3D Face Reconstruction from a Single Image Using a Single Reference Face Shape

Ira Kemelmacher-Shlizerman, *Member*, *IEEE*, and Ronen Basri, *Senior Member*, *IEEE*

Team member: Ruiqi Liu, Yintao Xu, Sihang Xu, Jiajian Zhang

Naïve model

Assumptions:

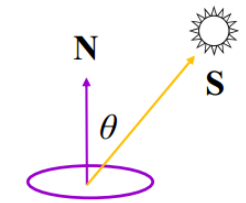
1.a Lambertian object
2.a set of images from the same view of the same object with known light directions.

Known: source vectors $\mathbf{S}_j$ and pixel values $I_j(x,y)$

Unknown: surface normal $\mathbf{N}(x,y)$ and albedo $\rho(x,y)$

$$B = \rho(\mathbf{N} \cdot \mathbf{S})$$

Diffuse reflection: Lambert's law



$$B = \rho(\mathbf{N} \cdot \mathbf{S})$$
$$= \rho \|\mathbf{S}\| \cos\theta$$

B: radiosity (total power leaving the surface per unit area)
ρ: albedo (反射率fraction of incident irradiance reflected by the surface)
N: unit normal
S: source vector (magnitude proportional to intensity of the source)

(from lecutre04 light)

Naïve model

Known: source vectors $\mathbf{S}_j$ and pixel values $I_j(x,y)$

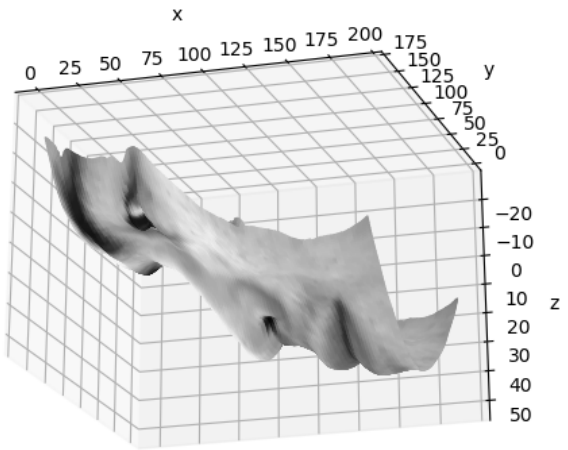
Unknown: surface normal $\mathbf{N}(x,y)$ and albedo $\rho(x,y)$

$$I_j(x,y) = k\rho(x,y)\left(\mathbf{N}(x,y)\cdot\mathbf{S}_j\right)$$
$$= \left(\rho(x,y)\mathbf{N}(x,y)\right)\cdot(k\mathbf{S}_j)$$
$$= \mathbf{g}(x,y)\cdot\mathbf{V}_j$$

For each pixel, set up a linear system:

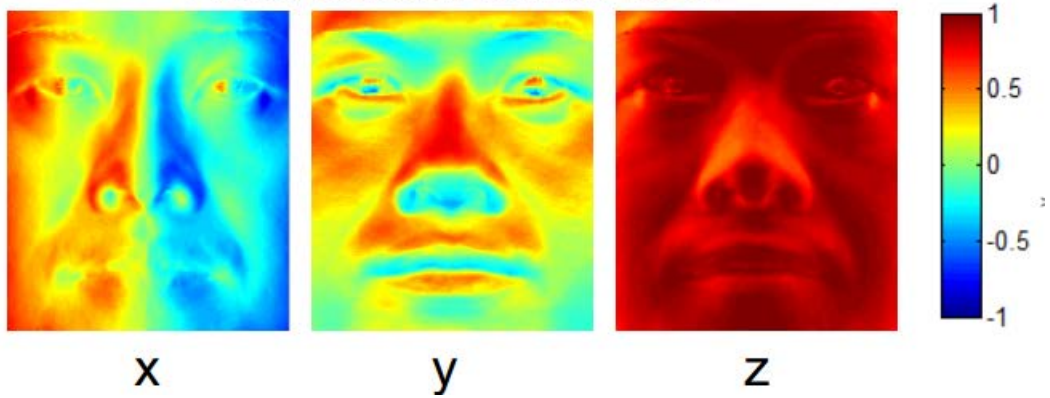
$$\begin{bmatrix} I_1(x,y) \\ I_2(x,y) \\ \vdots \\ I_n(x,y) \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \\ \vdots \\ \mathbf{V}_n^T \end{bmatrix} \mathbf{g}(x,y)$$

$(n \times 1)$ known — $(n \times 3)$ known — (3×1) unknown



"integrability"
有限差分近似微分

Recovered normal field

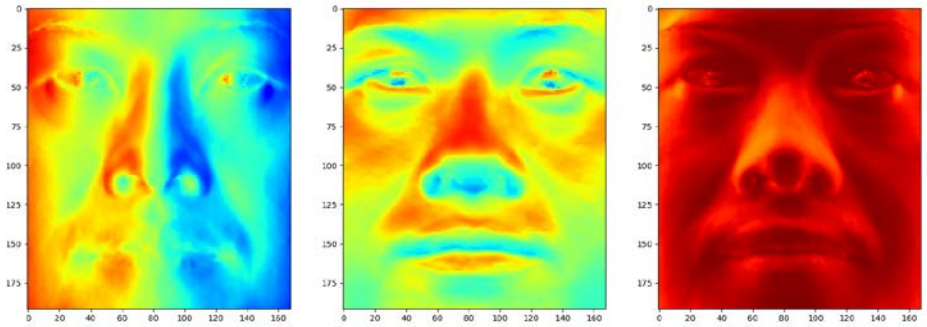


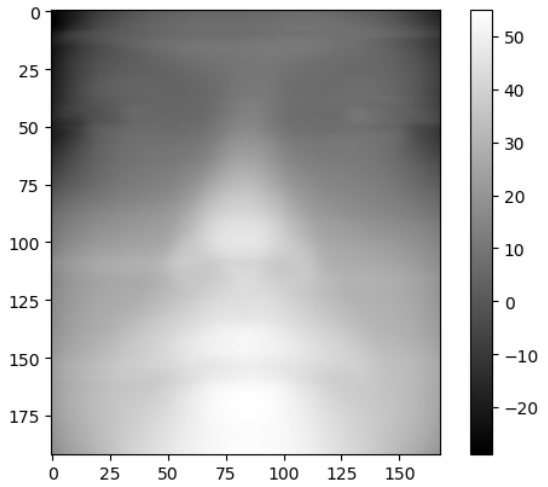
x y z

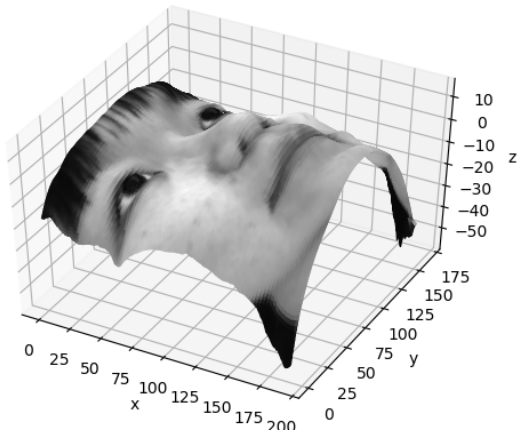
partly quoted from lecutre04 light

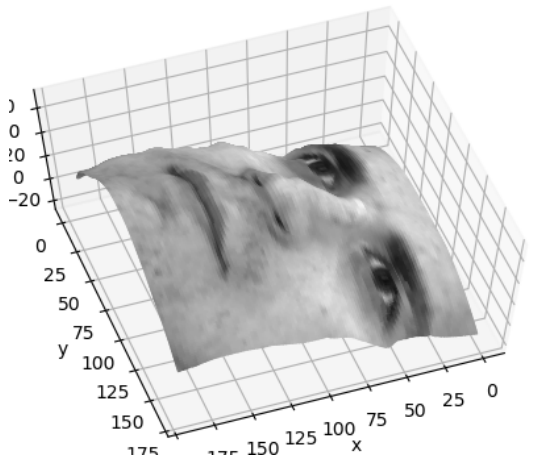
Research base on YaleB chopped database

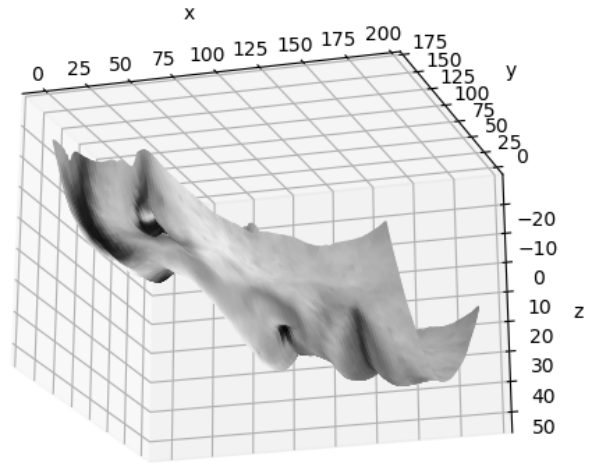
Experiments result









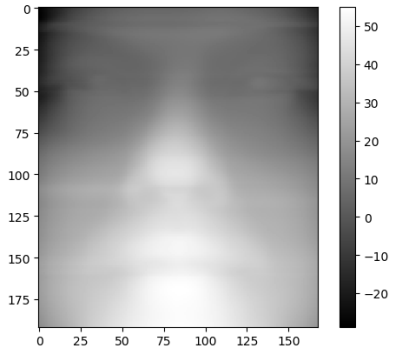


Research base on YaleB chopped database

Advanced model

Assumptions:

1.A single picture of a nearly frontal face
2.a reference 3-D face model with norms, albedo, depth.
3.No cast shadow, interreflection



√

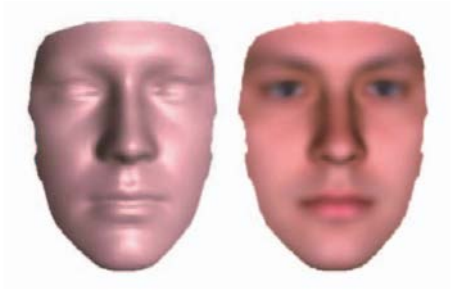


Fig. 2. The generic face model obtained by taking the mean shape over the entire USF database.

×

3D Face Reconstruction from a Single Image Using a Single Reference Face Shape

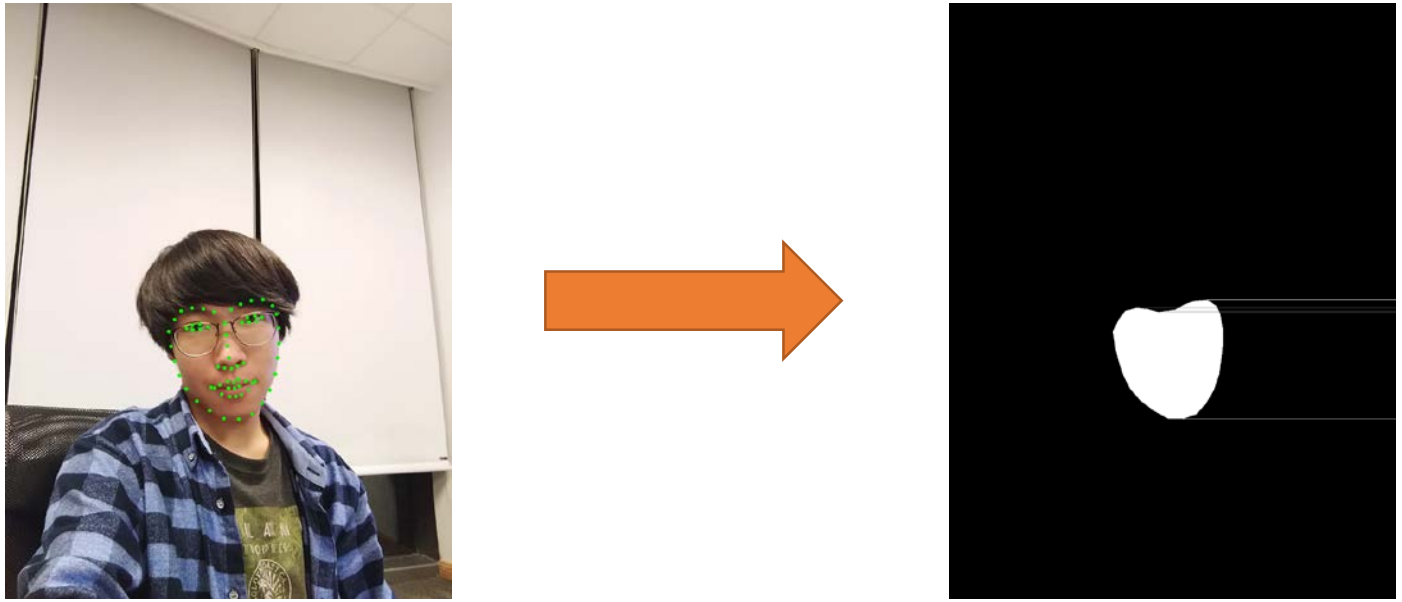
Ira Kemelmacher-Shlizerman, *Member*, *IEEE*, and Ronen Basri, *Senior Member*, *IEEE*

Region estimate

steps:

1.extract key points from faces(pretrained)

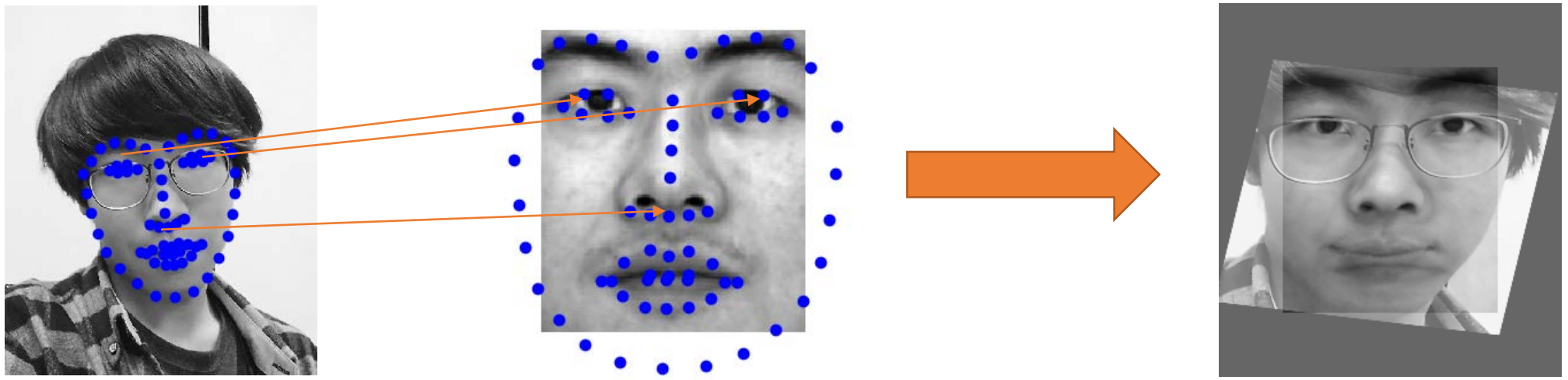
2.take the closed region of the rounding key points



Rough alignment

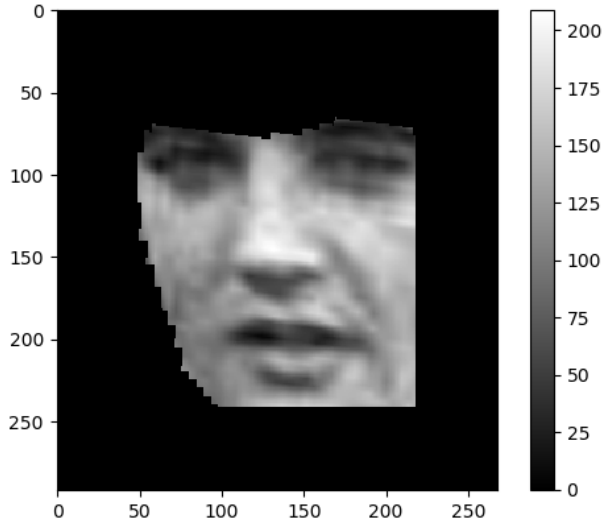
steps:

1.extract key points from faces(pretrained)
2.apply rigid transformation with three key points (two eyes and nose)
3.dereive the intersection of obj. face and ref. face



Rough alignment

The assignment of region Ω: intersection of obj face region and ref face region



Ω

Light coefficient (general Idea)

Nth order of spherical harmonic approximation, the light reflected by a Lambertian surface:

$$R(x,y) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} l_{nm} \alpha_n Y_{nm}(x,y),$$

2th order of approximation:

$$R(\vec{n}(x,y); \rho(x,y), \vec{l}) \approx \vec{l}^T \vec{Y}(\vec{n}(x,y)), \qquad \vec{Y}(\vec{n}) = (1, \boxed{n_x, n_y, n_z,} n_x n_y, n_x n_z, n_y n_z, n_x^2 - n_y^2, 3n_z^2 - 1)^T, \qquad I(x,y) = \rho(x,y) R(x,y).$$

Loss function:

$$\min_{\vec{l}} \sum_{(x,y) \in \Omega} \left(I(x,y) - \rho_{\text{ref}}(x,y) \vec{l}^T \vec{n}_{\text{ref}}(x,y) \right)^2.$$

Depth recovery(general Idea)

Data term:

$$I(x,y) = \rho(x,y)R(x,y). \quad R(\vec{n}(x,y);\rho(x,y),\vec{l}) \approx \vec{l}^T\vec{Y}(\vec{n}(x,y)),$$

$$\vec{Y}(\vec{n}) = \left(1, \frac{1}{N}p, \frac{1}{N}q, \frac{-1}{N}, \frac{1}{N^2}pq, \frac{-1}{N^2}p, \frac{-1}{N^2}q, \frac{1}{N^2}(p^2-q^2), \frac{3}{N^2}-1\right)^T,$$

$$N(x,y) = \sqrt{p^2+q^2+1}$$

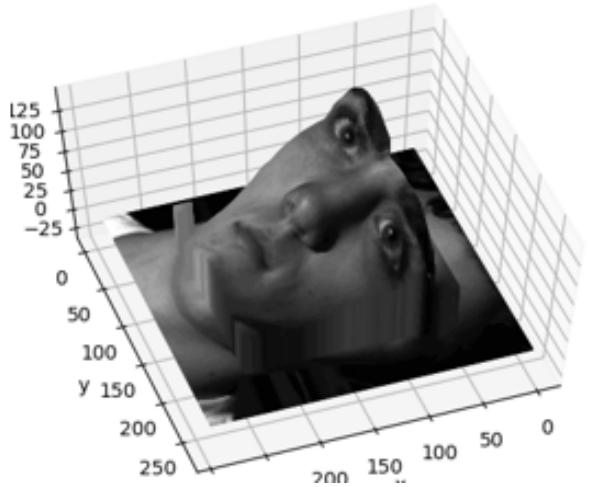
$$I = \rho_{\text{ref}}\, l_0 + \frac{\rho_{\text{ref}}}{N_{\text{ref}}}(l_1 z(x+1,y) - l_1 z(x,y) + l_2 z(x,y+1) - l_2 z(x,y) - l_3,$$

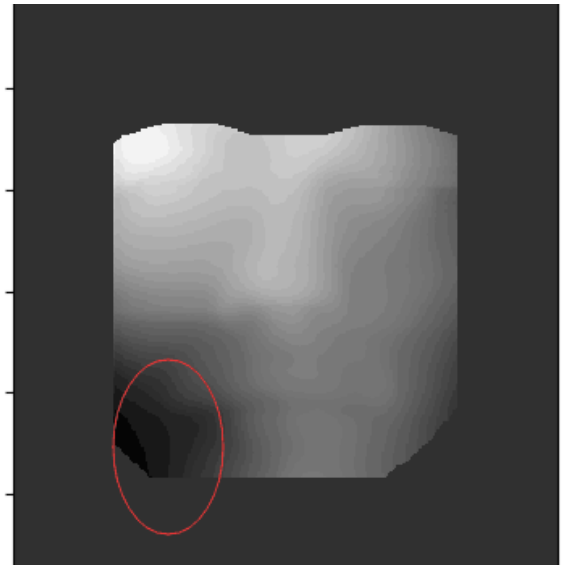
$$p = z(x+1,y) - z(x,y).$$
$$q = z(x,y+1) - z(x,y).$$

Regularization term:

$$\lambda_1(z(x,y) - G * z(x,y)) = \lambda_1(z_{\text{ref}}(x,y) - G * z_{\text{ref}}(x,y))$$

Experiments result☹



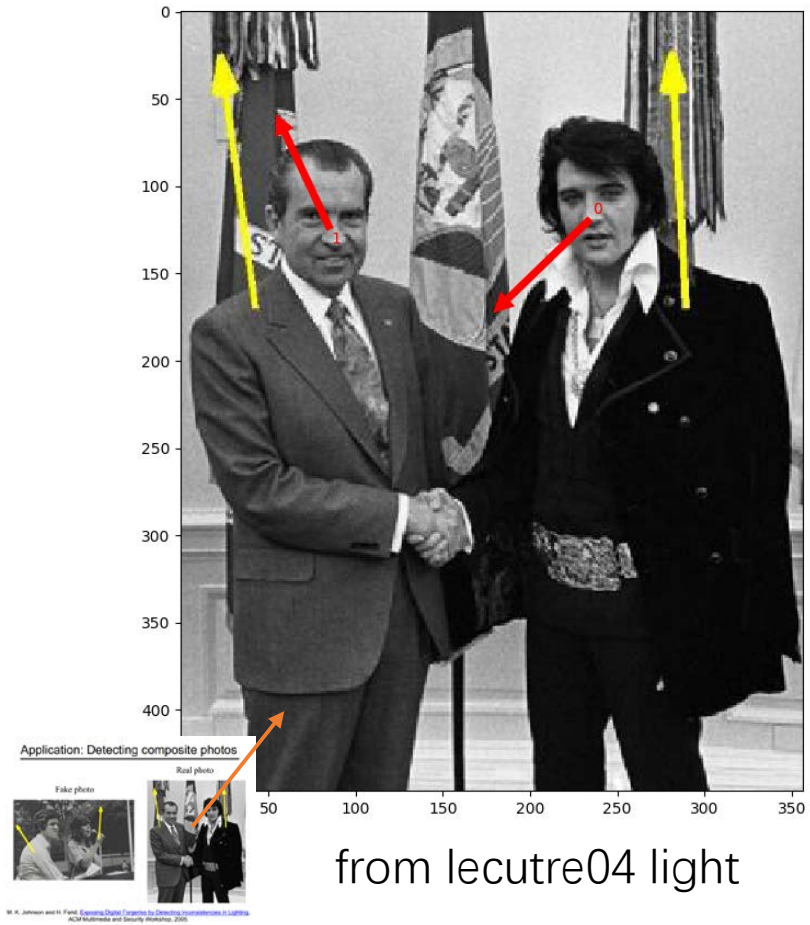


Causes

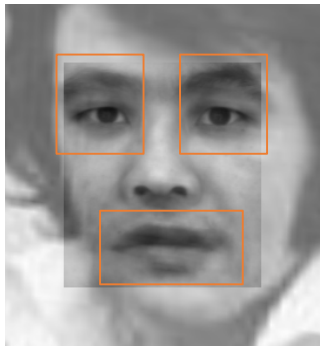
1. the reference model is not “mean” enough
2. Is the lighting coefficient estimating method effective? *
3. Is the depth assumption too rigid?
4. Ambiguous boundary conditions

Observations from lightening coefficients

- Unrobustness lightning evaluation （why?）

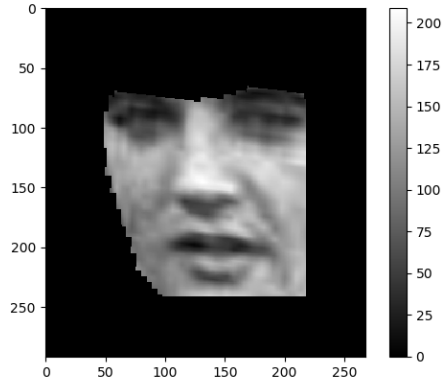


from lecutre04 light



huge difference between reference model and actual image at eyes at some regions

scaled reference model albedo

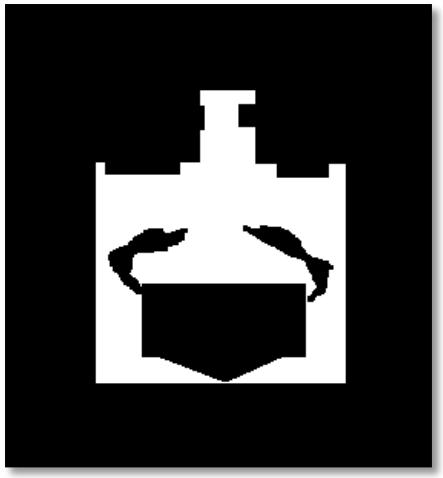


Intersection region Ω

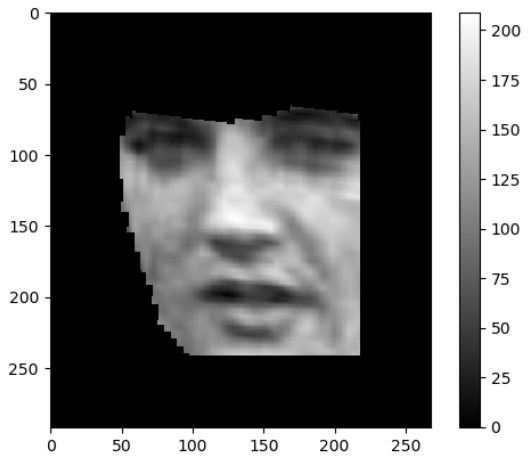
M. K. Johnson and H. Farid, Exposing Digital Forgeries by Detecting Inconsistencies in Lighting, ACM Multimedia and Security Workshop, 2005.

A more robust way

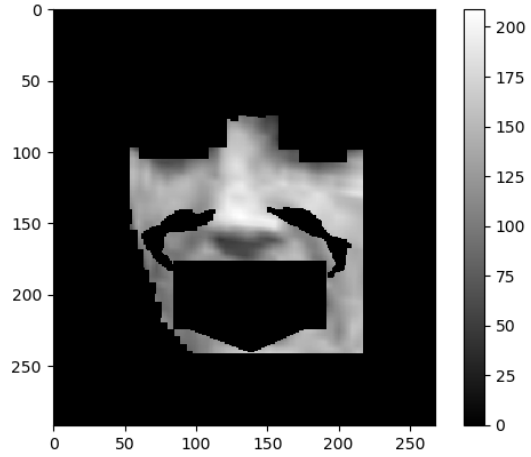
- only estimate the robust regions(manually assigned), the dark region is excluded from estimation.(i.e: to guarantee facial expression invariant)



mask



Intersection region Ω

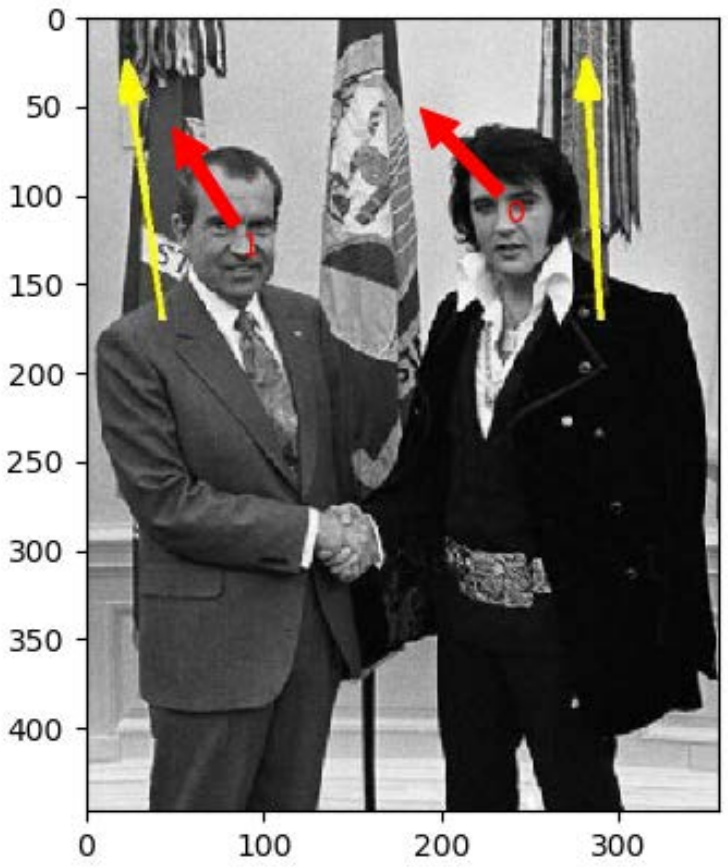


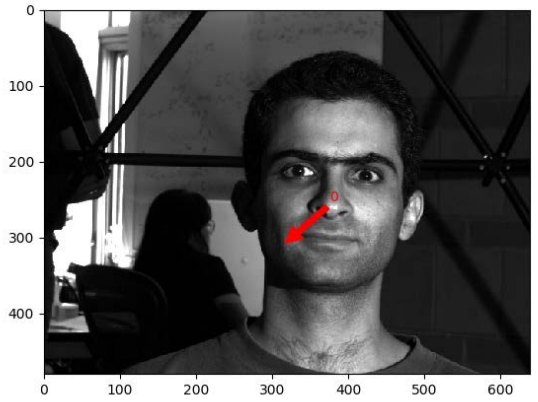
Intersection region Ω+

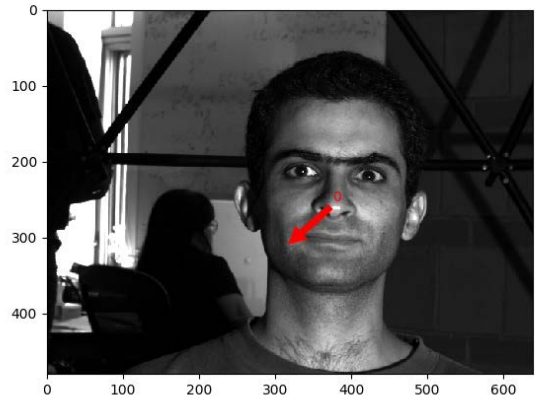
Experiment results

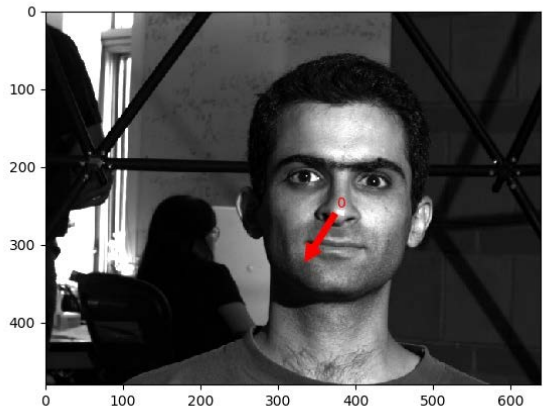
Images from extended yaleB database

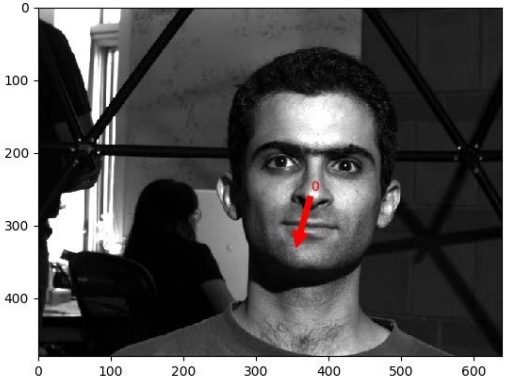
detailed information will be in report







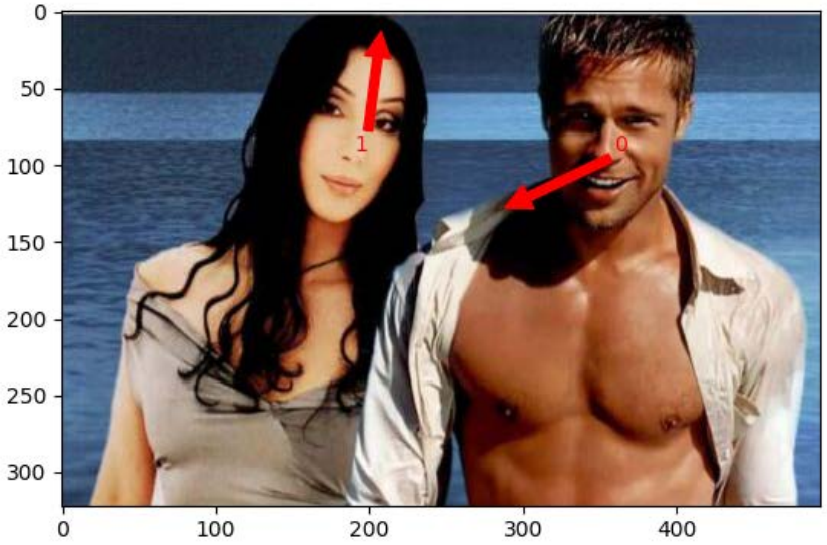




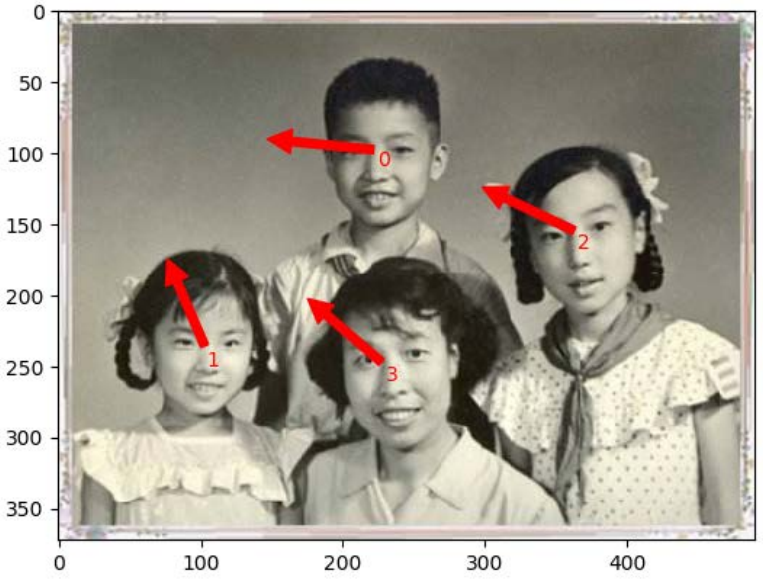
direction estimated from 1nd approximation term

Application scenario1

detecting composite photos



fake photo



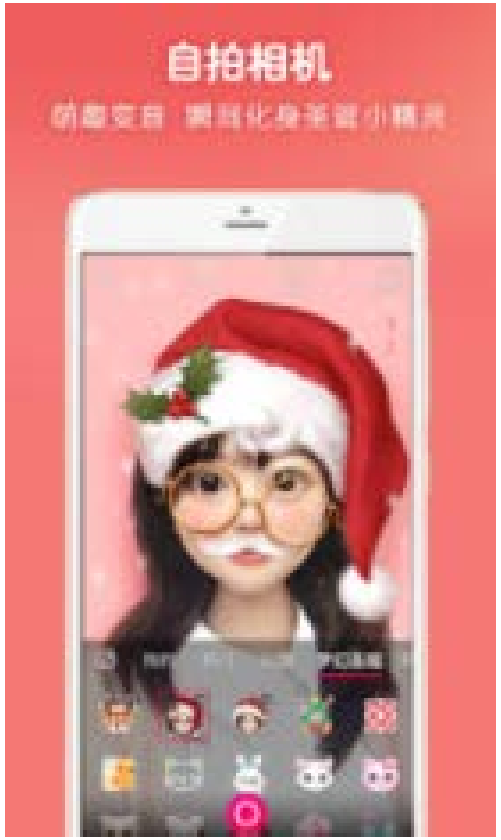
real photo

M. K. Johnson and H. Farid, Exposing Digital Forgeries by Detecting Inconsistencies in Lighting, ACM Multimedia and Security Workshop, 2005.

Pictures from the internet

Application scenario2

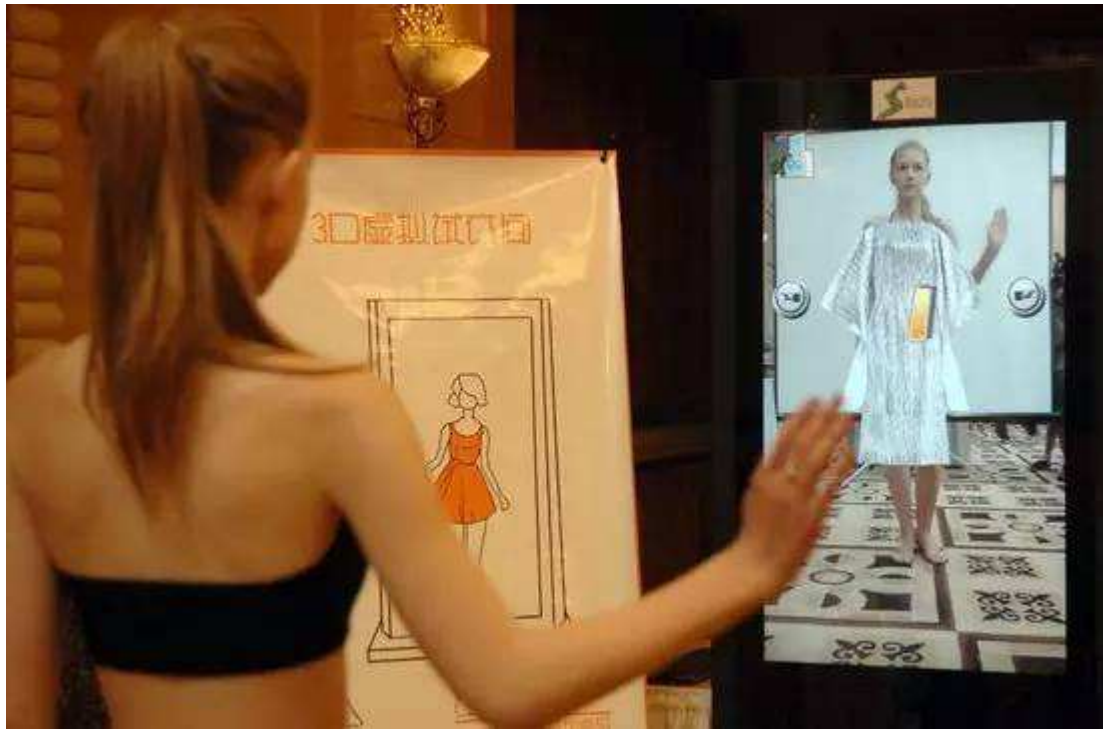
more natural compositing photos



√ with known surface norm information
√ with known albedo
√ with known light direction

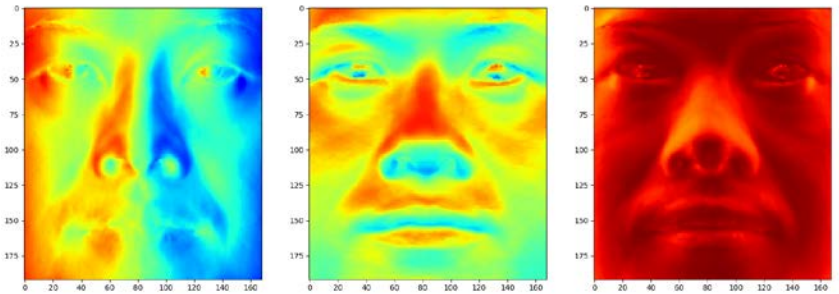
Application scenario2

more natural compositing videos



can apply lightning information to any 3-D model with known norms and albedo

We only have ..



Application scenario2

more natural compositing photos

estimated from 2nd order spherical harmonic approximation

Pictures from the internet

Q?&A!