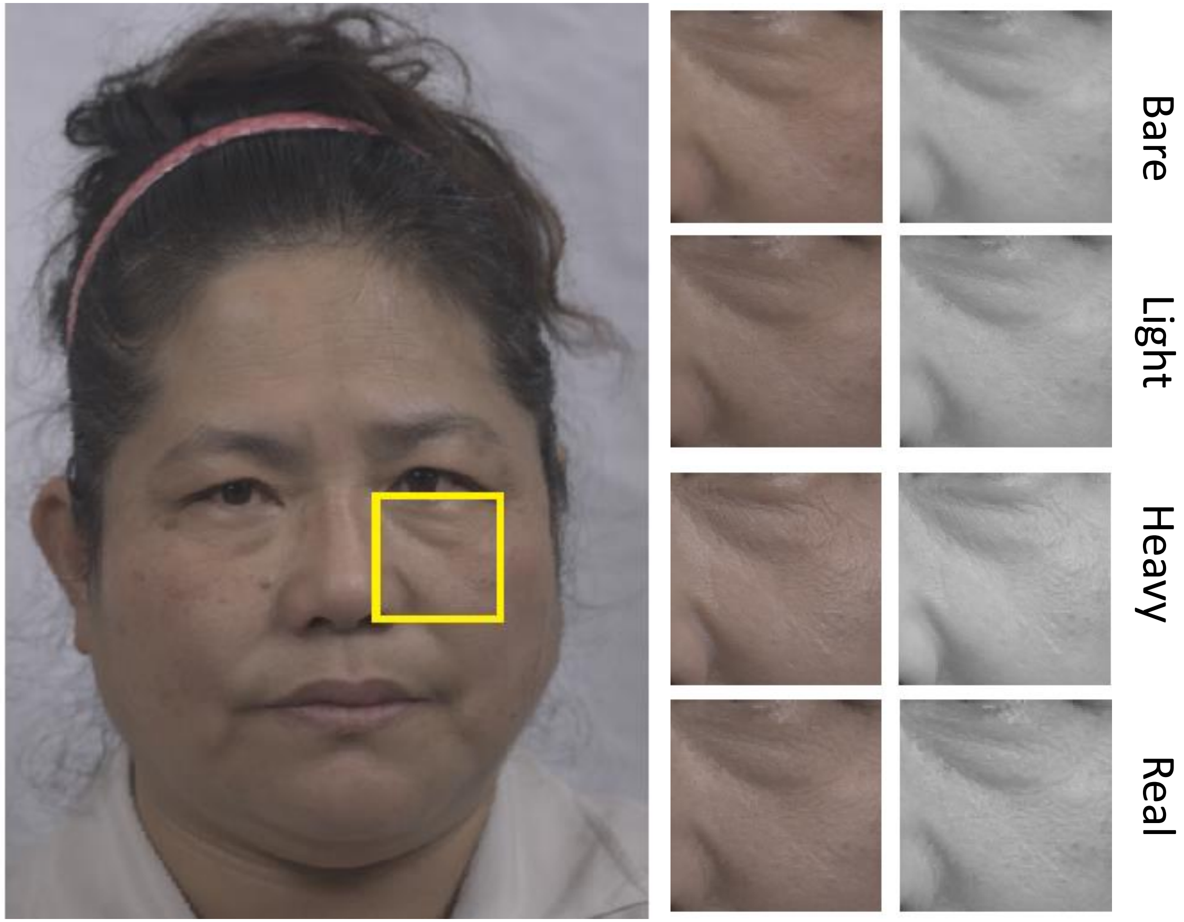
Subsurface scattering

$$\mathrm{I} = A \cdot \mathrm{N}^{\mathrm{T}} \cdot \mathrm{M} \cdot N + S$$

$$\Downarrow$$

$$\mathrm{I} = BSSRDF + S$$

Lambertian BSSRDF



C. Li, K. Zhou and S. Lin, "Simulating makeup through physics-based manipulation of intrinsic image layers," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 4621-4629.

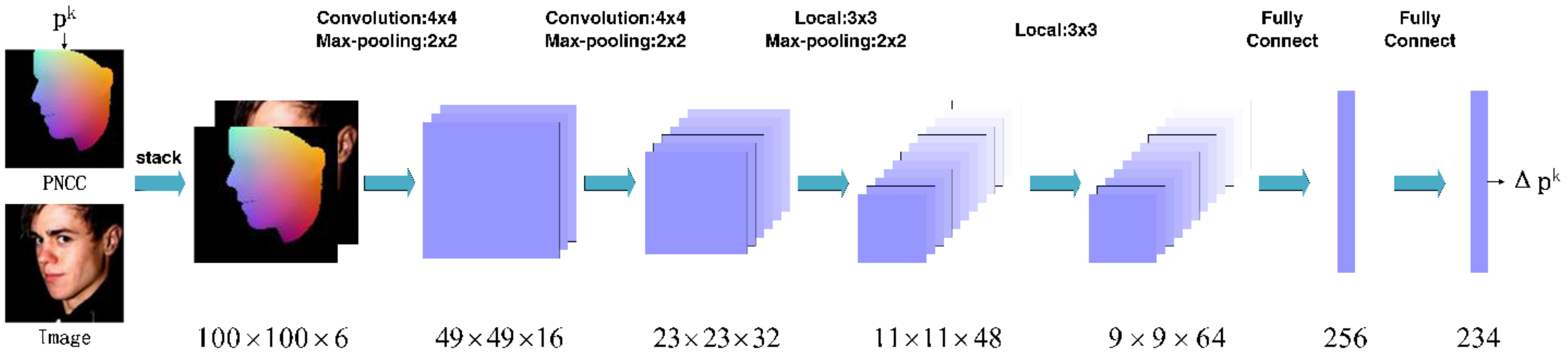
Rebirth in Deep Neural Network

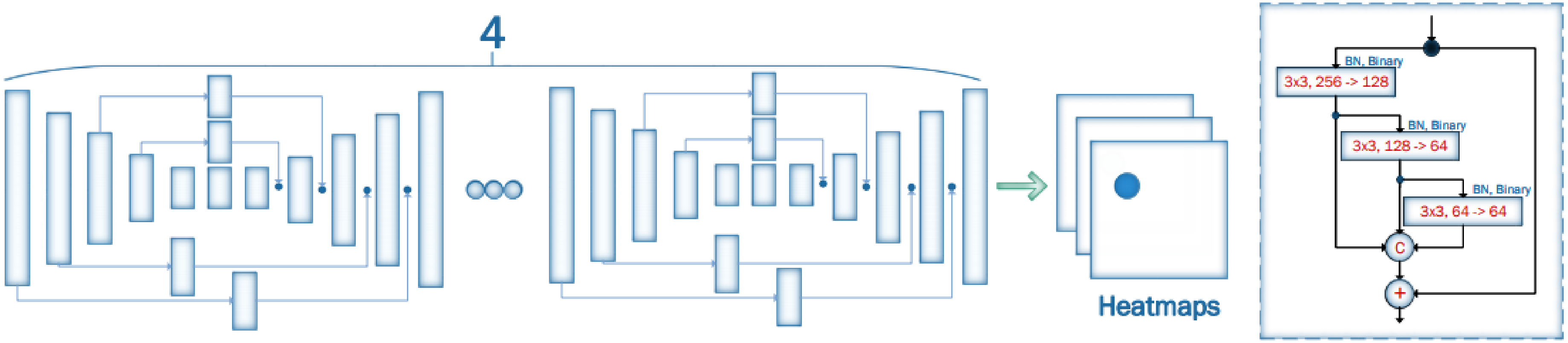
Rebirth in Deep Neural Network

- Frontal face to arbitrary poses
- Occlusion handling
- Separation of training and testing
- Synthetic data + unsupervised real images

Frontal face to arbitrary poses

X. Zhu, Z. Lei, X. Liu, H. Shi and S. Z. Li, "Face Alignment Across Large Poses: A 3D Solution," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 146-155.

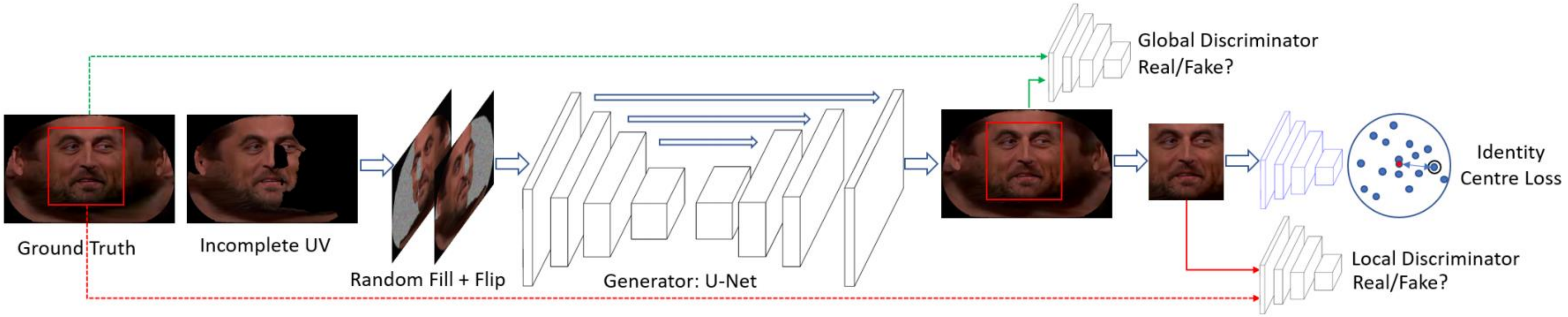


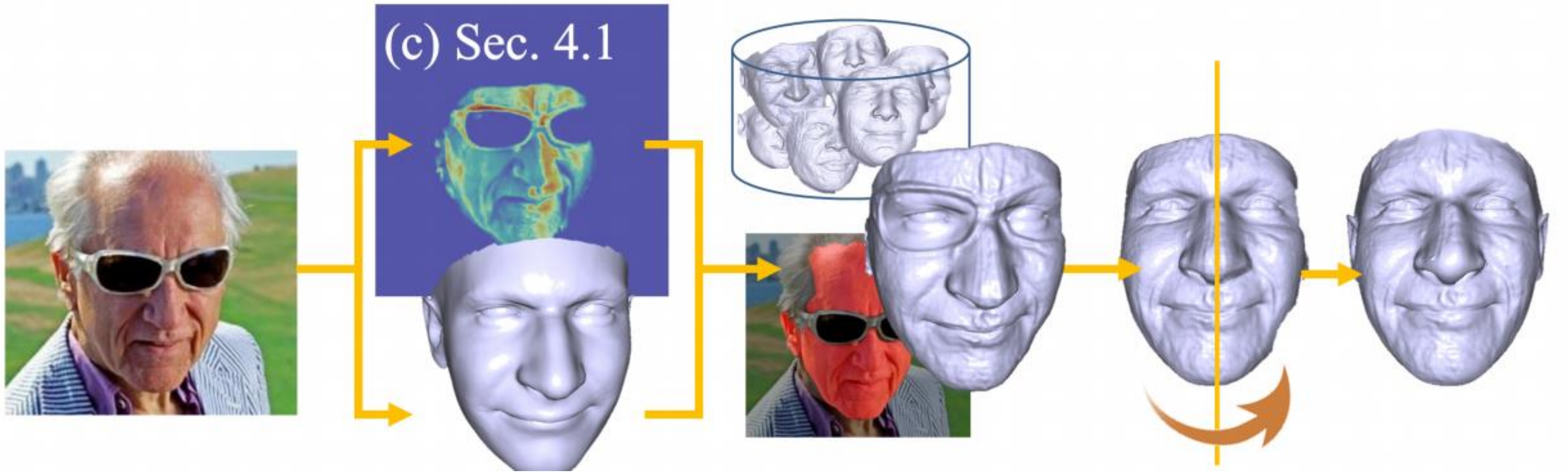


A. Bulat and G. Tzimiropoulos, "How Far are We from Solving the 2D & 3D Face Alignment Problem? (and a Dataset of 230,000 3D Facial Landmarks)," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 1021-1030.

Occlusion handling

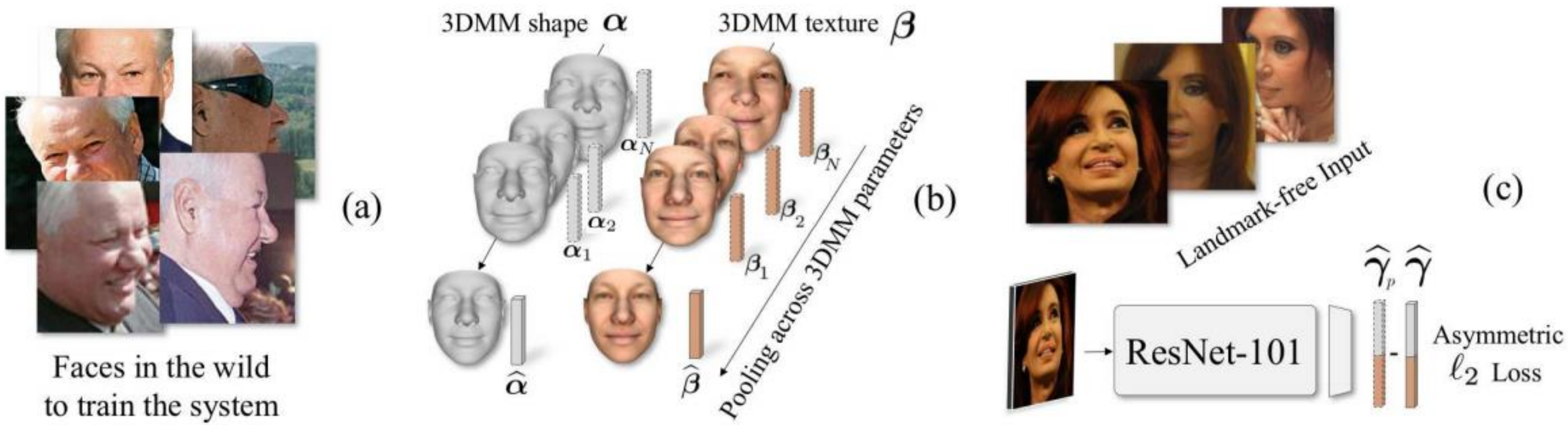
Deng, Jiankang and Cheng, Shiyang and Xue, Niannan and Zhou, Yuxiang and Zafeiriou, Stefanos, "UV-GAN: Adversarial Facial UV Map Completion for Pose-Invariant Face Recognition," 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).





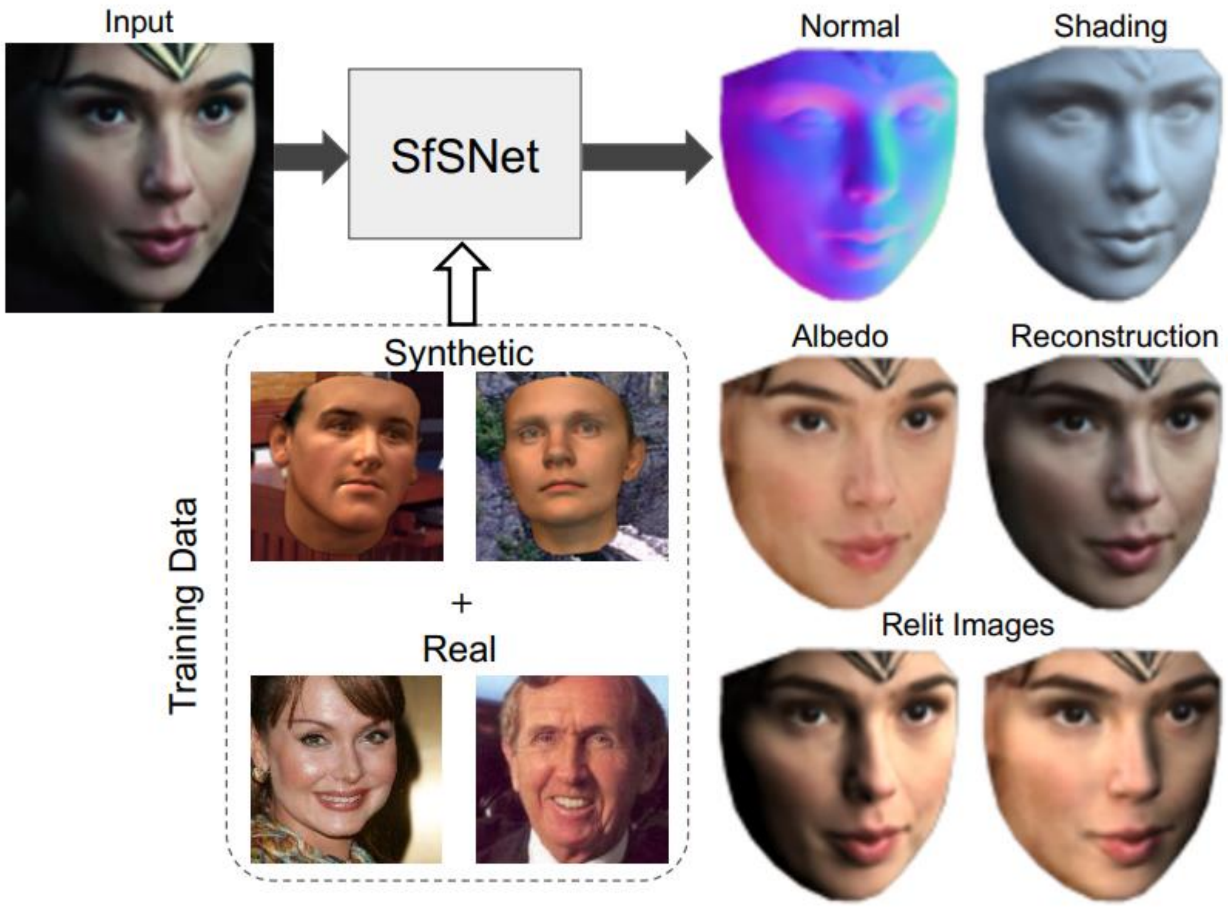
Anh Tuấn Trần, Tal Hassner, Iacopo Masi, Eran Paz, Yuval Nirkin, Gérard Medioni, "Extreme 3D Face Reconstruction: Seeing Through Occlusions," 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Separation of training and testing



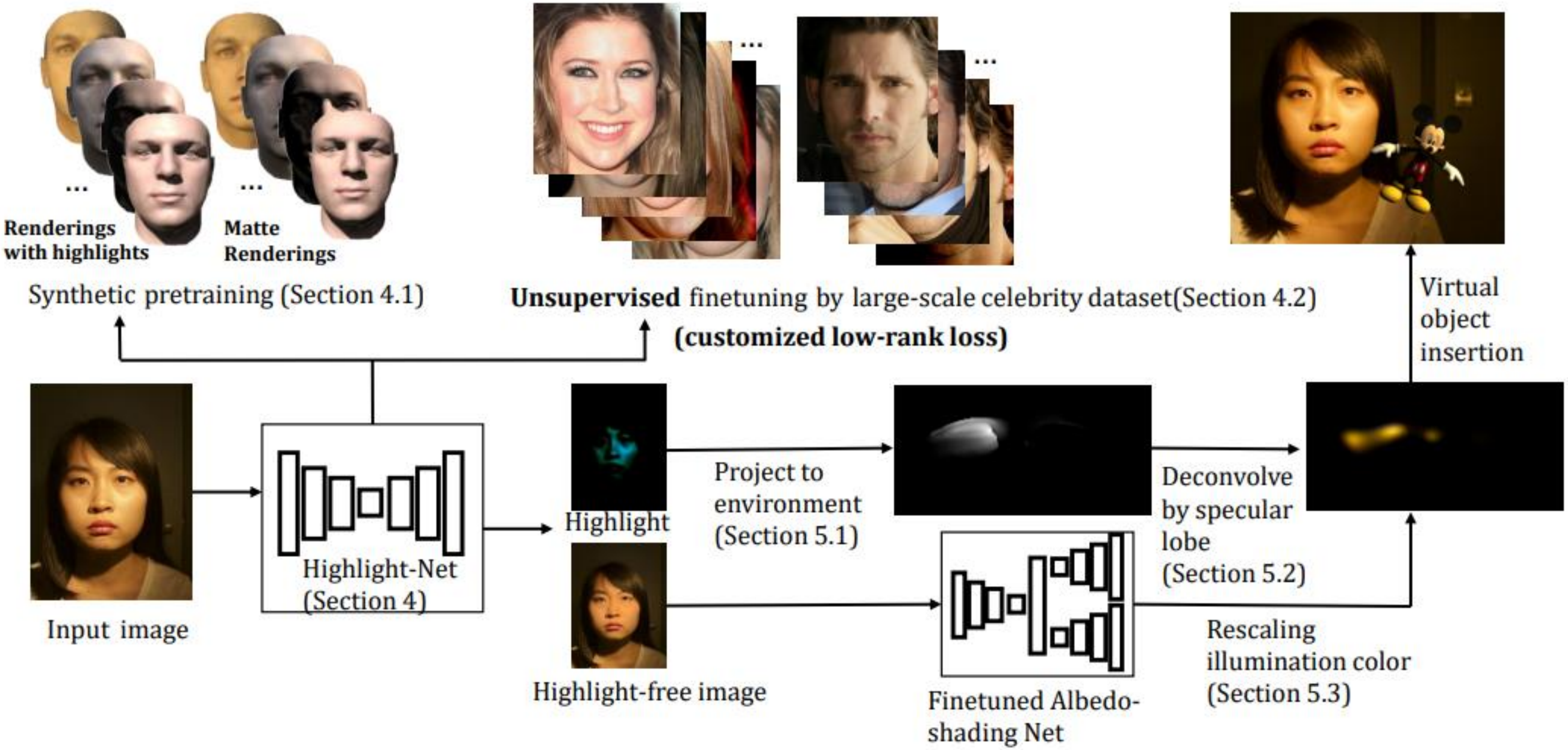
A. T. Tran, T. Hassner, I. Masi and G. Medioni, "Regressing Robust and Discriminative 3D Morphable Models with a Very Deep Neural Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 1493-1502.

Synthetic data + unsupervised real images



Soumyadip Sengupta and Angjoo Kanazawa and Carlos D. Castillo and David W. Jacobs, "SfSNet: Learning Shape, Refectance and Illuminance of Faces in the Wild," 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Synthetic data + unsupervised real images



Yi, Renjiao and Zhu, Chenyang and Tan, Ping and Lin, Stephen, "Faces as Lighting Probes via Unsupervised Deep Highlight Extraction," 2018 The European Conference on Computer Vision (ECCV).

Good Reference

- ZJU: Kun Zhou (http://kunzhou.net/)
- USC: Hao Li (https://www.hao-li.com)
- MSU: Xiaoming Liu (http://cvlab.cse.msu.edu/)
- MPI: GVV (http://gvv.mpi-inf.mpg.de/)